MICROECONOMETRIC ANALYSIS OF THE RESIDENTIAL LOCATION DECISION: THE CASE OF KANO, NIGERIA

Mohammed Aminu ALIYU

Thesis submitted for the degree of
Doctor of Philosophy

School of Economics
University of East Anglia, Norwich

February 2010

MICROECONOMETRIC ANALYSIS OF THE RESIDENTIAL LOCATION DECISION: THE CASE OF KANO, NIGERIA

Mohammed Aminu ALIYU

Thesis submitted for the degree of
Doctor of Philosophy

School of Economics

University of East Anglia

February 2010

Abstract

This thesis is concerned with the dynamics of the urban housing market in the context of developing countries, with a focus on the demand side. The focal point of the thesis is an attempt to answer the question of what are the factors that determine the individual's choice of residential location within a city of a developing country. Cross-section data on individual choices made by residents of Kano, Nigeria is used in this quest. Two factors of particular interest are the quality of water supply and electricity supply. The question of how important these factors are in the residential location decision may be reformulated in terms of an individual's Willingness-to-Pay (WTP) for an additional hour of water or electricity per day. This valuation is estimated using a discrete choice model, in which the choice of residential location is modelled directly, with water and electricity supply in different locations included as factors, in addition to rent, influencing this choice. This method results in significantly positive valuations of the two amenities: an additional daily hour of water supply for a period of one year is valued at around 650 Naira (about £5); an additional daily hour of electricity supply is valued at around 400 Naira (about £3). These values are of similar order of magnitude to the daily salary of a middle-level civil servant. (Comparisons are at the time of the data collection).

A second approach to the valuation problem that plays a prominent role in the thesis is the hedonic pricing approach, in which the two variables (water and electricity supply) are included as explanatory variables in regression models with price (annual rent) as the dependent variable. This method gives rise to considerably higher valuations for the two amenities than does the discrete choice model. However, a crucial point is that any estimate obtained using the hedonic pricing method must be interpreted as an *upper bound* to the total welfare improvement resulting from an improvement in provision of public utilities.

Following estimation of the choice model, the assumption of Independence of Irrelevant alternatives (IIA) is tested. The null hypothesis of IIA is broadly accepted for this application, meaning that a nested choice approach, or a multinomial probit

approach, is not required. This result led us to a new research question: for what sort of study is IIA most likely to be accepted? This question is answered using a form of meta analysis, in which the IIA test results from 181 different published and unpublished studies are combined and analysed in a regression framework. The key findings from this Chapter are that: studies of employment choice, health care/medicare choice, and environmental and natural resource valuation choice are the most likely to result in acceptance of IIA; the probability of detecting IIA violation rises with the sample size; the Hausman McFadden test is less likely to detect IIA violation (ceteris paribus) than its principal competitor, the Small-Hsiao test. This last result is consistent with evidence from previous work in the form of Monte-carlo studies. A probit model of publication is also estimated, which yields the interesting conclusion that the probability of a paper being accepted for publication is maximised when a choice set consisting of exactly three alternatives is modelled, when the Hausman-McFadden test is used to test IIA, and when estimates from a multinomial probit model are reported.

Acknowledgements

I would like to thank Bayero University Kano for granting me fellowship and paying my tuition fees (from the Bayero University/Macarthur Foundation scholarship scheme). I am particularly grateful to the following principal officers of the University. The former Vice Chancellor, Attahiru Muhammad Jega, the Deputy Vice Chancellor, Yahuza Muhammed Bello, and the former Deputy Vice Chancellor, Muhammad Kabir Bichi and former Registrar, Farouk Yanganau for their official and personal intervention.

Bayero University also awarded some funds from the university research grant to me in 2006, a modest sum, but very useful when I ran out of money during my field work. Former Deputy Vice Chancellor, Muhammad Kabir Bichi was particularly helpful facilitating the speedy release of these funds.

My friends and colleagues in Nigeria and diaspora helped in different ways towards this endeavour. They are: Mohammed Kabir (MK) Aliyu, Lukman Ibrahim Diso, Oladipo Fashina, Abdullahi Sule Kano, Abubakar Momoh, Danbala Danju and Muttaka Usman. The following handled the “home front” in my absence: my friends Shehu Aliyu Rano, Kamal Usman, Ahmed Ali Yakasai and my brother Muktar Aliyu Harazimi. I shall forever remain grateful to them.

At the Bayero University, Kano I received valuable comments from colleagues at the Department of Economics seminar and on my questionnaire from the following colleagues. Yusuf Adamu, Afolabi Falola, Emmanuel Olofin and Muhammad Yakasai. My data collection was coordinated by my good friend and former student, Aminu Fagge.

I appreciate the hospitality of the following colleagueas, fellow students in the UK. Saminu Ilyasu Bala, Farouk Sarkin-Fada, Ibrahim Idris, Abdulsalam Yayo, Umar Ndako Bida, Haruna Musa Dambatta, Surayya and their kids, Muktar Atiku Kurwara, Ahmed Ali Yakasai and Hamida, Nafiu Hussaini, Hassan Adamu Shitu, Rufai Ahmad, Nuhu Bello, Murtala Mohammed and Mohammed T. Jimoh.

My stay in Norwich was made joyful and memorable by the following Twaib Ali, George Allabison, Saeed Alkatheeri, Mubarak Aldwsry and their families, Lamin Fatajo, Momodu Jobe, Angela Leeding, Lesley Brown, Tony and Paulina, Sara Ganjaei, Mahsa Ganjaei, Sue Carpenter and numerous PGR colleagues.

I am grateful to Dr. Philomena Bacon for sparing her valuable time look at my draft thesis, and her two sons, Linford and Kieran, who did most of my data entry.

I am highly grateful to Sajid Javed, his wife Dr. Samina Yasmin, and their two lovely sons, Ahmad and Muhammad Hassan for their friendship. Sajid is a man of enormous patience and hospitality. He particularly tolerated what he considered “my near-puritan attitude to life”.

This research has been influenced by and I have personally benefitted from attending the following external courses. Two courses organised by the Courses in Applied Social Surveys (CASS) University of Southampton - Nonresponse Rates and Nonresponse Bias in Surveys, by Robert Groves; and Regression Methods for Survey Data, by Mac McDonald. I attended the Discrete Choice Analysis: Predicting Demand and Market Shares, EPFL Lausanne, Switzerland. I also attended five Cemmap training courses - Masterclasses, Manuel Arellano - Nonlinear Dynamic Panel Data Models; and Charles F. Manski - Identification and Decisions with Social Interactions; Discrete Choice Modelling Course, by William Greene; and Introductory Microeconometrics, by Alan Duncan; and Course on Policy Evaluation Methods, by Barbara Sianesi. Graduate Students Association (GSA) University of East Anglia, provided some funds to me (£120) on three occasions to support these workshop/course attendance.

I would like to register my unqualified gratitude to my supervisor Peter Grant Moffatt for his academic, moral and financial support. This gratitude notwithstanding, the usual disclaimer applies. While I share all positive aspects of this research, I accept responsibility for all its limitations.

Finally, my family, my parents - my ailing dad and my mom, my brothers and sisters endured my absence. I would like thank them for their love and emotional support. I apologise for the occasional, albeit ominous, silence.

Dedication

This thesis is dedicated to:

My parents, brothers and sisters. For their love and support.

and

My late uncle and neighbour Muhammad Aminu who died while I was in the UK and my aunt late Khadija Usman for her love and support. I still remember her daily breakfast for me and my sister, punctually served, in my final year in primary school.

Table of Contents

List of Abbreviations and Acronyms

ASC:	Alternative Specific Constant
ASCLM:	Alternative Specific Conditional Logit Model
ASMPM:	Alternative Specific Mixed Probit Model
BUK:	Bayero University Kano, Nigeria
CBD:	Central Business District
CDF:	Cumulative Distribution Function
CLM:	Conditional Logit Model
FT:	Financial Times
GCES:	Generalized Constant Elasticity of Substitution
GEV:	Generalized Extreme Value (Model/Distribution)
HM:	Hausman-McFadden IIA Test
IDEAS/RePEC:	RePEc - Research Papers in Economics Database
IIA:	Independence of Irrelevant Alternatives
LAD:	Least Absolute Deviation
Lowess:	Locally Weighted Scatterplot Smoothing
MLE:	Maximum Likelihood Estimate
MNL:	Multinomial Logit Model
MNP:	Multinomial Probit Model
nlcom:	STATA command - Nonlinear combinations of estimators
NLM:	Nested Logit Model
OLS:	Ordinary Least Squares
OPROBIT:	Ordered probit
RP:	Revealed Preference
RUM:	Random Utility Model
SH:	Small-Hsio IIA Test
SP:	Stated Preference
UEA:	University of East Anglia, Norwich
UNICEF:	United Nations Children's Fund
WHO:	World Health Organisation
WTP:	Willingness to Pay

List of Figures and Tables in this Thesis

Figures:

Tables:

Chapter One: Introduction

When this project first started, the objective was to identify and quantify the determinants of the residential location decision in an urban area within a developing country. At that time, we had many determinants in mind, ranging from positive factors such as proximity to family, place-of-work, schools, and shops, to negative factors such as the levels of air pollution, noise pollution and traffic congestion.

As the empirical work progressed, the focus was gradually narrowed to the effects of two particular household amenities: the reliability of water supply; and the reliability of electricity supply. These factors are represented by variables obtained through the questionnaire on the number of hours per day, on average, that the household has access to a supply of water or electricity. Our decision to focus on these two factors is for two principal reasons. Firstly, the empirical results suggest that the impact of these factors is very important: as we shall see, individuals appear to place a high valuation on an additional daily hour of water and electricity supply.

Secondly, we realised that the data set that was being analysed, being from a city within a developing country, was ideally suited to the estimation of this sort of valuation, simply because none of the sampled households enjoy 100% reliability of supply. In developed countries, by contrast, water and electricity supply to households tends to be taken for granted, and is typically 100% reliable. This means that a similar data set obtained from a developed country could not be used to obtain the valuations of these amenities. Hence we are exploiting the lack of reliability of supply in our study area in the estimation of individuals' willingness-to-pay for improvements.

1.1 Housing location decisions

The research started as an attempt to provide answers to the following questions. Why should households choose to reside close to noisy places such as airport and city centres, and hazardous places such as industrial areas? We know from theory that environmental quality is a normal good, i.e. at higher income people demand better environmental quality. Hence we would expect those with lower incomes to gravitate to the less desirable locations. However, what we see in our study area is a socio-economic mix of households residing in these apparently undesirable locations. What factors are guiding the choice of better off households who presumably have the means to reside in more desirable locations?

Is the household's location decision also based on myopia, heuristics or ignorance? These sorts of explanations are unsatisfactory especially if among the households members are educated middle and upper class individuals. Is the household's decision an expression of preference towards or willingness to tolerate environmental bad (noise, congestion and air pollution) in increasingly expanding urban areas?

To address these questions, we set out to undertake an empirical study of housing-location decisions by individuals in Kano city, Northern Nigeria.

A cursory look at the situation in Kano city would provide some of the plausible reasons for the relative (un)importance of pollution and environment hazards in the residential location decision. In most cases individuals choose to live near to their places of work in order to reduce the cost of commuting in terms of both time and expense. Workers and businesses are willing to accept some level of inconvenience, including pollution, in order to reduce expenses on transport. This is especially true in view of the poor state of the public transport system in Kano. Another reason is poor planning and lack of enforcement. Kano city has grown in such a way that it has swallowed the airport and industrial areas previously considered to be on the outskirts, with no demarcation and with all the attendant environmental consequences.

But for us, the most important reason is the spatial differences in the provision of public utilities. In most developing countries, economic growth is associated with a shift from subsistence agriculture into manufacturing, with resulting urbanisation and increases in investment in infrastructure. The marginal cost of the provision of infrastructure is very high because of the relative financial strength of developing countries. These problems combined with a conscious decision to promote uneven development in favour of the industrial sector, and the unbalanced growth strategy, have resulted in skewed provision of public facilities in most cities and in particular Kano city, our study area.

Industrial estates and airports enjoy good provision of public utilities. Communities neighbouring airports and industrial areas enjoy positive externalities. The industries and airports have in addition also created economies of scale with small-scale markets to cater for the new settlement. The observation that well-off households choose to reside in these locations suggests that these positive externalities are outweighing the negative externalities of poor environmental quality.

Therefore, the issue we intend to address is how does this skewed supply of public utilities affect residential housing location decisions, and how can willingness-to-pay (WTP) for these utilities be estimated. This research is all the more interesting because we know that Kano city has become more cosmopolitan; people with different socio-economic characteristics can be found in all parts of the city trying to capture the benefits of proximity to areas with better supply of public utilities. The data on public utilities (water and electricity supply) can be considered as representative of a reliability index within the city.

The rest of this chapter is set out as follows: In section 1.2 we articulate our motivation for undertaking this research. Section 1.3 provides a brief analysis of the provision of public water and electricity supply, covering issues such as availability and who is responsible for the provision of public water and electricity supply in the study area. A very brief history of Kano city follows in section 1.4. Finally, section 1.5 outlines the plan for the rest of the thesis, with a breakdown of chapters.

1.2 Motivation

The principal objective of this thesis is to estimate the WTP for two public utilities namely: public water and electricity supply. We use rent and housing location choice data. This is based on the fact that, when households rent a housing unit, they obtain not only the physical property, but because of its spatial fixity, the neighbourhood characteristics and public services (Arnott, 1987).

In the pursuit of this objective, two methods are employed. The first is the hedonic pricing method, which looks at the physical and neighbourhood attributes of housing units. House price data are used to determine consumers' valuations of housing attributes and housing-related public goods. The second method, discrete choice modelling, looks at the household's residential location choice decision, and considers how this choice depends on the locations' attributes and the households' characteristics.

The study of how households form a decision on where to live is a difficult one, involving a multitude of factors. However, there are reasons for believing that it becomes more complex in the context of a developing country. Some of the factors that makes developing country study complex are: relative economic underdevelopment; poverty; absence of coherent government planning; inadequate and skewed provision of public services; urbanisation and rapid expansion of the city, etc. From a policy perspective, it is important to understand how people behave, and the value they attach to the provision of public utilities, in order to set policy priorities. It is also an interesting area from the perspective of both the economist and the econometrician, to analyse household decisions empirically, and to test existing theories with a new database.

The hedonic pricing model, also called the bid-rent model, assumes that the price for a housing unit is attained at a point where there is a match between suppliers and consumers to obtain equilibrium for a particular housing unit with given attributes. Individuals are assumed to bid for housing units, based on a constrained utility maximization framework and housing units are occupied by the household with the highest bid for that particular unit. The hedonic pricing model assumes each individual economic agent is unable to influence the outcome and take equilibrium function as given which is a constraint on their choices. From these assumptions it is possible to obtain household's marginal decisions on housing units' (structural and neighbourhood) characteristics.

The discrete choice model yields estimates of indirect utility function parameters rather than bid-rent function parameters. If we assume that each individual's preferences can be characterized in a random utility framework, we can describe the probability that a particular individual would chose (i.e. is the highest bidder/willing to pay the asking price/rent for) a particular housing unit. If we can observe both the choice and the price, the discrete choice model can also be used to estimate individual's WTP for housing attributes (Bartik and Smith, 1987).

Several hedonic pricing studies have been undertaken for housing in developed countries but there are few applications to developing countries. Most of the discrete choice studies of residential location decision that we came across are theoretical works with very few empirical works even for developed countries cities. We could argue that even though several studies have been undertaken in Western developed countries it would be interesting to study a developing country city because individual’s economic decisions making have been found to differ significantly along cultural lines and geographical space. There is also evidence to show that both parametric and strategic decision making are sensitive to the national and ethic origins of subjects (Chuah et al, 2004).

More important, the previous literature focuses on house price data. This study, to the best of our knowledge, is the first to use rent data. Further, this is the first research to use water and electricity supply reliability data, measuring hourly supply of tap water and public electricity to households. Although there are a small number of water valuation studies in our literature review, ours is the first to estimate WTP for public water and electricity supply.

1.3 Water and Electricity Supply in Nigeria

Here we introduce the state of supply of water and electricity in Nigeria. In Chapters 6 and 7 we discuss the implications of these issues on individual households in relation to available alternatives and individual valuation and WTP for each of these utilities.

The study area like most developing economies is characterised by poor economic infrastructure. Interest rate and the price level (crucial to economics agent’s current and future decisions), are highly volatile. Inflation is high and increases on annual basis and the interest rate is higher than expected. In addition, there is persistent exchange rate misalignments and instability which invariably affects current and future demand and investment decisions.

Until recently most public utilities were provided by the government which uses non-market determined, populist utilities pricing policies. This is typical to both developing and emerging economies with a history of socialist regimes and independence/liberation struggles. In most former colonies, the post independent *dirigisme* [1] ideology and socialist ideas of the liberation movements led to the emergence of an "oversized state". The state was involved in the provision of all sorts of goods such as household goods, public transport, leisure goods and public utilities. The pricing policy of government enterprises is in most cases arbitrary. Since pricing is not dependent on marginal productivity and since resource allocation is not guided by "Pareto efficiency", efficient provision is hindered. It is therefore likely that households' true valuations of public goods and investments in public services are not reflected in the observed allocation of resources.

Inefficiencies associated with bureaucracy, corruption, the ascendancy of the market and the collapse of the "developmental state" (the rise of the "minimalist state") have led to the privatisation (and "commercialisation") of most public enterprises. However, public utilities especially pipe-borne water supply is still controlled by the government.

After the return of democratic rule in 1999, the federal government began the implementation of energy sector reforms. The central government-owned and controlled power supply company (the national electric power authority) was broken down into eighteen separate companies as precursor to its privatisation. The eighteen companies comprise of eleven distribution companies, six generating companies, and a transmission company. Several independent power projects were also started. These were financed largely by a number of state governments in partnership with a number of private companies.

---

[1] A term designating an economy in which the government exerts a strong directive influence.

However, few private investors have indicated ***genuine*** interest in the electricity generating sub-sector. It appears, only the distribution sub-sector has attracted serious interest from the private sector. This is largely because of the huge capital requirements, fiscal problems and long-term nature of investment in generating electricity, weak local economic base (capital market and money markets) and lack of interest from foreign investors in the Nigerian economy.

The fiscal problems are partly due the volatility in the world oil prices, the principal source of foreign exchange, which accounts for more than 90% of government revenue. However some have argued that the problem, as with so many problems in Nigeria's public sector, is not so much lack of money, but management. For example, more than $10bn had been devoted to the sector from 1999 to 2006 with no visible result. A recent Financial Times special report on Nigeria (FT 2008) provides a detailed assessment of the parlous state of the Nigerian electricity sector. The FT argues that the state of electricity supply is due to decades of underinvestment and corruption and mismanagement. According to the FT report, the problem has left the country with enough capacity, on average, to power one light bulb per person, or an average supply of electricity to businesses and households for about five hours a day.

According to FT (2008) the government has kept tariffs so low that power plants run at a loss and therefore it does not make business sense to invest in the sector. The government plans to phase out gradually this huge tariff subsidy until the sector can run on a purely commercial basis, with stepped increases from an average of N6 per kilowatt hour through to N10 from July 1, 2011. This would allow distribution companies to maintain low tariffs while still covering their costs. The question is how to determine the commercial rate by the government and whether the subsidy will be enough to encourage private sector investment, with some analysts arguing that it would be better to leave it to commercial power providers and customers to agree their own pricing, rather than involve a government regulator.

Clean water is considered necessary for sustenance of life and sanitation especially in Africa where most diseases are water-borne and preventable. Broadly, there are five sources of domestic water supply in Nigeria namely, pipe-borne, borehole, shallow well, water vendors, and stand alone street pipes. Another source of water, mostly found in rural communities, is water from rivers and streams. It is also possible for households to combine two or more sources and to store water in tanks and containers. The nature of the source of supply is determined by natural factors such water table and household economic status. Although it is possible to obtain water from different sources and to store water (e.g. in tanks and containers), in this research we are interested in quality water, public supplied piped-borne water.

In Nigeria, pipe-borne water supply is controlled by the state government, the second tier of government, in a three-tier federal system. The federal government, the highest tier, design policy, facilitates the finance of domestic water supply projects either directly through budgetary provision or through loans from international finance institutions. Federal government also provides funds for agricultural water supply which could also be used for domestic water supply in the host and neighbouring communities. Because Kano is situated in semi arid region, there is a huge potential for irrigation. Both the federal and state governments have sponsored irrigation projects which include large scale water provision such as dams and canals.

State governments establish agencies which are responsible for the treatment and distribution of domestic water which is distributed in cities through a network of pipes and dispensed through the tap. Different distribution mechanism obtains in semi-urban and rural areas.

1.4 The study area

The modern Kano was established, as a political entity, around the year 1000 AD. Smaller and less organised settlements had existed in the area long before this period. Ranked 10th most populated city in Africa, it is also one of the major commercial centres in Nigeria. Currently it is the third largest and second most populated city in Nigeria with a population of about 3 million inhabitants (UN-HABITAT 2010).

Kano city was for centuries the center of caravan routes. In particular, it was one of the centres of the ancient Trans-Saharan trade to and from Sudan in the east, and Gambia, Senegal in the west. It served as a link between the Islamic north and West Africa and there are separate connections with cities in Central Africa (Ellicott, 2002, pp. 442).

Its importance in the Trans-Saharan trade was due to geographical and economic factors. It is almost midway from the East to the West coasts, on the edge of the Sahel as opposed to the Saharan climate, host to a very big and arguably the oldest international market in the country – the Kurmi market, and a very strong and flourishing handcraft industry at that time.

After independence it became home to several industries as part of the import substitution industrialisation project. Industries that sprang up include textiles, iron-rod, confectionaries, soft drink canning, battery, household cleaning products and car assembly. With few exceptions, these industries are largely concentrated in two industrial estates, Sharada/Chalawa and Bompai industrial estates.

The city comprises of eight local governments councils or boroughs. Each local government council has special jurisdictions, such as primary education, local roads, drainage and sanitation etc. Secondary and (partly) tertiary education, town-planning, provision of water supply and the connection of local communities to the national electricity supply are some of the responsibilities of the state government. Generation and supply of public electricity is carried out on a national scale by a federal government owned company.

Kano airport, built in 1936, was the first airport in Nigeria. Until the mid 1980s it is second busiest airport in the country. Recent political development notably, neglect by military rule, collapse of several manufacturing industries due to lack electricity supply, roads and other public services, neglect of the Kano airport by the Federal Government, have led to a drastic decline in the number of international airlines that patronise the airport.

Although some of the early empirical studies assumed a mono-centric city with an expensive and highly sought after Central Business District (CBD), heavy polluting industrial location and noisy airports (Arnott, 1987), one could argue that, as we show in Chapter two, most cities are unique and therefore there is no universal concept of a city. Non-centric and polycentric cities exist. And there are, to a certain extent, depending on the particular city, non-industry polluting sites like solid/municipal waste, dirty slumps and noisy road traffic in most urban areas.

It would therefore be contentious to describe Kano as monocentric because of multiple clusters of business activities such as big markets, industrial estates and employment centres such as government departments in a country where government is the major employer of labour. There also a serious problem of infrastructure which affects sanitation, town planning implementation and the structure of the city.

1.5 Breakdown of Chapters

Chapter two is a literature review on the economics of urban housing market. The chapter discuses the structure of the city in developed and developing countries, theories of land-use structure of cities, neoclassical theory of supply and demand for urban housing services, and the chapter concludes with an analysis of the theory of housing as a differentiated good.

Chapter three introduces the existing literature related to probabilistic (discrete) choice models. We discuss their theoretical foundation, underlying assumptions and mathematical derivations. The objective is to provide a clear perspective about the models we use in our empirical chapter on discrete choice housing location decision and the theoretical basis for the IIA meta-analysis chapter.

We introduce the hedonic pricing model in chapter four, the last literature review chapter. This chapter summarises the theoretical foundation of, and empirical issues on hedonic pricing model. Specifically the chapter discusses the following topics, functional form, sub-markets, identification and welfare change analyses in the hedonic pricing model and their implications for empirical work. This provides the basis for chapter seven on valuing utilities provision using the hedonic price model.

Chapters six and seven are the first and second of the three empirical chapters. The objective is the same for both chapters, to estimate the WTP for public utilities (water and electricity) in Kano city. Hedonic pricing model is used in chapter six to estimate the WTP for water and electricity supply. Because the data we collected is interval data we use interval regression. We had to control for physical, structural and neighbourhood housing attributes. We also consider the impact on welfare of different policies.

Chapter seven is a discrete choice analysis, we estimate five models, alternative specific conditional logit, mixed logit, nested logit, alternative specific multinomial probit and "mixed probit" models. The estimated coefficients are analysed and used to estimate the WTP. The contentious issue with using nested logit to estimated WTP is, how to determine location choice nesting structure in the study area. We solve this problem by looking at the variance-covariance matrix of two of our discrete choice models, alternative specific probit and mixed probit models.

Chapter eight is meta-analysis of IIA studies. Our discrete choice model passed the IIA assumption test. We also observe some pattern among housing and location choice studies. We therefore decide to analyse this issue empirically. To our knowledge nobody has done this type of analyses. We use two models, binary probit model for rejection and acceptance of IIA and ordered probit model for reported p-values and lastly. A second binary probit model is used to estimate the probability of publishing a discrete choice study.

In chapter nine, the last chapter, we conclude the thesis, summarise our major findings, and propose some recommendations both for policy and future research.

Stata Econometric software version 11 is used throughout this thesis.

Chapter Two: The Dynamics of Urban Housing Market

2.1 Introduction:

In this chapter we review the literature on the structure of the city and argue that every city has its own distinct character, or what in economic geography is called the *spatial-fix* of a city. We analyse the physical character of the city from economic perspective, its land-use policy, residential and commercial and industrial locations, and the physical, social and economic infrastructure that ties everything together. We also discuss demand and supply for residential housing, both as composite and differentiated commodity.

2.2 Urban Morphology - structure of the city in developed and developing countries

The structure of a specific urban settlement is determined by its history, geophysical properties and level of economic development. The level of economic development affects the ability of local authorities to provide amenities, to put in place proper town-planning/regulations and to control expansion. The German economist Adolph Wagner theorises that economic development is accompanied by an increase in public expenditure because of increased sources of public income and increased demand for public services (Musgrave, 1969). This theory aptly fits cities in western industrial societies, where increase in public investment, good governance, transparency and accountability lead to better planning, more efficient provision of utilities and modernisation.

This is in contrast with cities in developing countries, which suffer from a combination of adverse factors namely: corruption, poor tax base, low public income, absence of proper regulation mechanisms, and increasing population because of high birth rates and rural-urban migration. This trend has resulted in massive and uncontrolled expansion of cities in the developing world with municipal authorities unable to cope with demand for amenities. Another reason for the poor supply (both quality and quantity) of public amenities in developing countries is their state of technology. The state of technology at a particular time always affects the marginal efficiency of public investment. This means cities in developed countries have a bigger quantum of resources and obtain a higher return on investment relative to cities in developing countries.

Spatial distribution of demand for residential housing is determined by land use pattern and the structure of a city, where jobs are located; cost and ease of travelling to work; physical configurations of the existing housing stock; existing pattern of house prices; household preferences for location and dwelling types.

In this section, we attempt to summarise the theories of the structure and "agglomeration" of the city. The oldest and most simple spatial model assumes a monocentric structure of the city. This model assumes a single Central Business District (CBD) and looks at the economic landscape relative to its proximity to the CBD. The cost of renting (or buying) a house in "the city centre" is higher than a comparable house in the suburb. In developed economies, a household is faced with a trade off between living in a high rise building in the noisy city and a low rise building in the suburb. The only difference with a city in less-developed countries is the modest height of buildings in the CBD.

It has been argued that the monocentric model only describe the structure of pre-1950 cities because cities have since been transformed into polycentric. However, polycentric structure of the city could be easily explained as, an increase in the number of employment concentration areas, without substantially altering of the spatial relationships in the cost of housing services and or the nature of settlements (Kraus, 2006).

Four models have been used to explain the land-use structure of urban centres and how they affect households' quality of life. These are: Burgess concentric circles model; Hoyt sector model; Ullmann and Harris multi-nuclei model; and Waugh model of cities in developing countries. These models attempt to explain the nature of the built environment, its impact on the economy and social relationships within cities (Lind and Hellström, 2003; Lees et al, 2008; Atkinson and Bridge, 2005).

1. Burgess was the first to analyse the structure of land use in modern cities based on his observation of Chicago in early part of the 20th century. His model assumes that there is a major (cheap and efficient) network of transportation linking other parts of the city with the CBD, which is the central and most accessible location within the city. Burgess identified five clusters, a set of concentric rings within a particular city: The CBD comprising high rise buildings, major centre of economic activity and entertainment; light manufacturing with cheap and dirty residential settlements; low class residential housing dominated by workers seeking to reduce commuting costs; medium class housing, higher quality residential housing mostly private semi-detached houses with gardens; and the suburb, high class housing, domicile for the rich who can afford to live in this area and to commute to the CBD.

2. The Hoyt model assumes multiple sectors or wedge-shaped patterns within the city connected with the CBD by a network of transportation. It is anticipated that the rich will live close to main roads and commute to work. Because they can afford to live in any part of the city, they will choose places where there are better amenities, most likely, away from the noisy and crowded CBD.

3. Harris and Ullman multiple-nuclei model is based on the fact that most modern cities are polycentric, big in size, with suburbs and heavy industries on the fringes of the city. There is a CBD, with smaller business districts which are multiple-nuclei, linked to the main CBD and other parts of the city through a network of roads and rail transport. Although this is a more complicated model, it seems to capture the structure of most western cities in late 20th century.

   Modern cities in developed countries have over time seen a new divide between what are commonly called “down-town”, “mid-town” “up-town”, “East-end” and “West-end”. This demarcation is along both economic and cultural lines. The poor and less privileged are likely to live in “uptown”, the older and less developed part of a city. The part of the western cities known as “down-town” arose mainly due to expansion by the emerging rich class away from the densely populated area (with all their attendant problems) due to gentrification.

Gentrification includes demolition of old housing and new constructions and the construction of new houses on parks, playgrounds and green-fields. Two theories have been used to explain this phenomenon, demand side theories and supply side theories. Demand side theories argue that the phenomenon is due to changing preferences and demographic factors which might lead to an increased demand from high income groups for centrally located or more expensive housing rather than green-field, parks or playgrounds. Supply side theories or "gap theories" attribute gentrification to the presence of a rent gap and/or a value gap. A gap exists when the current rent or property value is far less than the potential value of the property. This gap makes it profitable for investors to enter the market (and in some cases influence policy) which causes change in the housing supply and the structure of a city.

In contrast to the phenomenon of gentrification, some older cities are experiencing what is described as the “doughnut-effect” (Walford, 2001; The Economist, 2002). These are big cities, in most cases with traffic congestion problems, where shopping centres relocate away from the centre to the fringe of the city, either to move away from the congested inner city or to find more space for themselves and parking for their customers thereby creating new “business parks”. This is also because, people with means are more inclined to live outside the city and commute into the city, with access to shopping centres and village markets which have been made accessible by “ring roads” and increased ownership of cars.

A more comprehensive theory is required to explain the land use and growth pattern in the contemporary western urban areas because of post World War developments such as industrial de-concentration, sub-urbanization, urban/suburban sprawl, and urban decay. To our knowledge this theory does not exist.

4. Waugh Model: The structure of cities in developing countries has some elements of the three models of Western cities. Most of these cities are monocentric with a CBD, industrial parks and suburbs. The major difference between this model and Burgess’ concentric model is the nature of residential areas where the CBD

comprise of high-rise apartment buildings, domicile for the rich. Unlike in western cities the quality of housing services and provision of public utilities decreases with distance from the CBD dominated by the service sector, corporate headquarters and other commercial activities. Waugh observes a marked difference in living conditions for the well-off with the less privileged and migrant labour force, living in the surrounding areas of the city (Waugh, 2003).

A more generic model which could be applied to both western and developing countries land use pattern and city growth is the dispersed city, corridor and compact city model. A city is described as a "dispersed-city" if it is made up of single-use, segregated sections connected by network of roads. A "compact-city", is characterised by high-density, mixed-use sections. The "corridor-city" model describe a city with multiple-use, semi detached sections with natural and artificial corridors such as green-wedges, streams and rivers, railway connections, canals/waterways, highways and transmission lines creating corridors within the city (Frey, 1999; Hirt, 2007).

2.3 Supply of Urban Housing

Arnott (1987) describes four different housing production process namely: construction, maintenance, rehabilitation, and conversion. Supply of housing can be measured in numbers and/or in terms of capital stock. Supply of housing services is determined by three factors. First, the supply of housing from new building; second, conversion of existing housing from one type of accommodation to another by property developers and property owners; and third, actions of existing owners, such as renovation and/or sub-letting, which increase the volume of old housing coming onto the market at a given period of time. House owners put their houses into the market, for sale or letting, due to changes in circumstances such as income, household size and job mobility (DiPasquale, 1999; Quigley, 1979).

Although there is no empirical evidence from the available literature as to which of the three factors is the major source of addition to the stock of housing, Quigley (1979) asserts that, greater percentage of the urban "housing services comprised of dwelling units from pre-existing stock of housing" and new buildings are usually only a small proportion of total housing supply. Although there is resurgence of interest in housing transformation in urban centres – gentrification – with attempts to analyse the conversion of housing services in certain parts of modern cities, we have not come across an empirical study of the impact of renovation/refurbishment/regeneration on the stock of housing. One potential empirical problem would be endogeniety. Modelling the impact of housing improvement on housing supply would be difficult to identify because some homeowners are both suppliers and consumers of housing services.

Other factors that affect supply of housing include the following:

- government policy – tax policy, such as subsidies, vouchers, public housing scheme, and mortgage support;
- municipal planning restrictions;
- availability of credit/mortgage determined by national wealth, personal income and financial institutions portfolio management;
- demand for housing relative to other non-housing property;
- rent-profit margin and market share optimisation decision of house providers;
- rental income versus maintenance cost;
- economies of scale - technology and marginal cost of production/construction of housing;
- availability of land and skilled labour;
- high rise and low rise buildings; and
- cultural factors.

Government support, subsidy and other measures, affects both supply and demand for housing services. It would increase demand for housing services through increase in owner-occupier housing, own-houses and private sector investment to meet the increased demand.

Supply of new housing:

We adopt Grienson and Arnott’s (1982) simplified version of Smith’s (1976) model of housing supply. A house developer would choose quantity Q, and a density, D, of housing so as to maximise profit, π, per unit of land.

The objective function is:

$$\max \quad \pi = PQD - K(Q,D) - R \qquad (2.1)$$

QD is the housing per unit area of land, R is the unit cost of land, P is the location-specific selling price and $K(Q,D)$ is the housing technology.

First order condition (FOC):

$$\frac{\partial \pi}{\partial D} = PQ - K_d = 0; \qquad K_d = \frac{\partial K}{\partial D} \qquad (2.2)$$

and

$$\frac{\partial \pi}{\partial Q} = PD - K_q = 0 \qquad (2.3)$$

The FOC relates the profit-maximisation location-specific quantity and housing density (which, in a general sense could be treated as an aspect of housing quality) to the location specific housing price. Because it is possible to invert the relationship its is also possible to relate housing prices at different locations to the corresponding prices and densities.

$$P = P(Q,D) \qquad (2.4)$$

If we assume that in the long run the economy is in competitive equilibrium with constant return to scale in the production of housing, then housing producers make zero profits. The objective function could be re-written as follows:

$$R = PQD - K \qquad (2.5)$$

To derive the elasticity of housing supply, we differentiate the value of housing per unit of land (what Smith (1976) calls the total expenditure), $E = PQD$, the new objective function (equation 2.5 above) and the new FOC, equation 4 to obtain equations 2.6 and 2.7:

$$\frac{dE}{E} = \left(1 + \frac{1}{\varepsilon_q}\right)\frac{dQ}{Q} + \left(1 + \frac{1}{\varepsilon_d}\right)\frac{dD}{D} \qquad (2.6)$$

$$\frac{dR}{R} = \frac{E}{R}\left(\frac{dQ}{\varepsilon_q Q} + \frac{dD}{\varepsilon_d D}\right) \qquad (2.7)$$

2.4 Demand for Housing

Aggregate demand for housing is broadly determined by two sets of factors, financial and demographic factors. Typically, financial factors determine short-run demand while demographic factors affect demand in the long-run.

Because housing is a normal good, we expect positive income elasticity of housing (with some exceptions). Household income is a limiting factor on the quantity of all goods and services households may consume. This includes housing and housing related services. Related to this is price of housing, the price and availability of substitutes (own house versus rented house) and complementary goods e.g. amenities. Tenure choice creates substitutes in the housing market. The two major options are rented and owned house with some variations/sub-classifications. High rent (relative to average income) could lead to an increase in demand for owner occupation.

In most cases residential houses are purchased with a mortgage. The cost (interest rate) and availability of mortgage affects the number of transactions in the owner-occupier segment of the housing market. It also affects the household tenure choice. The number of mortgage institutions/building societies and their financial portfolio; government financed mortgages/owner occupier scheme; and employer housing schemes in some countries, all affect the quantum of resources available and the ease of accessing loanable funds to buy houses and tenure choice decisions. In the short run, expectations about future price inflation also affect demand for both own-house and rented house demand.

In some countries and/or regions, authorities provide rent-supplement schemes as income redistribution scheme especially in affluent societies where the willingness to pay for goods has been affected by higher standards living or artificially priced upward because of high demand. Housing benefit scheme increases household's ability to purchase more housing services and other consumption goods through the substitution effects.

In the long-run both household and aggregate demand for housing services depends on demographic factors. These include the number of households in a particular city, household size and distribution.

Straszheim (1975) analyses the conventional demand for housing within the framework of monocentric model. His starting point is a demand analysis that seeks to explain how much housing services households wish to consume and in what location. Models by L. Wingo, W. Alonso, R. Muth, pioneers of the monocentric urban model, argue that households choose locations which minimise the sum of transport and housing services costs in order to increase their consumption of "all purpose consumption goods". This may involve a trade-off between cheaper rent and longer commuting time, expensive rent (probably smaller house) and shorter trip to work.

If we formulate a utility maximisation model of the household subject to budget constraints with housing and all other goods, with transport cost included in the utility function, Alonso (in Straszheim, 1975) argues that the budget constraint will depend on household's income, price of all purpose goods, cost of housing, and transportation. The optimal location, amount of housing services and other goods to be consumed will depend on households utility function and the opportunity costs.

The Alonso model

$$Max : u = U(z, q, w)$$

st:

$$Y = P_z \cdot z + P_q(w) \cdot q + P_w(w)$$

z	–	all purpose consumption goods
q	–	quantity of housing services
Y	–	income

P_z – price of consumption goods

$P_q(w)$ – price of housing services at w distance from place of work

w – distance from residence to place of work

$P_w(w)$ – cost of transportation to distance w

$$\frac{U_w}{U_z} = \frac{q \cdot \frac{\partial P_q(w)}{\partial w} + \frac{\partial P_w(w)}{\partial w}}{P_z}; \tag{2.8}$$

$$\frac{U_z}{U_q} = \frac{P_z}{P_q(w)} \tag{2.9}$$

$U - denotes \quad partial \quad derivatives$

Households choose the price (and quantity) of housing services it consumes by altering its commuting plan. For the amount of housing services consumed, marginal rate of substitution of all-purpose goods, for travel time, must equal the ratio of acquiring more housing services or having a shorter trip to work. At optimum, the location chosen, the marginal rate of substitution of z for q must equal the ration of their prices.

The Muth (1960) model is slightly different. It assumes that households consume homogenous good, housing, with distance to place of work left out of the utility function.

$MaxU(z,q)$

st:

$Y = P_z Z + P_q(w)w + P_w(w)$

$$q \cdot \frac{\partial P_q(w)}{\partial w} = -\frac{\partial P_w(w)}{\partial w} \tag{2.10}$$

Distance to place of work affects travel cost by an amount equal to the associated change in housing expenditure. This is a good premise on which to build our demand model.

It is our view that Alonso model is not sufficient to explain demand for housing as a differentiated good containing multiple attributes. The more appropriate model is Rosen type model of hedonic pricing model in Follain and Jimenez (1985b).

They provide a model of demand for housing attributes which is appropriate for estimating the marginal rate of substitution in consumption of housing attributes and non-housing goods. They assume each household consumes Z, a vector of housing attributes and X, a composite of all non-housing goods, subject budget constraint where income is exhausted by purchase of Z and X. The problem with set-up is, the price of attributes is not observed, only market rent is for entire bundle is observed. A two-stage method is used to estimate the parameters of the following model, from which we can derive the demand for housing as a differentiated good within the context of household total expenditure.

$$Max\,U = \left[\sum_{j=1}^{m} \alpha_j Z_j^{\gamma_j} + X^{\varepsilon}\right]^{\phi} + \lambda\left[Y - P(Z) - X\right] \tag{2.11}$$

$\alpha, \gamma, \varepsilon, \phi$ are parameters to be estimated; while λ is the Lagrange multiplier.

The FOC:

$$\left(\frac{\partial U}{\partial Z_i}\right) \equiv \phi U \left[\sum_{j=1}^{m} \alpha_j Z_j^{\gamma_j} + X^{\varepsilon}\right]^{\phi-1} \left(\gamma_i \alpha_i Z_i^{\gamma_i - 1}\right)$$

$$= \lambda \left(\frac{\partial P}{\partial Z_i}\right) \quad i = 1, \ldots, m$$

$$\left(\frac{\partial U}{\partial X}\right) \equiv \phi U \left[\sum_{j=1}^{m} \alpha_j Z_j^{\gamma_j} + X^{\varepsilon}\right]^{\phi-1} \varepsilon X^{\varepsilon-1} = \lambda$$

$$Y = P(Z) + X$$

Simplifying these yields:

$$P_i = \left(\gamma_i P_i \varepsilon^{-1}\right) Z_i^{\gamma_i - 1}\left[Y - P\left(Z\right)\right]^{1-\varepsilon} = \lambda$$

$$P_i \equiv \frac{\partial P}{\partial Z_i} \tag{2.12}$$

The right hand side variable of equation 2.12 is the marginal rate of substitution in consumption between housing attributes Z, and non-housing goods X.

2.5 Developing Countries Experience

The stock of residential housing or new housing construction as a proportion of gross domestic product increases at an early stage of development but on average declines after a certain point. But in absolute terms, empirical results show that countries with developed financial markets invest relatively more in housing. When total housing stock is measured in terms of the number of housing units, its growth is determined by demographic rather than economic variables. But for the quality of existing housing stock is determined by economic factors. That is to say, demographics determine the total housing stock while incomes and prices determine the quality of available housing services (Malpezzi, 1999).

Some studies suggest that developing countries have inelastic supply for housing (Malpezzi, 1999). In other words cost is unrelated to share of housing investment because at the initial stage of development it is quantity rather than quality that matters in the consumption of housing services. But in both developed and developing countries, increases in national and household income are associated with higher probabilities of upgrading, what is also called “filter down”, an upward mobility for households, where housing units pass from richer households (owners or tenants) to lower income households. However, in some situations, it becomes a case of “filter up” when an area undergoes “gentrification” units pass from poor households to richer households. We have not come across any evidence to show that, except for sentimental loss, gentrification involves welfare loss for households.

On the demand side, based on the experience of developed countries, there is three-stage transition theory of demand for housing services in developing countries. It is anticipated that there would be a sequential upward mobility, where low income households would be owners of housing units, very low in quality but as their income increase they would move first into the formal rental sector with improve housing services, and eventually become formal homeowners with well developed housing attributes (Malpezzi, 1999). This trend is not automatic and may not be linear. This is because the type of housing individual household acquires is sensitive to its income and demographic characteristics (Arnott, 1987).

Specific factors that could affect demand for housing in developing countries include the rate of population growth, rural-urban migration, paucity or complete absence of mortgage and government support for low-income housing, and the absence of state institution to regulate urban sprawl.

We have not come across any comprehensive study of the structure of African cities, individually or collectively. Gilbert (1996), a compendium of housing conditions in Latin American cities is the closest that we could find. The book analyse the conditions in developing countries cities, which they argue are clearly far from good. Of relevance to us, the study argues that too many people lack services and basic infrastructure. Competent town planning and urban management is vital in big-cities. However, it is extremely difficult to achieve the desired levels of competence in cities located in developing countries. Any city which is in financial difficulty will have problems in providing adequate infrastructure and services.

Since 1960 when Nigeria became independent, Kano city has grown in size, and like most large cities in developing countries have "swallowed" nearby villages and towns. The original structure of the city has been distorted, creating multiple business centres competing and complementing the CBD.

2.6 Housing as a Differentiated Good

The conventional economic theory is built on the idea that optimisation decisions involve a choice of goods and services at their respective prices in a market limited by individual budget constraints. The "characteristics theory" of consumer behaviour assumes that utility is generated by characteristics or attributes of goods and services. Instead of utility being a function of indivisible products it becomes a function of utility derived from attributes of goods and services. Two major theoretical contributions to the characteristics/differentiated goods demand model are the Houthakker/Rosen and Lancaster models (Eastwood et al, 1986; Ratchford, 1975).

Differentiated goods are goods for which there could be significant differences between various units of the product but consumers consider them to be members of the same general product class (Day, 2001). The market normally reflects in the price of these goods, their differences and consumers willingness to pay for various attributes or constituent units. Basic examples of differentiated goods are cars, cereals and residential housing. A car could be differentiated according to its engine capacity, fuel consumption per mile, passenger capacity, air-conditioning, type of wheels, sun-roof, central lock, automatic break system, air-bag, number of doors etc. A house could be characterised by the size and number of rooms, access to amenities, heating system, garden, garage and neighbourhood. Cereals could be soft, crunchy, with or without sugar, possess more or less calories, be made of wheat, rye, barley, oats, millet, rice, and maize etc.

The Lancaster model (Lancaster, 1966 and 1991) propose a theory of consumer utility based on characteristics rather than the good itself because, goods do not give utility to the consumer, but it is their characteristics/attributes which give utility. To put it differently, the consumer might not be interested in a good as a bundle, but its disaggregated constituents. It is from these characteristics (most of which are consumed collectively) that the consumer derives utility. Individual consumers, subject to budget constraints, seek to maximise their utility by choosing goods that will give them the best combination of desired characteristics.

This model is built on two propositions. First, all products possess measurable attributes relevant to consumer choice among different products and secondly, individuals differ in their valuation of different attributes rather than their assessment of the levels of attributes produced by the various products. That is to say that, individuals possess preferences for collections of attributes and the preference for products are indirect, valued because they contain attributes sought by the individual consumer (Eastwood et al, 1986).

In order to explain the decision making process involving multiple goods, this approach assumes that, the consumer's utility function is separable. The consumer is expected to allocate resources between "groups" of goods, and attempt to optimise within each group by selecting the best combination of characteristics within the group. The individual consumer will allocate resources between groups, for example accommodation, leisure, food, transport etc; she will subsequently make a choice within a particular group, to obtain the combination of characteristics which maximises her utility at the least cost.

The Lancaster approach recognises a more complicated analysis, where goods have many attributes, and these attributes could be shared by more than one good and that, combined together, goods possess attributes different to their individual attributes (joint demand attribute) (Wong, 2002).

To apply this analysis to housing consumption, we note that, the household determines its consumption of housing in conjunction with and reference to other non housing goods:

$$Y = pZ.Z + pX.X + S \qquad (2.13)$$

Where Y is disposable income; pZ - price of housing; Z – bundle of housing services; pX – price of non-housing goods; X – all non-housing goods; and S – savings and investment goods.

The Lancaster approach to the utility derived from differentiated goods (housing services) consumption bundle of the consumer can be expressed as:

$$U = U(Z_i, X, S) \qquad (2.14)$$

Where U is the utility derived by the household; Z*i* – bundle of *housing attributes;* X – the composite commodity, all non housing consumption; and S – savings and investment goods.

Housing attributes could be categorised as: dwelling; neighbourhood quality; and accessibility:

$$Z_i = D_i, N_i, L_i \tag{2.15}$$

Where Z*i* is the individual house characteristics – physical structure, rooms, size, toilets etc; D*i* – dwelling characteristics; N*i* – neighbourhood; and L*i* – location

Another way of looking at this, Houthakker postulated that a commodity could be described by two variables, its physical quantity and quality (Eastwood et al, 1986). In this sense, commodities with different attributes are treated as the same (in quantity) but variable in quality. With this premise, it is possible to estimate the price of attributes/quality from the consumers explicit choice.

It is possible to observed the market clearing price – the interaction of consumers with heterogeneous taste for different combination of attributes and the supply of goods with given attributes – and specific amounts of attributes associated with each to derive individual implicit or hedonic prices. The (modern) hedonic pricing, which is due to Rosen (1974), provides the functional relationship between the market clearing price of a good and its constituent attributes.

A rational consumer is expected to maximise her utility by consuming goods with given attributes subject to budget constraints. Consumer's willingness to pay will depend on her income and taste which determines her preferences for given combination of characteristics. The solution to this optimisation problem would require that the marginal rate of substitution between characteristics and the price of the good must be equal. The consumer's willingness to pay for an attribute must be equal to the implicit price of the attribute in the market.

Neoclassical, maximalist utility theory analyse how households rationalise housing needs given income constraints. Analysing housing demand by households who select from a menu of characteristics based on preferences in order to maximize their welfare, and housing supply by landlords, who produce houses with different characteristics who, thanks to providence, inherit some characteristics, and price their property based on costs incurred, with the aim making profit. The household is assumed to have an organised system of preferences, considerable knowledge and skills to evaluate alternatives and selects the alternative which yields highest utility (Wong, 2002).

But house characteristics could be divided into observed and unobserved characteristics - locational, structural and neighbourhood characteristics. Unobserved product characteristics are *independent* of the observed product characteristics. Even with many observations on a consumer's choices (such that the consumer's entire demand function is known) it is not always possible to *uniquely* determine consumer preferences (Bajari and Benkard, 2001). This problem of unobserved characteristics, coupled with consumer heterogeneity are some of the basis for one of the discrete choice models of consumer preferences.

2.7 Optimisation Decision in Differentiated Market

As in the market for homogenous goods, the market for differentiated goods consists of large number of buyers and sellers. The market clears at equilibrium through the normal price mechanism. However, unlike in the normal market, where one (equilibrium) price is determined, the equilibrium price for a particular product depends on its characteristics. For example, the price of a house is determined by number of rooms, their size, access, neighbourhood, garden size, parking space etc. It is the matching of supply and the market price for the commodity containing different combination of characteristics with the corresponding demand for the commodity by consumers with different taste for characteristics that leads to multiple equilibria.

Since the value of a good depends on the amount of characteristics it possesses, its price will be a function of the characteristics:

In short-run the supply of the commodity is fixed in quantity but qualitatively variable. While the supply of houses is fixed in the short-run, the aesthetics - e.g. paint, blind, garden, the heating facilities could be changed which alters Z and affects its price.

Unlike in homogeneous goods market, where a consumer is a price taker, in a market for differentiated goods a consumer can choose to pay different prices. A consumer has an option to purchase a good which contains low levels of $z_1, ..., z_n$, for which she pays a low price or high levels of $z_1, ..., z_n$, for which she pays a high price. The price she pays for a product which contains a given combination of characteristics is given by the price function, which can not be influenced by the action of any single individual consumer. Larger quantities of "desirable" or good characteristics would attract higher prices, while "undesirable" characteristics would attract lower prices. Although what constitute "desirable characteristics" is a subjective issue, it is expected that, these are also good characteristics which, in their own right, in a homogenous goods market would attract higher prices.

In a market for composite differentiated goods arbitrage is not feasible. That is, the differentiated good must be consumed as a whole. It is impossible to disaggregate a house and take units of z_1 (e.g. living within the town centre or the central business district) and combine it with units of z_2 (e.g. living next to the sea). This implies that, consumers are unable to repackage the product or breakdown products into constituent parts and consume the characteristics separately.

Day (2001) provide a basic example of this problem by comparing the choice between a house with two bedrooms with two houses each containing one bedroom; renting a house with four bedrooms for six months and a two bedroom for another six months, which is not equivalent to renting a three bedroom house for one year. The reason for this difference is because marginal prices of characteristics are not constant. Another reason is joint demand for characteristics; price of one characteristic may depend on the quantity of another.

2.8 Equilibrium in Housing Market

The equilibrium price is the matching of individual consumers, given their preference for a combination of housing attributes and income constraints, with suppliers of a given type of housing. It is therefore the maximum price consumers are willing to pay for a set of characteristics equated to the market price, which is the minimum landlords are willing to accept for a house with a given combination of characteristics.

While we could establish equilibrium hypothetically, there is a debate on whether in reality equilibrium exists in the property market. However, the most crucial information for many environmental issues is contained not in equilibrium but in the consumer side of the market. Ignoring the producer side does not create theoretical or econometric problems. The property market is likened to stock-flow model in which changes in stock is a function of price, but prices are determined only by available stock at the time. Supply side is less crucial in property market studies because, the quantities of the characteristics in the existing property are predetermined and difficult to alter, and therefore the equilibrium price schedule is completely demand driven. In most empirical works in housing economics, cross section and aggregate data are used and therefore need not bother about the supply side (Palmquist, 1984), (Freeman, 2003).

Equilibrium may not exist because of the unique nature of the housing market and some of the characteristics of housing namely: housing as a necessity good which takes a large proportion of household's income and wealth; complex bundle of attributes comprising necessary, luxury, asset, consumption goods, and different elements of housing services; durability and imperfect malleability/spatial fixity, which makes adjustment on the supply side slow; indivisibility in consumption; property as investment asset (longevity of the investment and specificity/irreversibility of the asset); dependence on finance market by both consumers (mortgage) and producers (loan); the existence of market imperfections on both demand and supply sides - transaction cost, and information asymmetry; importance of on housing in social policy (Anas and Arnott, 1991), (Watkins, 2001)

Chapter Three: Probabilistic Choice Models

3.1 Introduction

In this chapter we introduce probabilistic choice modelling in general, paying particular attention to the multi-alternative discrete choice models that we use in Chapter 7 to analyse our residential location choice data. The principal example used throughout this chapter is the choice of travel mode. This is chosen as an example, partly because it is the standard example used in the discrete choice literature, and partly because many of the insights gleaned from this example are directly transferrable to the residential location choice problem considered in Chapter 7.

In the context of travel mode choice, we pay particular attention to how the estimation results may be used to estimate a commuter's "value of time". We focus on these techniques, because the same techniques are used in Chapter 7 to obtain estimates of households' valuation of water supply and electricity supply.

3.2 The Discrete Choice Framework

The Random Utility Model (RUM) is the behavioural side of the discrete choice model. It helps us to derive and interpret discrete choice models. Because discrete choice involves choice process, we have a **decision maker**, the individual person or household (can be a business establishment or a corporate organisation). In the context of the travel mode choice example, the decision-maker is an individual commuter. Every decision maker can be associated with a vector of **characteristics**. For individual decision-makers, we are most interested in socioeconomic characteristics such as age, income, gender, education, etc.

The decision maker selects amongst **alternatives** from a (finite) universal choice set (all possible alternatives), and it is usually assumed that one alternative must be chosen, i.e. the universal choice set is exhaustive. If some individuals choose none of the alternatives, "none of the above" could simply be added to the choice set, making it exhaustive. It is sometimes assumed that a given individual has access to only a subset of the universal choice set, and we refer to this subset as the feasible choice set.

The choice set can be the same for all individuals in the population or individual specific because some alternatives are not available to all individuals. A decision maker with only one alternative is labelled a "captive" because she has no choice. It is desirable for the alternatives in the choice set to be mutually exclusive (i.e. only one alternative may be chosen by an individual). Where they are not mutually exclusive the researcher should redefine them to be mutually exclusive or set a primary alternative.

Alternatives are characterised by **attributes** which take different values for different alternative and for different individuals. Examples of attributes of transport alternative are cost of travel, speed of travel, and comfort. These definitely vary across different modes of transport, but also vary between individuals, since individuals have different travel routes and perceive comfort in different ways.

The way in which the alternatives are evaluated by the individual is called the ***decision rule*** or the decision protocol. There are a variety of decision rules: following Camerer (1995), we can classify them into: consumer utility maximisation; heuristics; mental shortcuts or "rules of thumb" that simplify thinking and decision making. There other non-conventional decision rules such as ***dominance of alternative***. Sometimes people follow a ***satisfactional*** rule, whereby they choose alternatives which are good enough, rather than searching through a large number of possible alternatives for the one that is truly the best. It is also possible to have a combination of these decision rules. There are instances in which the individual choice is constrained when the choice set contains too many or complex alternatives. This sometimes creates a negative attitude towards the "freedom to choose" (Sen, 1988; Baharad and Nitzan, 2000).

From consumer theory, we know that consumers select alternatives (from the individual-specific feasible choice set) that yield maximum utility. An important question is how do we measure utility? If we could quantify utility for goods, the problem would be much simpler, and the answers to our questions could be obtained by conducting regressions with utility as the (continuous) dependent variable. However, it is well-known that utility cannot be measured directly. The problem that we have is how best to use the available information on the discrete choices made by individuals to *infer* features of the utility function.

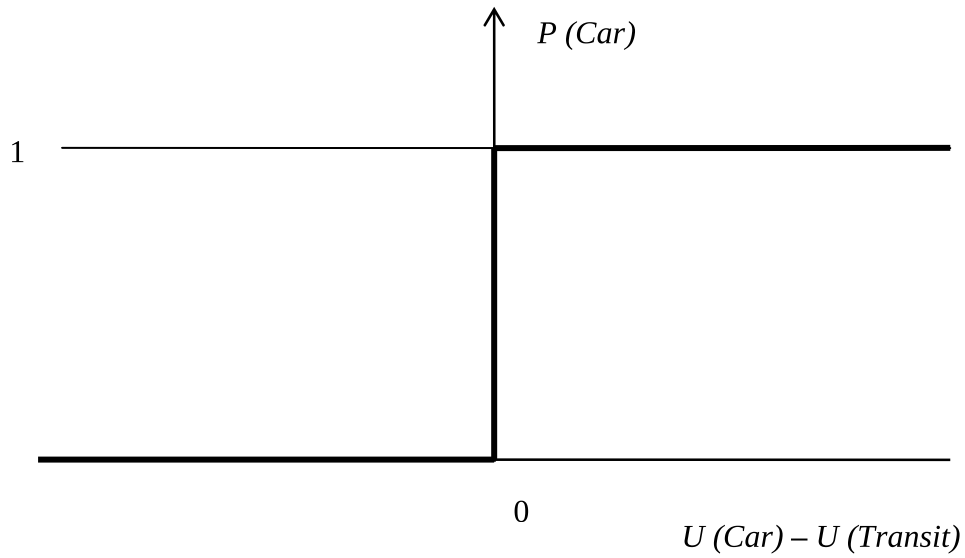

Figure 3.1: Deterministic Binary Choice Utility Step-function

In the case of only two alternatives (car and transit, for example), we would say that we are in a “binary choice” setting, and what matters is the sign of the difference between the utilities $U(car)$ and $U(transit)$. If this difference is positive then we would predict that car will be chosen with certainty. If it is negative then we would predict that transit will be chosen with certainty. This sort of analysis is called *deterministic* binary choice. The decision rule is represented by a step function, as shown in Figure 3.1.

3.3 Probabilistic Choice

Even if the individuals are rational and maximise utility in the way that we hypothesise, the decision process can never be treated as completely deterministic. We may specify the model incorrectly, we are making simplifying assumptions about functional forms, and we do not observe/measure all the attributes. Because of these problems, and also the fact that it is natural for human decision makers to make errors in decision making, we always use a probabilistic choice model.

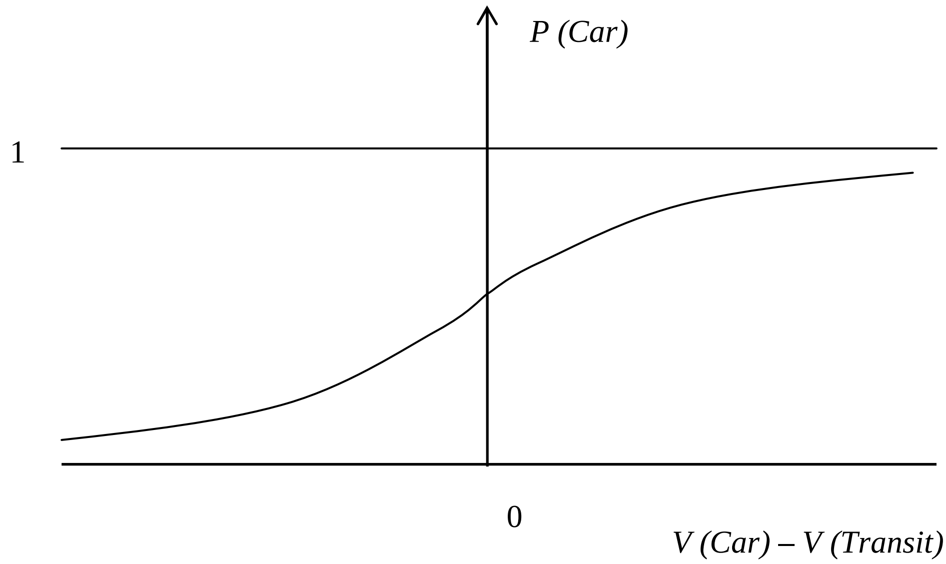

Figure 3.2: Probabilistic Choice Utility Function

In the probabilistic model that is known as the Random Utility Model (RUM), utility consists of two components, the deterministic and the random variables. The deterministic component (*V*) is made up of attributes of the alternatives, interactions with socioeconomic characteristics, and parameters. The random component (ε), can be thought of as the part that results from errors by the decision-maker, from incorrect measurement of observed attributes, and also from unobserved attributes such as personal tastes. Instead of the step function that we see in *deterministic* binary choice, in the probabilistic model we have an S-shaped decision curve, as shown in Figure 3.2. This function now represents the *probability* of choosing one of the two alternatives.

Along the lines of Ben-Akiva and Lerman (1985), the 2-alternative RUM may be written as:

$$U_{is} = V_{is} + \varepsilon_{is} \quad s = 1, 2 \tag{3.1}$$

V_{is} is the systematic utility expressed as a function of observed variables, consisting of attributes of alternative *s*, socioeconomic characteristics of individual *i*, and unknown parameters.

For the current expositional purposes, let us assume that only attributes of the two alternatives are the only relevant determinants of the choice, so we have:

$$U_{is} = z_{is}'\alpha + \varepsilon_{is} \quad s = 1,2 \tag{3.2}$$

Where z_{is} is a vector of attributes of alternative s, and α is a corresponding vector of parameters.

Although there are two alternatives, only one choice variable is required. Let us label this choice variable as y, and define it as:

$$y_i = 1 \; if \; individual \, i \, chooses \, alternative \, 2$$
$$y_i = 0 \; if \; individual \, i \, chooses \, alternative \, 1$$

Combining equations (3.1) and (3.2), the probability of individual i choosing alternative 2 is:

$$\begin{aligned} P(y_i = 1) = P(U_{i2} > U_{i1}) &= P(z_{i2}'\alpha + \varepsilon_{i2} > z_{i1}'\alpha + \varepsilon_{i1}) \\ &= P\left[\varepsilon_{i1} - \varepsilon_{i2} < (z_{i2} - z_{i1})'\alpha\right] \end{aligned} \tag{3.3}$$

That is, the probability of choosing the second alternative depends only on the <u>differences</u> in attributes between the two alternatives.

The functional form of the probability depends on assumptions made about the distribution of the error terms ε_{is}. If the error terms are normally distributed, then the difference $\varepsilon_{i1} - \varepsilon_{i2}$ is also normally distributed, and we arrive at the <u>simple probit model</u>, or as suggested by Maddala (1983), the “Normit” Model:

$$P(y_i = 1) = \Phi\left[(z_{i2} - z_{i1})'\alpha\right] \tag{3.4}$$

Where $\Phi(.)$ is the standard normal c.d.f.

If the error terms follow a type I extreme value (Gumbel) distribution, then the difference $\varepsilon_{i1} - \varepsilon_{i2}$ is follows a logistic distribution, and we arrive at the <u>simple logit model</u>:

$$P\left(y_i=1\right)=\frac{\exp\left[\left(z_{i2}-z_{i1}\right)'\alpha\right]}{1+\exp\left[\left(z_{i2}-z_{i1}\right)'\alpha\right]} \qquad (3.5)$$

Note that both the probit and the logit probability formulae give rise to S-shaped curves as shown in Figure 3.2.

From a practical point of view, there are no great differences between probit and logit. As Greene (2003, p.667) writes, "in most applications, the choice between these two seems not to make much difference". However, as we shall see later in the Chapter, when the problem is generalised to deal with more than two alternatives, very important differences emerge between the two model-types.

As mentioned, whichever of these two models is chosen, the choice probability is a function of the differences in attributes between the two alternatives. So, in the travel mode example, the probability of choosing car depends on: the difference in cost between car and transit; the difference in speed between car and transit; and so on.

Finding differences in attributes clearly requires that the attributes are known for both alternatives. For example, if an individual uses the car, we not only need to know the cost of their car-journey, but we also need know the cost that they would incur if they made the same journey by transit. The later is known as a <u>counterfactual</u> variable. It is natural to expect that the counterfactual information is not known, and needs to be imputed in some way. This particular problem is encountered in our residential location choice model of Chapter 7.

3.4 Stated Preference and Revealed Preference

The data that we use in this thesis is Revealed Preference (RP) data, since it is data on actual decisions. An alternative approach is to use Stated Preference (SP) data, which are survey responses to hypothetical questions. An example of a SP question would be: if the car journey cost £1 and took 20 minutes, while the transit journey cost 40p and took 40 minutes, which would you choose? The great advantage of the SP approach is that the counterfactual problem described above is avoided; the counterfactual data is known exactly since it has been determined in the design of the choice experiment.

There is much controversy over the use of SP data. One problem is that responses are sensitive to the wording of questions. Another problem is that people have no incentive to respond truthfully when asked hypothetical questions; indeed there is often an incentive to respond falsely if respondents feel that their responses may provide support for desirable (or undesirable) policy changes.

RP data is, by definition, truthful data and this is one of the reasons why we work exclusively with RP data in this thesis.

3.5 Multiple Choice Data

Here we extend the analysis from the case of two alternatives to the case of multiple alternatives. There are many textbook treatments of the analysis of multiple choice data, including Maddala (1983), Ben-Akiva and Lerman (1985), Cramer (1991), and Cameron and Trivedi (2005). In what follows, the ideas from these textbook treatments that are most relevant to the present study are introduced in the context of a standard example.

As already mentioned, the standard example is travel mode choice by commuters. In this context, there may be five alternatives:

Bus:	$y_{i1} = 1$ if individual i chooses bus; zero otherwise
Train:	$y_{i2} = 1$ if individual i chooses train; zero otherwise.
Car:	$y_{i3} = 1$ if individual i chooses car; zero otherwise.
Bicycle:	$y_{i4} = 1$ if individual i chooses bicycle; zero otherwise.
Walk:	$y_{i5} = 1$ if individual i chooses to walk; zero otherwise.

Note that we are only concerned with alternatives that are *not ordered*. An example of ordered alternatives is:

Alternative 1:	Don’t own a car
Alternative 2:	Own 1 car

Alternative 3: Own 2 cars

Alternative 4: Own more than 2 cars

For this sort of data, the ordered probit or ordered logit models would be appropriate. As we see in Chapter 8 some researchers have applied choice models to ordered data. We simply make the point that we consider this to be an inappropriate choice of modelling strategy.

3.6.1 Explanatory Variables in Multiple Choice Models

A very important aspect of the models under consideration is the need to distinguish between two different types of explanatory variable: characteristics of the individual; and characteristics of the alternative.

Characteristics of the individual are the variables we might normally expect to appear in a microeconometric model: age, gender, income, marital status, etc. These obviously vary from one individual to the next, but they do not vary between alternatives. Also, they have different impacts on the probabilities of different alternatives. For example, older people may be more likely to use a car, but less likely to use bicycle, than younger people.

Characteristics of the alternatives include variables such as cost, time, safety, and comfort. These obviously do vary between alternatives, but they vary between individuals as well. Obviously, the time taken and the cost incurred from taking the bus depends on where the commuter lives and works. Also, the "comfort" associated with using a bicycle is very individual specific. These variables must therefore be measured as the characteristic of an alternative as perceived by a particular individual. Another important point is that the effect of one of these variables is uniquely defined; the effect does not change across alternatives. Such an effect must simply be interpreted in terms of the impact on utility of a unit-increase in the characteristic.

3.6.2 Notation

We continue to use the subscript i for individuals (of whom there are n). We use the subscripts s and t for alternatives (of which there are S).

We define a set of latent variables as follows:

y_{is}^{*} is the utility individual *i* derives from choosing alternative *s* (*s = 1,...,S*).

If the explanatory variables are characteristics of the individuals, we specify:

$$y_{is}^{*} = x_i{}'\beta_s + u_{is} \quad s = 1,\cdots,S \;\; i = 1,\cdots,n \tag{3.6}$$

x_i is a vector of the characteristics of individual *i* (age, gender, income,...), β_s is a corresponding vector of parameters, and there is a different β vector for each alternative. The first element of each β_s vector is an intercept. If the random component u_{is} is assumed to follow a type I extreme value distribution, defined by the distribution function $F(u) = \exp(-\exp(-u))$, $-\infty < u < \infty$, then the model defined in (3.6) is the Multinomial Logit Model (MNL).

If the explanatory variables are characteristics of the alternatives, we specify:

$$y_{is}^{*} = z_{is}{}'\alpha + u_{is} \quad s = 1,\cdots,S \;\; i = 1,\cdots,n \tag{3.7}$$

where z_{is} is a vector containing the characteristics of alternative *s* (e.g. cost, time, comfort,...), as perceived by individual *i*. There is no intercept in (3.7). (3.7) is known as the Conditional Logit model (CLM), made popular by McFadden (1973).

Equation (3.7) is reminiscent of Lancaster's (1966, 1991) "characteristics approach" to demand theory, in which consumers are assumed to derive utility from the characteristics contained in goods, rather than from the goods themselves. (3.7) also reminds us of the Hedonic Pricing Model (the topic of chapters 4 and 6 of this thesis). However, the application of a hedonic pricing model requires that a market price variable is available to be used as the dependent variable in (3.7). It is when no market price data is available, but only information on whether alternative *s* has been chosen, that the CLM is useful.

If some of the explanatory variables are characteristics of the individual, and the remainder are characteristics of the alternatives, then we specify:

$$y_{is}^{*} = x_i{}'\beta_s + z_{is}{}'\alpha + u_{is} \quad s = 1,\cdots,S \;\; i = 1,\cdots,n \tag{3.8}$$

(3.8) is simply a combination of (3.6) and (3.7), and is known as a mixed logit model.

Use of the name "mixed logit model" for the model defined in (3.8) is in conformity with the terminology of Cameron and Trivedi (2005). However, it should be noted that some other authors have a different idea of what "mixed logit" is. For example, Train (2003) uses the name "mixed logit" for a MNL with random parameters.

3.6.3 Counterfactual Data

It may be that data on the characteristics of the alternatives (z_{is}) are only available for the chosen alternative. For example, an individual may have reported that they travel by bus, the journey lasts for 30 minutes and the cost is £1.50. It is unlikely that the survey design would be such that they would also be required to report the time and cost that they *would* experience *if* they chose each of the other modes. Nevertheless, in order to estimate the models discussed below, the characteristics of all alternatives must be known.

What is needed here is a system for determining "counterfactual" observations. This does not need to be complete guesswork. The data itself is useful. For example, if the data on car users and bus users reveals that car journeys are, for a given journey length, two times as fast as bus journeys, we might simply divide the observed bus-time by two in order to obtain the counterfactual car time.

This method for obtaining counterfactual data is clearly quite crude. Needless to say, more sophisticated methods are available.

3.6.4 Estimation

We write (3.6), (3.7) and (3.8) in the common notation:

$$y_{is}^{*} = V_{is} + u_{is} \quad s = 1, \cdots, S \quad i = 1, \cdots, n \qquad (3.9)$$

In (3.9), V_{is} is the <u>deterministic component</u> of utility (sometimes called the Indirect Utility Function), and u_{is} is the <u>random component</u>. The form of V_{is} depends on which of the three models (3.6), (3.7) or (3.8) is being used.

We then assume that each individual chooses the alternative which yields the highest utility. Formally, the observed variable is y_{is} and:

$$y_{is} = 1 \text{ if } y_{is}^* = \max\left(y_{i1}^*, y_{i2}^*, \cdots, y_{iS}^*\right)$$

$$y_{is} = 0 \text{ otherwise.}$$

We assume that no two alternatives ever give the same level of utility.

We next need to obtain a formula for the probability that individual *i* will choose alternative *s*. Although the formula that we derive is very well known in the literature, the formula, and certain aspects of its derivation, are so central to this thesis, that we consider it appropriate to include a complete and detailed derivation here.

3.7 Derivation of Probability Formula for MNL/CLM/Mixed Logit

In accordance with (3.9) above, the utility individual *i* derives from choosing alternative s is given by:

$$y_{is}^* = V_{is} + u_{is} \quad s = 1, \cdots, S \;\; i = 1, \cdots, n \qquad (3.10)$$

where V_{is} is the deterministic component of utility and u_{is} is the random component.

The random component u_{is} is assumed to follow a type I extreme value distribution, defined by the distribution function:

$$F(u) = \exp\left(-\exp(-u)\right) \quad -\infty < u < \infty \qquad (3.11)$$

or the density function:

$$f(u) = F'(u) = \exp\left(-u - \exp(-u)\right) \quad -\infty < u < \infty \qquad (3.12)$$

An important assumption made here is independence between alternatives: u_{is} and u_{it} are distributed independently for $t \neq s$.

Individual *i* chooses alternative *s* if the utility she derives from alternative *s* is higher than the utility she derives from any other alternative. Ties are not allowed. If *i* chooses *s*, we say y_{is}=1. We therefore have:

$$y_{is} = 1 \; if \;\; y_{is}^* > y_{it}^* \;\; \forall t \neq s \,.. \qquad (3.13)$$

or $y_{is} = 1 \; if \; V_{is} + u_{is} > V_{it} + u_{it} \quad \forall t \neq s$ (3.14)

or $y_{is} = 1 \; if \; u_{it} < u_{is} + V_{is} - V_{it} \quad \forall t \neq s$ (3.15)

Let us henceforth suppress the *i* subscript, so that we have:

$$y_s = 1 \; if \; u_t < u_s + V_s - V_t \quad \forall t \neq s \tag{3.16}$$

and let us proceed to find the probability that $y_s = 1$. Let us first condition on the value of u_s. Using (3.16), and exploiting the independence of the u's:

$$\begin{aligned}
P(y_s = 1 | u_s) &= \prod_{t \neq s} P(u_t < u_s + V_s - V_t | u_s) \\
&= \prod_{t \neq s} F(u_s + V_s - V_t) \\
&= \prod_{t \neq s} \exp\left[-\exp(-u_s - V_s + V_t)\right] \\
&= \exp\left[-\sum_{t \neq s} \exp(-u_s - V_s + V_t)\right] \\
&= \exp\left[-\exp(-u_s - V_s)\sum_{t \neq s} \exp(V_t)\right] \\
&= \exp\left[-\exp(-u_s - V_s)\left\{\sum_t (\exp(V_t)) - \exp(V_s)\right\}\right] \\
&= \exp\left[-\exp(-u_s - V_s)\exp(V_s)\left(\frac{\sum_t \exp(V_t)}{\exp(V_s)} - 1\right)\right] \\
&= \exp\left[-\exp(-u_s)\left(\frac{\sum_t \exp(V_t)}{\exp(V_s)} - 1\right)\right] \\
&= \exp\left[-\exp(-u_s)\left(\lambda_s^{-1} - 1\right)\right]
\end{aligned} \tag{3.17}$$

where $\lambda_s = \dfrac{\exp(V_s)}{\sum_t \exp(V_t)}$.

To obtain the marginal probability from the conditional probability, we use:

$$P(y_s = 1) = \int_{-\infty}^{\infty} P(y_s = 1 | u_s) f(u_s) du_s \tag{3.18}$$

Placing (3.17) and (3.12) into (3.18), we obtain:

$$
\begin{aligned}
P(y_s = 1) &= \int_{-\infty}^{\infty} \exp\left[-\exp(-u_s)\left(\lambda_s^{-1} - 1\right)\right] \exp\left[-u_s - \exp(-u_s)\right] du_s \\
&= \int_{-\infty}^{\infty} \exp(-u) \exp\left[-\exp(-u)\left(\lambda_s^{-1} - 1\right) - \exp(-u)\right] du \\
&= \int_{-\infty}^{\infty} \exp(-u) \exp\left[-\exp(-u)\lambda_s^{-1}\right] du \\
&= \lambda_s \int_{-\infty}^{\infty} \lambda_s^{-1} \exp(-u) \exp\left[-\lambda_s^{-1} \exp(-u)\right] du \qquad (3.19) \\
&= \lambda_s \exp\left[-\lambda_s^{-1} \exp(-u)\right]_{-\infty}^{\infty} \\
&= \lambda_s (1 - 0) \\
&= \lambda_s \\
&= \frac{\exp(V_s)}{\sum_t \exp(V_t)}
\end{aligned}
$$

which is the well-known formula for choice probabilities in MNL./CLM/Mixed logit.

The virtue of (3.19) is its simplicity. In particular, no integration is required in the evaluation of the probability. However, a potential drawback is that it is based on the (under some circumstances) unreasonable assumption of independence between alternatives: u_{is} and u_{it} are independently distributed for all $s \neq t$. This means that, for example, if we know that an individual particularly likes travelling by car ($u_{i,car}$ is high and positive), this knowledge does not alter the expectation of ($u_{i,bike}$), the random component of the utility from using bicycle. This independence assumption is closely related to the widely-discussed assumption of Independence of Irrelevant Alternatives (IIA) which is given attention later in this Chapter.

3.8 The Likelihood Function

Now that we have established the probability formula (3.19), we may construct the likelihood function for a sample of size n:

$$
L = \prod_{i=1}^{n} \left[\frac{\exp(V_{i1})}{\sum_t \exp(V_{it})}\right]^{y_{i1}} \left[\frac{\exp(V_{i2})}{\sum_t \exp(V_{it})}\right]^{y_{i2}} \cdots \left[\frac{\exp(V_{iS})}{\sum_t \exp(V_{it})}\right]^{y_{iS}} \qquad (3.20)
$$

This may be written more compactly, as:

$$L=\prod_{i=1}^{n}\left[\frac{\left[\exp\left(V_{i1}\right)\right]^{y_{i1}}\left[\exp\left(V_{i2}\right)\right]^{y_{i2}}\cdots\left[\exp\left(V_{iS}\right)\right]^{y_{iS}}}{\sum_{t}\exp\left(V_{it}\right)}\right] \qquad (3.21)$$

So the log-likelihood is:

$$\begin{aligned} LogL &= \sum_{i=1}^{n}\left[y_{i1}V_{i1}+y_{i2}V_{i2}+\cdots+y_{iS}V_{iS}-\log\left(\sum_{t}\exp\left(V_{it}\right)\right)\right] \\ &= \sum_{i=1}^{n}\left[\sum_{t}y_{it}V_{it}-\log\left(\sum_{t}\exp\left(V_{it}\right)\right)\right] \end{aligned} \qquad (3.22)$$

(3.22) is the log-likelihood function for MNL, CLM or mixed logit. It is maximised with respect to the parameter vector(s) β and/or α to obtain MLE's of these parameters.

3.9 Some Identification Issues

Two important results relating to identification need to be stated explicitly.

Result 1: The intercept term is <u>not identified</u> in the Conditional Logit Model (CLM).

Proof of Result 1

The CLM may now be written as:

$$P\left(y_{is}=1\right)=\frac{\exp\left(z_{is}'\alpha\right)}{\sum_{t}\exp\left(z_{it}'\alpha\right)} \qquad (3.23)$$

Let us consider what happens when we add an intercept, α_0:

$$\begin{aligned} P\left(y_{is}=1\right) &= \frac{\exp\left(\alpha_0+z_{is}'\alpha\right)}{\sum_{t}\exp\left(\alpha_0+z_{it}'\alpha\right)}=\frac{\exp\left(\alpha_0\right)\exp\left(z_{is}'\alpha\right)}{\exp\left(\alpha_0\right)\sum_{t}\exp\left(z_{it}'\alpha\right)} \\ &= \frac{\exp\left(z_{is}'\alpha\right)}{\sum_{t}\exp\left(z_{it}'\alpha\right)} \end{aligned} \qquad (3.24)$$

which is the same as (3.23). This means that any change in the value of the intercept, α_0, does not have any effect on $P(y_{is} = 1)$. This means that observations on behaviour (the y_{is}'s) cannot be used to estimate the value of this parameter. It is not identified.

Q.E.D.

Result 1 is not surprising if it is remembered that the terms in brackets in the RHS of (3.23) and (3.24) are (indirect) utility functions. It is well known from basic consumer theory that adding a constant term to a utility function cannot have any effect on implied behaviour.

Result 2: In the Multinomial Logit Model (MNL), one of the S β_s vectors must be normalised, i.e. its value must be set *a-priori*.

Proof of Result 2

The MNL model is defined by:

$$P(y_{is}=1)=\frac{\exp(x_i'\beta_s)}{\sum_t \exp(x_i'\beta_t)} \tag{3.25}$$

Let us consider what happens when we add a constant vector γ, to each of the β_s vectors:

$$\begin{aligned} P(y_{is}=1) &= \frac{\exp(x_i'(\beta_s+\gamma))}{\sum_t \exp(x_i'(\beta_t+\gamma))} = \frac{\exp(x_i'\beta_s + x_i'\gamma)}{\sum_t \exp(x_i'\beta_t + x_i'\gamma)} \\ &= \frac{\exp(x_i'\gamma)\exp(x_i'\beta_s)}{\exp(x_i'\gamma)\sum_t \exp(x_i'\beta_t)} = \frac{\exp(x_i'\beta_s)}{\sum_t \exp(x_i'\beta_t)} \end{aligned} \tag{3.26}$$

which is the same as (3.25). This means that changing all the β_s vectors by the same amount γ doesn't change $P(y_{is} = 1)$. So clearly the vector γ is not identified. It follows that the β_s vectors are not separately identifiable. Only the differences between them, e.g. $\beta_3 - \beta_2$, are identified. Q.E.D.

To deal with this problem, one of the β_s vectors needs to be normalised, in order for the remaining S-1 to be identified. The convention is to normalise the first vector to zero, that is:

$$\beta_1 = 0 \tag{3.27}$$

With this normalisation, the choice probabilities in the MNL become:

$$P(y_{i1}=1)=\frac{1}{1+\sum_{t=2}^{S}\exp(x_i'\beta_t)}$$
$$P(y_{is}=1)=\frac{\exp(x_i'\beta_s)}{1+\sum_{t=2}^{S}\exp(x_i'\beta_t)} \quad for\ s=2,\cdots,S \tag{3.28}$$

A consequence of the normalisation is that the first alternative becomes the base alternative, and the interpretations of the estimated β_s's are made in comparison to the base alternative.

For example, if the base case is "bus", and "car" has a positive age-coefficient, this just means that age has a greater positive effect on the probability of car use, than it does on the probability of bus use. If "bicycle" has a negative age coefficient, this means that age has a greater negative effect on the probability of bicycle use, than it does on the probability of bus use.

3.10 Multinomial Logit Model with only two alternatives

Consider the MNL with only S=2 alternatives. The probabilities are:

$$P(y_{i1}=1)=\frac{\exp(x_i'\beta_1)}{\exp(x_i'\beta_1)+\exp(x_i'\beta_2)}$$
$$P(y_{i2}=1)=\frac{\exp(x_i'\beta_2)}{\exp(x_i'\beta_1)+\exp(x_i'\beta_2)} \tag{3.29}$$

And, when we normalise β_1 to zero, these become:

$$P(y_{i1}=1)=\frac{1}{1+\exp(x_i'\beta_2)}$$
$$P(y_{i2}=1)=\frac{\exp(x_i'\beta_2)}{1+\exp(x_i'\beta_2)} \tag{3.30}$$

If we now define a binary variable y_i which is 1 if the second alternative is chosen, and zero if the first alternative is chosen, we have:

$$P(y_i=1)=\frac{\exp(x_i'\beta_2)}{1+\exp(x_i'\beta_2)} \tag{3.31}$$

which is, of course, the definition of the simple logit model (3.5) discussed above in Section 3.3. The point here is that when there are only two alternatives, the MNL model simplifies to the simple logit model which is used for binary data. MNL can thus be seen as a generalisation of simple logit.

3.11.1 The "Independence of Irrelevant Alternatives" property (IIA)

The Independence of Irrelevant Alternatives (IIA) property is a feature of the MNL/CLM framework that may present severe problems in certain applications. Many authors have given attention to the problem, including Amemiya (1985) and Train (2003).

Consider the probabilities of two of the alternatives, *s* and *t*:

$$\begin{aligned} P\left(y_{is}=1\right) &= \frac{\exp\left(V_{is}\right)}{\sum_r \exp\left(V_{ir}\right)} \\ P\left(y_{it}=1\right) &= \frac{\exp\left(V_{it}\right)}{\sum_r \exp\left(V_{ir}\right)} \end{aligned} \tag{3.32}$$

and consider the ratio of these two probabilities:

$$\frac{P\left(y_{is}=1\right)}{P\left(y_{it}=1\right)} = \frac{\exp\left(V_{is}\right)}{\exp\left(V_{it}\right)} = \exp\left(V_{is}-V_{it}\right) \tag{3.33}$$

(3.33) is sometimes called the "odds ratio" of alternatives s and t. The "log-odds ratio" is the log of the odds ratio, which is V_{is} - V_{it}.

In the multinomial logit model, the log-odds ratio is:

$$V_{is}-V_{it} = x_i'\beta_s - x_i'\beta_t = x_i'\left(\beta_s-\beta_t\right) \tag{3.34}$$

and in the conditional logit model, the odds ratio is:

$$V_{is}-V_{it} = z_{is}'\alpha - z_{it}'\alpha = \left(z_{is}-z_{it}\right)\alpha \tag{3.35}$$

The important thing about the odds ratio (3.33) is that it does not involve the parameters or characteristics of any alternatives other than *t* and *s*. That is, the ratio of the probabilities of *t* and *s* is independent of all of the other alternatives. This is the IIA property.

The IIA property is a consequence of two factors: the assumption of statistical independence of the error terms between different alternatives (which was central to the derivation of 3.17 above); and the choice of the type I extreme value distribution (3.11) for the stochastic components of the model. Note that choice of another distribution, such as the normal, would lead to a violation of IIA.

An extreme example serves to illustrate why IIA is a problem. This example is already well known in the literature.

3.11.2 The "Red Bus/Blue Bus" Problem

The "red bus/blue bus" problem was introduced by McFadden (1973). The initial situation is that two alternatives are available, car and red bus. And the probabilities of these alternatives being chosen are:

P(car) = 0.5

P(red bus) = 0.5

Now a new mode of transport is introduced: blue bus. In reality, the commuter is indifferent between the two types of bus, so we would expect the probabilities to become:

P(car) = 0.5

P(red bus) = 0.25

P(blue bus) = 0.25

However, in the MNL/CLM framework, the introduction of the new mode cannot change the ratio of the probabilities of the existing modes. This ratio is one. So, under MNL/CLM, the probabilities become:

P(car) = 0.33

P(red bus) = 0.33

P(blue bus) = 0.33.

These numbers are clearly not sensible; we do not expect P(car) to fall when Blue bus is introduced. The problem here is that blue bus and red bus are perfect substitutes. Perhaps this example is too extreme. However, similar problems arise when two of the alternatives are close substitutes.

3.11.3 Testing IIA

One method to test the IIA assumption is using the Hausman (1978) testing procedure. The model is estimated twice, first on full set of alternatives, then on a specific subset of alternatives. If IIA holds, the two sets of estimates should not be significantly different (McFadden, 1987). Hausman and McFadden (1984) suggest that, if a subset of the choice set is really irrelevant, omitting it from the model should not change the parameter estimates systematically. Hence the test has come to be labelled the "Hausman-McFadden Test".

Small and Hsiao (1985) show that the Hausman-McFadden test is asymptotically biased and propose another testing strategy. The Small-Hsiao IIA test is a likelihood ratio test which divides the data set randomly, into two subsets. Like in the Hausman test, the model is estimated twice, first the unrestricted/full model is estimated for one of the subsets, and then a restricted model in which one of the alternatives is dropped is estimated on the second subset. The two results are compared by means of a test statistic which is asymptotically distributed as Chi-squared, with degrees of freedom equal to the number of explanatory variables plus one.

There is some *Monte Carlo* evidence on the comparative performance of the two testing strategies (see, for example, Fry and Harris, 1996). In Chapter 8 of this thesis, the performance of these two tests is considered in the context of a meta-analysis regression. That is, all of the IIA tests we can find in the literature are collected together, and the determinants of the test results (including testing method) are identified. The results we report in Chapter 8 are in broad agreement with previous results from *Monte Carlo* studies.

3.12 The Nested Logit Model (NLM) and the Multinomial Probit Model (MNP)

Generalisations to MNL/CLM are available which relax the IIA restriction. One such generalisation is the Nested Logit Model (NLM), in which alternatives are first allocated into groups according to degree of substitutability (or similarity)(see Cramer, 1991). For example, the grouping of travel mode choices might be:

group 1: car, taxi
group 2: train, underground
group 3: bus
group 4: bicycle, walk, rollerblade

For further examples of the nesting process, see Fox (2006), Heiss (2002), Christiadi and Cushing (2007).

Under the nested logit model, the commuter is first assumed to choose between the four groups. Then, after this choice has been made, the choice is made between alternatives within the chosen group.

The NLM has additional parameters, representing the degree of similarity within each group. McFadden (1978) and Maddala (1983) refer to these parameters as "dissimilarity parameters", while Cameron and Trivedi (2005) use the term "scale parameters".

The NLM is an ideal approach when there is a clear nesting structure. However, according to Cameron and Trivedi (2005) and Greene (2003), there is usually no obvious structure. Given this, NLM seems hard to justify.

One model that gives complete flexibility in terms of the nesting structure is the multinomial probit model (MNP). This is the model that arises when the error terms associated with each alternative, $u_{i1}, u_{i2}, \cdots, u_{iS}$, are assumed to be joint normally distributed. This model permits a very rich correlation structure. However, the significant drawback is that numerical or simulation methods are required to accommodate integrals of dimension S-1 (see Train, 2003).

Until recently, for reason of computational complexity, the MNP approach is hardly used in empirical studies. However, in Chapter 8 we find that use of the MNP model is one of the key determinants of whether a paper is published. Developments in econometrics software and increase in computer processing capacity and speed have made the estimation of MNP feasible. We apply both NLM and MNP in Chapter 7, our discrete choice analysis chapter.

3.13 Extracting WTP from Conditional Logit Model (CLM) Results

If the two characteristics are cost and time, then the utility function (3.2) underlying the CLM becomes:

$$y_{is}^{*} = \alpha_1 cost_{is} + \alpha_2 time_{is} + u_{is} \tag{3.36}$$

where $cost_{is}$ is the cost (in pence) to individual i of using alternative s, and $time_{is}$ is the time taken (in minutes) by individual i if alternative s is chosen. Maximum Likelihood Estimates of the parameters α_1 and α_2 may be estimated using the clogit command in STATA, or using the logit command in TSP (see Section 3.14 below on Software Issues).

Since we expect cost and time to be characteristics which reduce utility, we would expect both α_1 and α_2 in (3.36) to take negative values. Since (3.36) is linear in cost and time, the implied indifference curves are parallel straight lines, each with (negative) slope $-\alpha_2/\alpha_1$. See Figure 3.3.

The absolute value of this slope has a very important economic interpretation: it is the number of pence a commuter is willing to give up for a one-minute saving of commuter time, i.e. it is the Willingness-to-Pay (WTP) for a one minute saving (McFadden (2000); Sonnier et al (2007). Therefore, it can be interpreted as a measure of the <u>value of time</u>.

If $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are respectively the estimates of α_1 and α_2, then the ratio $\hat{\alpha}_2 / \hat{\alpha}_1$ is an estimate of the value of time. The standard error of this estimate may be obtained using the delta method (Greene, 2003).

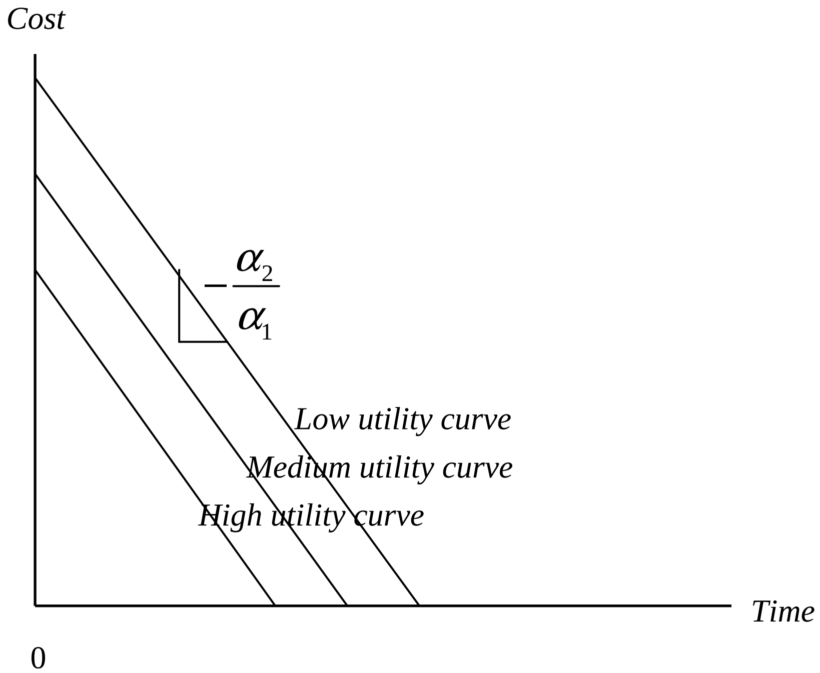

Figure 3.3: Indifference Curves in time-cost space

We do not expect the value of time to be the same for all individuals. For example, we might expect the value of time to rise with income. After all, an extra minute spent commuting is a minute that could have been spent earning money. Looking at it this way, the value of time is very closely related to income.

To allow value of time to depend on income, we would introduce an interaction variable as the product of income and time:

$$y_{is}^{*} = \alpha_1 cost_{is} + \alpha_2 time_{is} + \alpha_3 \left(income_i * time_{is} \right) + u_{is} \qquad (3.37)$$

By rearranging (3.37), we find that the implied value of time is:

$$\frac{\alpha_2 + \alpha_3 income}{\alpha_1} \qquad (3.38)$$

A negative value of the parameter α_3 would confirm that the value of time rises with income.

3.14 Software Issues

In order to estimate the MNL model in STATA, the data needs to consist of one row per observation, with an integer valued variable (mode) representing the choice made by each individual. So, the first few rows of the data set might look like this:

indiv	***mode***	***age***	***income***
1	2 (bus)	30	220
2	1 (car)	55	420
3	3 (bicycle)	24	350
:	:	:	:

Then the mlogit command would be used:

```
mlogit mode age income
```

To estimate the CLM, the data needs to be in "long" form (see below). This means that for each individual, there need to be *S* rows, one for each choice. There also needs to be a variable "choice", taking the value 1 if this choice was made, zero otherwise. So, if *S*=3, the first few rows of the data might look like that shown below.

The command that is required to estimate the CLM is:

```
clogit choice cost time, group(indiv)
```

The mixed logit model (in the sense defined in this chapter, as containing a mixture of the two types of explanatory variable) can be estimated in TSP (Hall and Cummins, 2005) using the command:

```
LOGIT (NCHOICE=3,COND) Y COST TIME | C AGE INCOME
```

indiv	*mode*	*choice*	*cost*	*time*	*age*	*income*
1	1 (car)	0	75	10	30	220
1	2 (bus)	1	50	20	30	220
1	3 (bicycle)	0	25	30	30	220
2	1 (car)	1	220	40	55	420
2	2 (bus)	0	180	55	55	420
2	3 (bicycle)	0	120	90	55	420
3	1 (car)	0	110	25	24	350
3	2 (bus)	0	90	45	24	350
3	3 (bicycle)	1	60	60	24	350
:	:	:	:	:	:	:
:	:	:	:	:	:	:

Same result can be obtained in STATA using the alternative specific conditional logit with individual attributes added to the model using the command:

asclogit choice cost time, case(indiv) alternatives(mode) casevars(age income)

We also estimate NLM and MNP models ("mixed" probit and alternative specific probit models) as a means of relaxing the IIA assumption, and thereby avoiding the cost it imposes on choice modelling (and estimations). NLM and alternative specific multinomial (conditional and mixed) probit model(s) can be estimated using Stata 11 model using the respective commands:

asmprobit chosenarea rent hourswater hourselectricity, case(id) alternatives(area)

asmprobit chosenarea rent hourswater hourselectricity, case(id) alternatives(area) casevars(age income yearsofedu)

nlogitgen nlo = area (city: 1, lowden: 2, pollution: 3|4|5, other: 6) nlogit chosen rent hourswater hourselectricity || nlo: yearsofedu income age, base(city) || area:, noconstant case(id)

Chapter Four: Theory of Hedonic Pricing

4.1 Introduction

In this chapter we review the literature and some of the conceptual issues relating to the estimation of hedonic pricing models. The first question to ask would be why are we interested in this type of model? Because it provides a direct means of meeting the main objective of this study: to estimate WTP of housing amenities.

4.2 Model for heterogeneous good

We start with a conceptual definition of heterogeneous/differentiated good. Heterogeneous goods are products whose characteristics differ to create a distinct product variety even though they belong to the same product 'family' and are sold in one market. Standard examples, provided by Taylor (2003) are cars, computers and houses. The variation in product attributes gives rise to variation in product prices. From the market transactions for these varieties we are able to estimate the willingness to pay for attributes and related (market and non-market) goods. Another important use of the hedonic pricing method is in the construction of a quality adjusted price index.

In a market for a heterogeneous good, the explicit market with observed prices and resulting transactions is for a bundle of properties. The explicit market therefore contains several implicit markets for individual bundles. One way to analyse the implicit market is to regard demand for goods not for themselves but for the attributes they contain. In this sense households purchase goods and use them as inputs transforming them into utility. This approach which is due to Lancaster (1966) places emphasis on household's production/transformation of, and demand for attributes.

The second interpretation is that goods are traded in a single market after they are carefully packaged, but they are heterogeneous and are approximated by a single price. That price lies within a range of prices that depend on the types and quantities of attributes the good possesses. Although conventional economic theory could not be used to analyse this market, it is possible to employ hedonic pricing method which assumes that heterogeneous goods are composed of aggregates of homogenous parts, and while aggregates may not have single price, the component attributes do have a common price structure.

In general, there are two major motivations for the estimation of hedonic price functions; the first is related to quality adjusted price indices; and second, consumer demand/willingness to pay for attributes of heterogeneous goods (Sheppard, 1999)

The main objective of this chapter is to provide the theoretical foundation of hedonic pricing which would be used to estimate, from rent data, the willingness to pay/avoid, air quality, proximity to place of work, security/law and order, schools, and basic public utilities in the study area. The marginal price of these housing characteristics could be treated as analogous to consumer's willingness to pay.

Rosen (1974) provides the basis for the modern theory of hedonic pricing. Unlike previous hedonic pricing studies, his hedonic pricing model for analysing a market for composite differentiated goods is based on utility theory. The theory recognises that housing comprises of various characteristics which are not directly traded but that the implicit marginal price of the constituent characteristics can be derived by hedonic regression. In consuming housing goods, a rational consumer is expected to maximise her utility by selecting a given bundle of characteristics (subject to budget constraints), which includes prices for these characteristics; other (non-housing) goods; and savings/investment.

Consumer's willingness to pay will depend on her income and taste which determines her preference for a given combination of characteristics. It is expected that, the market, given current state of technology, will generate various combinations of house characteristics. The price-characteristic relationship is identified through the "exchange" between buyers and sellers. The transaction resulting from supply and demand interactions could be used to generate data on prices and characteristics. From this we can generate the contribution to price or the marginal value of a characteristic, which is the partial derivative of the price equation with respect to a particular characteristic (Palmquist, 1984), (Taylor, 2003).

In addition to producer generated structural characteristics; there are also spatial characteristics which are due to external factors, outside the producer's control, and those generated either by providence or public policy.

$$P_j = p(Z_i) \qquad (4.1)$$

Equation 4.1 is the hedonic price function. The price of a house P is a function of Z_i, a vector of price. Price of structural, spatial, and neighbourhood characteristics. Differentiating P with respect to Z_i would yield the implicit price of a constituent attribute/characteristic. Hedonic pricing model therefore provides a link between a differentiated product and constituent characteristics through its price.

$$P_j = \beta_0 + \beta_1 Z_1 + ... + \beta_n Z_n + \varepsilon \qquad (4.2)$$

When the vector of house price is regressed against the vector of house characteristics, the coefficients, also called the hedonic weights, β_i, (the part of a product's overall price attributable to a given characteristic) are usually interpreted as the price of the corresponding characteristic (Day, 2003), (Hulten, 2002).

We first summarise the algebraic foundation of the hedonic pricing method using notations from Sheppard (1999).

Consumers derive utility from the consumption of a heterogeneous commodity (housing services) that contain a vector of attributes Z and a vector of composite (non-housing) goods Y. The households utility function is given by equations 4.3 and 4.4.

$$u = u(Z, Y, \alpha) \qquad (4.3)$$

Where α is the vector of parameters that characterise consumer preferences which could be observed or unobserved. Households are characterised by their income M and the parameter vector α with distribution over possible values given the joint probability $f(\alpha, M)$. From this the household's willingness to pay for a hetrogenous goods can be obtained as a function of the embodied attributes. The household's *bid rent* function $\beta(Z, M, u, \alpha)$ is defined as:

$$u = u(Z, M - \beta, \alpha) \qquad (4.4)$$

The term bid rent is due to Rosen, (1974) (Day, 2001). Differentiating the *bid rent* function with respect to attributes gives the rate at which the household would be willing to change expenditure for an increase in house attribute i.

Following Day (2001), these simplifying assumptions are as follows:

- All consumers perception of the amount of characteristics embodied in a product are identical, though consumers may differ in their subjective valuation of alternative packages.
- The set of properties in the market is fixed – no new houses are built in the short run but characteristics of existing houses could change. It is possible to relax this assumption in a dynamic setting like measuring welfare change.
- Houses are produced and supplied by landlords. Homeowners are treated as landlords that rent from themselves. There is no provision for second hand market and the possibility of resale of property does not exist.
- Each individual temporarily purchase (rent) one property. The location choice decision of landlords who own more than one property is independent of their supply decision.
- Each individual consumer is a price taker, they make decision on where to live but could not affect price.

The overall household optimisation decision involves the choice of a house with attributes Z and consumption of composite goods Y:

$$\max \; u(Z, Y, \alpha) \qquad \textit{subject to } M \geq P(Z) + Y \qquad (4.5)$$

The first order condition require that:

$$\frac{u_i}{u_Y} = P_i \;\; \forall_i \qquad (4.6)$$

where:

$$u_i = \frac{\partial u}{\partial Z_i} \qquad \textit{and} \qquad P_i = \frac{\partial P}{\partial Z_i} \qquad (4.7)$$

The derivative P_i is referred to as hedonic price of attribute i and function $P(Z)$ as hedonic price function.

Solving these equations would yield the optimal choice of housing attributes which is the equality of the slope of the bid rent and the hedonic price for each attribute.

$$\frac{\partial\beta}{\partial Z_i} = \frac{u_i}{u_Y} = P_i \qquad (4.8)$$

Sheppard (1999) concludes that, this is part of the justification for the hedonic approach because it indicates that if we can observe/estimate the hedonic price of an attribute and the choice made by household, "then under the assumption of optimising behaviour, the observation provides local information about consumers preferences or willingness to pay for attributes in the neighbourhood of the observed choice" (Sheppard, 1999 pp. 1601).

On the supply side, a producer of a heterogeneous good is characterised by the following optimisation problem. A cost function $C(Z,N,\gamma)$, that depends on the amount of attribute Z of the heterogeneous good (house) produced, N the number of housing units produced and γ a vector that capture production technology and characterises each producer. We assume a market with multiple producers described by the probability density $g(\gamma)$.

The profit function is:

$$\pi = P(Z)\cdot N - C(Z,N,\gamma) \qquad (4.9)$$

The multiple producers each assumes the price function as given and attempt to solve the following optimisation problem:

$$\max \; P(Z)\cdot N - C(Z,N,\gamma) \qquad (4.10)$$

Solving for this problem, the first order condition require that each producer equates the marginal cost of each attribute to its hedonic price and builds housing units until the marginal cost of building another unit is equal to the value of the house.

$$P_i = C_i \ \forall_i \qquad \text{and} \qquad P(Z) = C_N \tag{4.11}$$

4.3 History of the Hedonic Pricing model

There are two interpretations of the history of hedonic price modelling. Some scholars argue that G. C. Haas was the first in 1922, to estimate hedonic price model for farmland using distance, from and, the size of a city as characteristics. Others argue that A. T. Court study of the American automobile industry in 1939 was oldest published hedonic price analysis. Even if A. T. Court was aware of G. C. Haas he did not acknowledge it. It is also possible that there were previous hedonic price studies prior to these two. However, these two have had the most significant influence in setting the stage for a widespread application of hedonic price model in the analyses of differentiated goods market, Wen et al. (2005), Colwell and Dilmore (1999), Goodman (1998).

If we recognize Haas's study as the oldest hedonic price study, we could argue that hedonic price was first applied to studies explaining the spatial productivity of land. It was argued that, spatial difference yields differential rents to land and therefore differential land values. Competition for good land pushes up its price, with potential buyers/renters willing to bid above the market rate. It is expected that this will continue until the rent differential eliminates profit and or when the rent differential is equal to the productivity differentials (Colwell and Dilmore, 1999).

But some environmental characteristics also affect the productivity of land. Factors like air, water quality and neighbourhood attributes all affect quality of land. It means therefore, included in the structure of rents and prices are some environmental factors. It was discovered that, we could extract the value of environmental characteristics from the land value. This implies that, environmental factors could affect land prices, and by using knowledge therefrom, we can predict changes in land prices when any such factors change. In addition it is possible to use this information to measure resulting welfare changes.

Another application of hedonic price method is in labour market analysis. *Hedonic wage functions* reflect the relationship between wages and job characteristics.

Different types of workers have different tastes for risk. Iso-profit curves show combinations of wages and risk that yield identical levels of profits for firms. Lower iso-profit curves mean higher profits; iso-profit curves slope upward; iso-profit curves are concave reflecting diminishing marginal returns to producing safety. Firms differ in their abilities to produce safety, just as workers differ in their tastes for safety or risk. Equilibrium sorts workers and firms such that workers with a low preference for risk are matched with firms that have less difficulty producing safety and workers with a high preference for risk are matched with firms that have difficulty in producing safety. This matching process can be observed empirically and is called the hedonic wage function (Roback, 1988).

In recent times, the most popular application of hedonic price model is in the housing market, where environment quality is one of the housing characteristics. The most extensive, although dated survey on hedonic price model application to environmental economics is provided by Cameron (1998).

It is possible to use hedonic price to extract information on the value of the environmental characteristics from the market for houses. Because environmental quality varies across space, individuals would choose their exposure to pollution through their residential location choice. This is because residential housing price may include premium for clean, accessible and quite areas and discounts for noisy, dirty and inaccessible areas. From this we could estimate the demand for and price of public goods and environmental quality in particular, from the demand and the price differentials revealed in the housing market (Grafton et al, 2004).

That is to say, in theory, by looking at the aggregate behaviour of individuals in consuming housing services, we could determine, using the price and rental value of the various houses, the values of non-market environment good (or bad) like air pollution, noise, water quality, etc. As pointed out earlier, this method is based on the variety in the housing market, different sizes, and different locational and environmental characteristics. The housing market in this context is treated like a huge supermarket offering a variety of products; only that housing is a fixed durable good, individuals can only increase the amount of some of the house characteristic by moving to another location offering more of the desired characteristics (Day, 2001).

In order that we can estimate such a model empirically we make some simplifying assumptions about the housing market to remove the possibility of double counting of multiple property owners, omission of own-property owners, and ensure that, the market is not dominated by a monopsony buyer or a cartel of buyers who could influence price by individual or collective action (Day, 2001). On the whole, Rosen (1974) argued that these assumptions represent an enormous simplification of the problem which ensures that the market does not explode.

We assume that each individual has a utility function containing a bundle of housing commodities; a vector of location specific characteristics; a vector of structural characteristics of the house such as: size, number rooms and their size, garden, age, design; and a vector of relevant neighbourhood such as: access to market, crime rate in the area, pollution, quality school, parking space. A combination of these factors, determines the individual's demand for residential housing, which includes the demand for location specific factors, physical characteristics of a house and its neighbourhood.

4.4 Functional Forms

Rosen (1974) shows why in the general case, theory cannot specify the appropriate functional form for hedonic functions, except that it is monotonically increasing in desirable characteristics. First reason is that the differentiated products are sold in separate but highly interrelated markets. Secondly, linearity is unlikely as long as there is increasing marginal cost of characteristics for suppliers and as long as it is not possible to unbundle and repackage the characteristics of the products. Repackaging or arbitrage is not possible in the short-run in property markets, long-run complete adjustment has not been found in hedonic models, and it is therefore not possible empirically to force linearity.

Palmquist (1991) showed that, although theory does not preclude linearity of the hedonic function - it is purely an empirical issue, to be determined from analysis of the data – nonlinear functional forms can be made linear by transforming the variables. The most common transformation being semi-logarithmic, inverse semi-logarithmic, log-linear and quadratic Box-Cox – a flexible form that could take the form of translog, log-linear, quadratic, linear, Leontief and semi-log. Box-Cox has the following general form:

$$P_h^{\theta} = \beta_0 + \sum_{i=1}^{m} \beta_i Z_{hi}^{\lambda} + \tfrac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \delta_{ij} Z_{hi}^{\lambda} Z_{hj}^{\lambda} + \varepsilon_h \qquad (4.12)$$

Where P is price, and Z attributes. P^{θ}, and Z^{λ} are Box-Cox transformations. The generalised Box-Cox form allows for transformation of both the dependent and independent variables. P_h^{θ} is the *h*th observation on the transformed price variable, Z_{hi}^{λ} is the *h*th observation on the *i*th transformed attribute, and there are *m* attributes in total. β and δ are coefficients from the regression. ε_h is the disturbance term. Halvorsen and Pollakowski (1981) show that for the purpose of identification the following restriction is required $(\delta_{ij} = \delta_{ji})$.

In deciding the functional form, theory is the first step. Accordingly, the equilibrium price is derived from the interaction of individuals' preferences for property and its underlying characteristics and the suppliers cost and profit functions. Because hedonic price function is an equilibrium relationship derived from the interaction of demand and supply function, the necessary condition for the functional form of the hedonic price function is that the first derivative with respect to characteristics be positive for good characteristics and negative for bad characteristics. We then rely on the simplifying assumptions on preferences and supply to derive our solutions.

The best functional form to be used would be determined by how close to reality these assumptions are. The chosen functional form must also allow the marginal implicit price of characteristics to depend on the levels of other characteristics of the house (Freeman, 2003).

Cropper et al (1988) simulated the performance of housing market from real data on buyer and housing characteristics from the Baltimore, U.S. using data from the 1980 Census of Housing and Population. They considered alternative functional forms and characteristics for the utility functions and the distribution of its parameters, buyer's characteristics and housing characteristics. After considering cases in which the estimated equations were correctly and incorrectly specified, they found that, when the hedonic equation was specified correctly, the quadratic and log-linear Box-Cox forms yielded close estimates, but when the hedonic equation was wrongly specified because of unobserved or proxied variables, the simpler Box-Cox function form performed better. Their conclusion was that, since correct specification may be difficult to achieve, the linear Box-Cox functional form is preferable.

4.5 Sub-Markets

Whilst it is possible to treat urban property markets as a single market and estimate a single price function to describe the equilibrium price within the market, it has been suggested that, in order to make the determination of hedonic equation, we divide the market into smaller homogenous markets. This is to make the hedonic equation measurable because, if a house price dataset contains data from more than one market segment, it is likely that the hedonic price functions for each segment are different (Day et al, 2003). Estimating a pooled hedonic price model may bias estimates of the true hedonic price functions.

Day (2003) links the existence of clusters and sub-markets for properties exhibiting different pricing structures to imperfections in the market mechanism. This could be due to either supply or demand related factors, the normal arbitrage that would be expected to equalize prices both within and across metropolitan areas may work either slowly, or not at all, varying attribute prices, the presence of independent price schedules, and the existence of a segmented market.

The spatial nature of property goods makes it different from other differentiated goods, this means impacts of environmental factors could be global or localised. If it is localised it may not be detected in a pooled hedonic price function making it imperative to estimate separate hedonic price models.

Name	***Equation***	***Implicit Prices***
Linear	$P = a_0 + \sum \beta_i z_i$	$\partial P / \partial z_i = \beta_i$
Semi-Log	$\ln P = a_0 + \sum \beta_i z_i$	$\partial P / \partial z_i = \beta_i . P$
Double-Log	$\ln P = a_0 + \sum \beta_i \ln z_i$	$\partial P / \partial z_i = \beta_i . P / z_i$
Quadratic-Log	$P = a_0 + \sum_{i=1}^{n} \beta_i z_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} \delta_{ij} z_i z_j$	$\partial P / \partial z_i = \beta_i + \frac{1}{2} \sum_{j \neq i} \delta_{ij} z_j + \delta_{ij} z_i$
Linear Box-Cox	$P^{\theta} = a_0 + \sum_{i=1}^{n} \beta_i . z_i^{\lambda_i}$	$\partial P / \partial z_i = \beta_i z_i^{\lambda - 1} P^{1-\theta}$
Quadratic Box-Cox	$P^{\theta} = a_0 + \sum_{i=1}^{n} \beta_i . z_i^{\lambda_i} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} \delta_{ij} . z_i^{\lambda_i} . z_j^{\lambda_j}$	$\partial P / \partial z_i = \left(\beta_i z_i^{\lambda - 1} + \sum_{j=1}^{n} \delta_{ij} z_i^{\lambda - 1} z_j^{(j)} \right) P^{1-\theta}$

Table 4.1: Possible functional forms and corresponding implicit prices: Modified from Taylor (2003)

Submarkets, could also arise where exogenous factors constrain individuals to participate in segments of a larger market. There many reasons for restrictions on demanders, factors like income, sub-group preference - high income and low income groups - and the difficulty experienced during search process, racial differences and social capital. It is assumed in hedonic models, that the individual economic agent is familiar with all the information necessary to evaluate all feasible exchanges as part of making her choice. It is also possible in the short-run, given the inelastic demand for housing, for spatial and structural factors to, independently and jointly generate submarkets (Michaels and Smith, 1990).

Watkins (2001) discusses numerous methods of identifying a sub-market, its boundaries and constituents. But, recognised that, most urban areas are not homogenous, definitions of sub-markets therefore, could vary from study to study. Municipal boundaries, school districts, racial division, housing types, income clusters have been used in different studies. While some studies use time series others use cross-section in delineating sub-markets using different statistical methods. This affects the definition of the sub-market.

A more explicit and universal approach, was suggested by Day et al (2003) using to the following classifications.

- Structure type: Households demand for a property of a certain type. For example, the market might segment between households looking to purchase houses with gardens, garages and those looking to purchase flats or maisonettes;
- Structural characteristics: Households may have strong preferences for a particular property characteristic. For example, households who only consider buying period properties with "original features" whilst others only consider purchasing modern homes;
- Neighbourhood characteristics: households may have strong preferences for localities providing certain amenities. For example, certain households may desire proximity to transport links or good quality schooling whilst others find no advantage in such proximity.

Gentrification has been identified as creating sub-markets and one of the major causes of change in the structure of urban areas. Slater (2002) defined gentrification as an "invasion of working class areas" by the middle class, who upgrade modest housing to an elegant residence, resulting in displacement of all, or most of, original occupiers. There are two types of displacements "direct displacement" where people are evicted and "indirect replacement" where people move out because of higher prices/rents and a new social structure.

Gentrification includes demolition of old housing and new constructions and construction of new houses on parks, playgrounds and green-fields. Two theories have used to explain this phenomenon, demand side theories and supply side theories. Demand side theories argue that the phenomenon is due to changing preferences and demographic factors which might lead to an increased demand from high income groups for centrally located or more expensive housing, housing rather than green-field, parks or playgrounds. Supply side theories or "gap theories" attribute gentrification to the presence of a rent gap and/or a value gap. A gap exists when the current rent or property value is far less than the potential value of the property. This gap makes it profitable for investors to enter the market (in some cases influence policy) and change the housing supply and the structure of a city (Lind and Hellström, 2003).

The "city" therefore can not be treated as one large "housing market". We have to estimate separate hedonic prices for specific locations. This could be necessary because, as Vandell (1995) noted, the housing characteristics being studied may be fixed in one location, more common in certain locations, or the customers, for some other reasons are located in a particular location.

Using hierarchical clustering technique to identify property sub-market, Day (2003) estimates the impact of road traffic noise on the market price of property in Glasgow and reports that in all but one of the sub-markets traffic noise have negative impact on property prices.

While the presence of market sub-markets may create a problem in analysing proximity to amenity (or disamenity) when such amenity is localised, it is found to be useful in the second stage regression and the estimation of demand equation, where identification becomes an issue (Day, 2003) (Goodman and Thibodeau, 1998) (Taylor 2003).

4.6 Identification:

In analysing welfare change using hedonic pricing model, a second stage regression is required. Two major identification problems arise in the second stage estimation, this is due mainly to the fact that in the second stage estimation the estimated implicit price may not contain information beyond the first stage estimation. First, the willingness to pay is not directly observed but calculated from marginal implicit price from the estimated hedonic price function. It is possible to have identical functions for both the willingness to pay and the estimated coefficient in the hedonic price function. The second problem arises because both the quantity of characteristics and their implicit price are exogenous in the hedonic price model. The implicit marginal price simultaneously, determines both the willingness to pay and the quantity of characteristics (Murty et al, 2003).

Various methods have been proposed to deal with the problem of identification most of which attempt to find ways to ensure that the marginal implicit price of characteristics vary independently of the demand shift variable. In most of the studies we have come across, the necessary condition for identification is multiple markets. Ekeland et al (2004) and Day (2001) are the most extensive literature reviews and treatment of this issue. The former is theoretical, while the later is practical application to the housing market.

The problem with a single market is that it assumes all consumers face the same equilibrium price schedule. Unlike in multiple markets where separate hedonic equations exist, where it is possible to obtain the necessary variation in price schedule to which individual consumers are reacting, in a single market it is difficult to obtain the necessary price variation for the estimation of parameters. Even though, multiple markets are not sufficient unless hedonic equations differ significantly between the markets, increasing the amount of exogenous price variation increases the reliability of the parameter estimates (Palmquist, 1991).

Even with varying marginal prices, it is also important, to impose structure or generate a new set of data in order to distinguish between the equilibrium marginal price schedule and consumers marginal bids functions. One way to do this is by imposing a structure on the system of equations or use multiple markets to generate multiple equilibrium price schedules. Palmquist (1991) observes that, identification of the demand equation is more difficult but, many studies achieved identification by restricting the functional forms or the variables. Rosen (1974) suggests the use of a non-linear hedonic price function because it generates varying marginal prices.

Identification could be achieved in a single market using the functional form restriction because, within a single market, individuals choose between different bundles and different marginal prices due to differences in socioeconomic characteristics. From Cassel and Mendelsohn (1985) we can deduce that, in addition to the direct relationship between marginal prices and housing characteristics, there is an indirect relationship through socioeconomic characteristics. It is this indirect relationship that is exploited to achieve identification. Quigley (1982) estimated the parameters of the utility function in a single market using an identical generalised constant elasticity of substitution (GCES) utility function, which has a homothetic functional form. This was possible because, prices of marginal characteristics vary using this method and homotheticity allowed the consumer choice to be standardised, so that they became observations along a common indifference curve.

It is also possible to assume a single price schedule in a city-wide single market, where some characteristics vary geographically throughout the city. An example is where the cost of obtaining amenities includes the cost of commuting, depending on ones place of residence within the city. This variation in the cost schedule between individuals allows identification without other restrictions. However, the data requirements for this method are higher (Palmquist, 1991).

In both multiple and single markets, spatially or temporally separated markets, it is possible if different marginal price schedules can be observed, for identification to be achieved.

Day (2001), following Bartik (1988) and Murray, (1983) suggest that, endogeneity (and identification) can be better handled through the application of instrumental variable techniques. Each of the endogenous variables in the demand equation is regressed on a set of exogenous variables – instruments. The "ancillary regressions" is used to calculate predicted values for the endogenous variables. Demand equations are estimated using these predicted rather than the actual values of the endogenous variables. Suggested instruments are household's socioeconomic characteristics namely: the number of members of the household, their ages and educational status.

4.7 The Hedonic Market and Marginal Prices for Housing Attributes

The outcome of any hedonic pricing study for a housing market which includes neighboourhood attributes such as environmental quality (i.e. attributes which are outside the influence of the landlord), should be interpreted with caution. This is due to three factors: the nature of the housing market, in particular, the spatial fixity of the housing property; the attempts by both households and landlord to optimise their utility before and after the change in environmental quality in ways which change the individual house attributes; and whether the change in environmental quality is local or city-wide. This issue has been extensively discussed theoretically by Bartik (1988) and analytically by Day (2001). Here we summarise the issues they raise.

As previously pointed out, changes in location (neighbourhood) attributes could be minor or substantial; they could be local or city-wide. Marginal, localised changes would have little impact on the housing market as a whole. Normally we expect an increase in the rent on properties in the improved area since the attributes of those properties have changed. If we assume that there are zero transaction and relocation costs, the improvement and resulting increase in rent would compel some households to relocate to a new house/location that is affordable to them. However, if we relax the assumption of zero transaction and moving cost, and assume a small increase in rent, in the short-run, we expect households to remain at the current location. In the longer run, we expect changes in both household size/characteristics and the dynamics of the housing market, which would affect household's demand for housing attributes.

If the environmental improvement is significant and city-wide, the relationship between the hedonic price function and the housing market is more complicated because the changes in supply and demand of housing will change the market-clearing price. Housing markets would respond to even small localized changes in environmental quality, because a change in the conditions of supply of a particular attribute in one part of the city would affect the market clearing implicit prices across the whole city. This is because of the factors we pointed out above, which make housing a unique commodity, especially, its geographical fixity.

If we assume an improvement in the environmental quality in one section of the urban area, the hedonic price function will not be affected and the change will simply give rise to an increase in rent in that locality. However, if the improvement in environmental quality is large in scale and spread, we would expect a shift in the whole hedonic price function, bringing about a reduction in the price per unit (implicit price) of the particular attribute (of environmental quality), across the entire market. In turn this would lead to a new market clearing rent.

Cēterīs paribus, we expect rent to be positively related to localised improvements in environmental quality, but negatively to city-wide improvements. Even though some properties may not be directly affected by the environmental improvement, market adjustments may well result in changes in their rental value.

Of course the overall impact on the hedonic price function will not be restricted to adjustments in the environmental quality coefficients. It seems likely that a number of concomitant effects will cause shifts in the supply and demand for housing characteristics. For a start, demand for property characteristics that are substitutes for the environmental attribute will decline. For instance, demand for double-glazed properties will decline in an area in which noise pollution has been reduced. Similarly, demand for complementary attributes will increase. For example, a reduction in air pollution might increase demand for houses with gardens. The implicit prices for these substitutes and complements will themselves have to adjust in order to ensure that the demand for these attributes is balanced by the supply.

Further, in response to the shifts in the hedonic price function, households, realizing that they are no longer at their optimal residential location, may choose to move to a new property. Indeed, we would expect that landlords at certain locations would find that the characteristics of the households wanting to rent their property would change. For example, reductions in the implicit price of environmental quality will encourage lower income households to demand properties in areas that they previously could not afford, so that, at any given level of environmental quality, there will be an increase in demand from lower-income households. Bartik (1988) hypothesises that lower-income households will have lower demands for other housing characteristics and landlords will change their levels of investment in properties to maximise their profits. For areas that experience large increases in environmental quality the reverse may be true. High income households will be attracted to the area and their higher demands for other property characteristics will encourage landlords to invest in property improvements that will increase their rental value.

It is evident that the overall change in the hedonic price function and the resulting change in rents and locational choice are extremely complex. For any one property, the eventual rental value will not be determined solely by the change in environmental quality experienced at that location. Instead it will be determined by the complex interaction of supply and demand across the entire market.

4.8 Analysing Welfare Change

After obtaining an estimate of the hedonic rent function, we plan to use it in a cost benefit analysis. By altering the values of the amenity variables in the hedonic rent function, we may investigate how rents might be expected to change in response to changes in amenity levels. However, we need to consider carefully whether this change in rent may be interpreted as a welfare change. In order to address this crucial point, we again refer to the contribution of Bartik (1988). We shall see that using the hedonic rent function gives rise to an upper bound to benefits from an improvement in amenities.

Let us assume that we have access to a hedonic rent function that is obtained from (pre-improvement) data on observed market rents, property characteristics and amenity levels. Assume also that we are interested in measuring the welfare change that results from an improvement in amenities in a small locality. In the context of the example we focus on, we might assume that daily hours of water supply to properties located in a particular small area increases by a certain number of hours, and we are interested in measuring the resulting welfare increase. We need to take careful account of the fact that, in accordance with the analysis outlined in Section 4.7, the improvement in amenities is itself expected to bring about a downward shift in the hedonic function.

In order to demonstrate that use of the (pre-improvement) hedonic rent function gives rise to an upper bound to benefits, it is useful to decompose the effects of the amenity improvement into four imaginary stages.

Stage 1: The amenity improvement occurs in the small locality. The hedonic rent function is constrained to remain unchanged. Rents increases at the improved sites, because they are moving upwards on the fixed hedonic function. Landlords are made better off by the increase in rent. Households are worse off because they are forced to pay more for an increase in amenity levels that they did not freely choose.

Households and landlords at <u>unimproved</u> sites are unaffected because nothing changes at these sites.

Stage 2: The hedonic function remains unchanged. Landlords in the improved area are allowed to optimally adjust their housing supply, e.g. to extend properties in the locality in which the improvement has taken place. Such adjustment must increase landlords' profits. However, households are no better off than before the improvement.

Stage 3: All landlords and households (both inside and outside the improved area) are assigned to the location that they will choose after the hedonic has shifted, although the rent that they actually pay is assumed to be determined by the original hedonic. As a result of this change, both landlords' profits and households' utilities fall.

Stage 4: Rents adjust in accordance with the new hedonic. The new hedonic is lower than the original hedonic, that is, the market rent for any given amenity level, ***ceteris paribus***, is now lower. However, whatever rent changes occur during stage 4, the landlords' gains (or losses) are exactly matched by households' losses (or gains). Hence there are zero net changes to efficiency benefits in stage 4.

The net efficiency benefits from all four stages are given by:

stage 1 and 2 increases in landlord profits

+ stage 3 loss in landlord profits

+ stage 1, 2, and 3, utility losses incurred by households

Hence we see that an upper bound to benefits is provided by the stage 1 and 2 increases in landlord profits (since all other components are negative). Unfortunately this upper bound is difficult to implement because it is difficult to estimate the changes in landlord supply occurring in stage 2.

In contrast, the stage 1 profit increase is easy to measure, and can usually, according to Bartik (1988) be used as an upper bound to benefits. The condition for this to be a valid upper bound is:

stage 2 profit increase

< absolute value of (stage 3 profit loss + stage 1, 2, 3 utility loss)

There are reasons to expect this inequality to hold. For example, many housing characteristics are difficult to alter, so stage 2 profit increases are expected to be small.

The way to measure the stage 1 profit increase is to insert old and new amenity levels into the pre-improvement hedonic function, and to compute the difference. This is the approach that is followed when we come to perform the cost-benefit analysis in Chapter 6. There, we will be careful to note that the estimates we reach are, as proven in this section, an upper bound to true benefits.

An empirical study by Bartik (1986) suggests that actual benefits are reasonably close to the upper bound just prescribed.

4.9 Previous Hedonic Pricing Studies

Espey and Lopez (2000) estimated the relationship between residential property values and airport noise and proximity to an airport in Nevada (USA). They find that proximity to the airport has negative impacts on property values. This is in contrast to Tomkins et al (1998) whose study of Manchester airport find airport to be an amenity.

Pennington et al. (1990) argues that, although airport noise affects property value, not all parts of the airport produce noise. Jud and Winkler (2006), examine the influence of the announcement of a new airport hub on house prices near the airport. Their results indicate that residential property prices in the neighbourhood of the Greensboro/High Point/Winston Salem metropolitan airport declined in the post-announcement period. Other airport and noise related hedonic price studies include Mieszkowski and Saper (1978), Cohen and Coughlin (2006), McMillan et al. (1980) and Nelson (2003).

There are numerous air quality hedonic price studies. Examples are, Chattopadhyay (1999), Trijonis et al. (1985), Batalhone et al. (2002), and Ridker and Henning (1967). It is known that air pollution affects health, irritates the eyes, nose and throat, and cause corrosion to metal and stone, contribute to dirty buildings and smelly neighbourhoods. Ridker and Henning (1967) provide empirical evidence to show how air pollution affects property values and how it affects household's location decisions.

Several studies have estimated the impact of school characteristics on house prices. Downes and Zabel (1997) recognise, the difficultly for individuals to decide which school characteristics to consider when deciding where to reside. On whole they report from their study in Chicago that, schools test score have significant impacts on house values. Similar results were reported by Haurin and Brasinton (1996) Cheshire and Sheppard (2002) and Jud and Watts (1981).

We expect a significant positive impact from the provision of public goods on the value of residential property. This is because public goods and residential property are complimentary. Houses in locations where there is efficient provision of amenities/utilities/municipal services, ceteris paribus, are likely to attract higher prices compared to areas where these facilities are poor or non-existent. Utilities like water supply, electricity, waste disposal, recreation centres, parking spaces, outdoor and street lighting are sought after in residential location decision. If we generalise our definition of "public good" to include neighbourhood and environmental quality, we expect locations with negative externalities to command lower prices. Bhattarai et al. (2005) estimate the demand for public goods in the Ohio (USA) housing market and report positive impact of public goods on residential property prices. They also found from cross-elasticity estimates that, school quality are substitute for environmental quality and neighbourhood safety.

Following King and Mazzotta (2001), we can summarise the advantages and disadvantages of using hedonic pricing method. In the context of property market study there are three advantages of using the hedonic pricing method. The method can be used to estimate values based on actual choices, this is especially because, property markets respond to changes quickly, so can be good indicators of value. Secondly, property records, rent and house prices (where they are kept) are very reliable. Data on property sales and characteristics can be obtained through many sources, either directly from authorities or field work. Finally, the hedonic pricing method is flexible, and can be adapted to consider several possible interactions between market goods and environmental quality.

Some of the limitations of the hedonic pricing model in property market study are as follows. First, the scope of environmental benefits that can be measured is limited to things that are related to housing prices. The method will only capture people's willingness to pay for perceived differences in environmental attributes, and their direct consequences. If people are not aware of the linkages between the environmental attribute and benefits to them or their property, the value will not be reflected in home prices.

Secondly, the method in its simplest form assumes that households have the opportunity to select the combination of features they prefer, and adjust/respond to price and changes in attributes, given their income. However, the housing market may be affected by outside influences, like taxes, interest rates, and several other factors.

Thirdly, the method overlooks mitigation action taken by households against externalities, example the use of double glazing to mitigate noise pollution.

Finally, at the empirical level, the results depend heavily on model specification. Another problem in estimation is the possibility of multicollinearity amongst characteristics. That is to say that, it is possible to have more than one characteristic, jointly-present, concurrently within a sub market and/or across the market being studied. This problem - which is a very common problem in empirical analysis - could be tolerated as long as it is not a very high correlation.

But we argue that, the most serious limitation of the hedonic pricing model is that, it seeks to explain consumers behaviour in the housing market by studying demand for housing alone, overlooking or assuming fixed supply of housing. The model assumes, as pointed out earlier that, all the required information on the housing market exists in demand side.

This approach, which, at best could be a partial analysis of the housing market is fundamentally flawed. This is because we can not ignore the impact of supply factors in explaining the economic agent's behaviour in the housing market. Since we have no information about supply we assume perfect adjustment in the market, with the market always in equilibrium. However, we know that, the equilibrium is only hypothetical in the absence of market adjustment; the resulting implicit price would be affected by market imperfection; and the marginal price paid for an attribute could be higher or lower. But because our research is a cross-section study of the housing market, we assume that, the supply of residential housing is fixed for the period of our study.

Chapter Five: Research data

5.1: Introduction

In this chapter we discuss our primary data. Two of our empirical chapters, the hedonic pricing model (Chapter 6), and the discrete choice model (Chapter 7) are based on this data. One data-related issue is left out of this Chapter: we do not discuss our IIA meta-analysis (Chapter 8) data in this Chapter, because it is a quasi-primary data. Instead, section 8.5 of that Chapter provides detail on the data used there.

5.2 Rent versus House Price data

There are two possibilities for the choice of price variable for residential housing: house price data and rent data. The choice of which to use would depend on data availability, proportions of households in rented and owner occupier residence, and the nature of the research project. In this research we decided to use rent data for two reasons. First, it is anticipated that rent and its differentials will contain more relevant information than house price. Households living in rented accommodation are more mobile, due to lower moving costs, and therefore more likely to move when there is a change in their economic circumstance, or a change in the housing attributes in their current location or in another location.

Second, obtaining house price data would be difficult in the study area. This is because house owners may not know and/or may be unwilling to disclose the value of their property. Although most transactions in the real estate market are formal in Nigeria, unlike some countries, the parties to the transaction are not obliged to make it public. We could also argue that, even where house price data are kept and made public, not all properties are offered for sale at the same time and estimates by estate agents/realtors are only a rough guess if the property has not recently been traded. It might be possible to use estate agents records of traded properties after taking account time differences but, the property market in the study area is very complex and in some cases secretive.

5.3 Pilot Study and Feedback

The first draft of the questionnaire was administered to faculty members and research students in the Schools of Economics and Development Studies at the University of East Anglia (UEA), Norwich during the summer of 2006. The objective was to ensure that questions are presented in an understandable way and that the data could be coded and estimation carried out from the coded data. Utility provision (water and electricity supply) is efficient in Norwich and therefore would not make for an interesting discrete choice or hedonic pricing research topic. Therefore, the questionnaire administered in Norwich only contained questions suitable for estimating housing location choice probabilities.

Although there was poor response from the target population, we were able to estimate a multinomial logit model of residential location among Norwich academics, with postcodes (NR1-NR8) as alternatives. One of our findings was that older academics are more likely to live away from the city while younger academics and graduate students more likely to reside close to the university and the city centre.

The Kano questionnaire was circulated among some lecturers at the Bayero University, Kano for comments. There was also a one-hour session on the questionnaire and problems of field-work/data collection with the selected research assistants and research supervisors. The feedback from these two consultations was used to produce the final version that was administered among renting households in Kano.

5.4 Data collection

The primary data was collected in the study area (Kano, Nigeria) between October and November 2006. A little over 3000 questionnaires were distributed. The complete questionnaire is shown in Appendix V. Typically, a single questionnaire took around ten minutes to complete. The target respondents were households living in rented properties in Kano city, Northern Nigeria. This is because, as pointed out above, households living in rented houses have higher mobility than households who own the house they live in.

Research assistants were employed to administer the questionnaire. Part of their role was to actively encourage respondent participation, and to explain the questions if necessary; this was considered important because we were keen to avoid losing respondents through non-literacy.

To facilitate the data collection exercise, 34 research assistants and eight coordinators, one coordinator per local government, were used. Research assistants were paid an amount close to monthly minimum wage; supervisors were paid a little extra and were supplied with mobile top-up in order to facilitate communication and coordination. The survey was part-funded by the Bayero University Kano (BUK) from its internal university research grant.

Research assistants were selected from Final year Economics and Sociology undergraduate students at the Bayero University Kano, on the basis of previous experience such as administration of the national population census, or national or local elections. Supervisors were drawn from postgraduate students and academic staff of higher-education institutions in Kano State.

Our advertisement for the post of research assistant attracted 45 applications, out of which 34 were selected, of whom two were upgraded to supervisors on the strength of their higher-level qualifications and work/research experience.

We are fortunate to conduct the research at the time we did, because there was relative peace in Kano and Nigeria, a place that is noted to be politically charged. A few months after our survey, one of our data collection centres was engulfed with political crisis. We also drew experiences from the National population census which was conducted few months before our survey, in terms both logistics and experienced personnel. The 2006 population data was collected in case we need it, to control for sampling bias in choice based sampling. Our population data is therefore from the latest census and contemporaneous, collected around the same time with our survey. This solves the problem of having to rely on estimated population data from different sources and/or time. After careful scrutiny of the data, we concluded that our data do not have sampling bias.

5.5 Curbstoning

Some of the research assistants cheated in data collection exercise, a phenomenon very commonly known in demography as "curbstoning" (also kerbstoning). After rigorous cleaning of the data, and dropping blank or incomplete questionnaires, about 19% of the questionnaire was discarded and 2438 (largely) complete questionnaires retained. This is the sample size in much of our analysis.

We used a simple data cleaning technique in excel, sorting the data and deleting multiple entries. This method has precedence in population census data collection. It is possible that in some cases we have committed "type one error", i.e. the rejection of valid responses, but we are more inclined to do that rather than to retain invalid responses.

5.6 Research instrument – the questionnaire

The questionnaire, printed in an 8-page portfolio format, consists of 30 questions. It is reproduced in Appendix V of this thesis. The objective of the whole data collection exercise is to obtain information about the individual household (socio-economic characteristics) and the residential property (housing attributes). This information is required for both the hedonic pricing research and the discrete choice analysis. These socio-economic and location attributes are necessary to estimate choice probabilities using either MNL, CLM, or mixed logit.

Using carefully worded questions, to avoid concerns that could arise due to ethical, cultural and related sensitivities, the following socio-economic characteristics of the individual were included in the survey (we had earlier obtained ethical approval for the questionnaire from the relevant university committees): place of residence; previous residence; age; household/family size; marital status; number of children in each of the following category 0 – 4, 5 – 6, 7 – 12, 13 – 18, 18 and above; respondent's highest education qualification; nature of respondent's current occupation; respondent's other major source of income; respondent's gross annual income from employment; respondent's spouse highest education qualification; and respondent's spouse current occupation.

Table 5.1 presents definitions and descriptive statistics of all household attribute variables, while Table 5.2 presents the same for all housing attribute variables.

Variable	Obs	Mean	S-Dev.	Min	Max	Definition
County	2437	4.02	2.06	1	8	Categorical Variable for the eight Local Governments/Counties within Kano City/Metropolitan
Area	2438	3.62	2.14	1	6	Categorical variable representing six major location classifications for this research. Old City; Low Density/Government Reserved Area; Close to Airport; Close to one of the two Industrial Estates; and Other
Mode of transport to work	2393	3.04	1.60	1	7	Categorical variable: Car; Bus; Cycle; Walk; Others; and Not Applicable
Time commuting to work	2271	1.50	0.78	1	5	Continues variable – reported estimated daily average
Last area of residence	2383	1.82	0.87	1	3	Categorical variable: Area above; Another City; No previous residence
Years of education	2416	10.22	5.18	0	18	Continues variable – estimated based on reported highest qualification
Spouse years of education	2366	5.95	2.82	1	11	
Annual income	2380	2.33	1.61	1	10	Income from main occupation reported in range
Other sources of income	2396	5.18	2.79	1	8	Categorical variable: Support from family members; Providence; Fixed Assets; Financial Investment; Moonlighting; Secondary Occupation (e.g. Part-time); Private Consultancy; Other (please specify); and None
Respondent's Current occupation	2409	3.97	3.70	1	16	Categorical variable: Manual; Businessman/woman; Civil servant; Teacher; Corporate Sector/White Collar; Lecturer/Researcher; Farmer; Security Worker; Law Enforcement Agent; Medical Doctor; Nurse/Midwife/Social Worker; Retired; Unemployed; and Other
Spouse current occupation	2354	9.65	5.64	1	16	
Spouse years of education	2366	5.95	2.82	1	11	Continues variable – estimated based on reported spouse highest qualification
Marital status	2400	1.89	0.49	1	5	Categorical variable: 1 Single: 2 Married; 3 Divorced; 4 Widowed; 5 Separated
Age	2418	39.10	10.10	20	60	Continues variable from group data - 20 and Below, 21–29, 30–39, 40–49, 50 –59, 60 and above
Family size	2274	5.93	4.23	1	40	Continues variable: Number of children in each category.
Children Age 0-4	1569	1.59	0.75	0	7	
Children Age 5-6	1285	1.50	0.90	0	14	Obs: The number of households with at least one child in respective category
Children Age 7-12	1059	1.84	1.08	0	11	
Children Age 13-18	683	1.99	1.26	0	10	
Children Age 18 and above	548	2.89	2.29	0	17	

Table 5.1: Variable definition and descriptive statistics (Household Attributes)

Variable	Obs	Mean	Std. Dev	Min	Max	Description
Rent	2415	42.71	29.70	0	250	Annual rent in local currency ’000 - Interval data
Number of bedrooms	2426	2.93	1.30	1	6	Number of bedrooms, bathrooms/toilets. Size of floor area garden/courtyard in sq ft - Interval data
Number of toilets	2411	1.43	0.62	1	4	
House floor area	2392	2597.39	249.90	2500	4251	
Garden/Courtyard	2365	139.16	256.35	0	2501	
Electricity supply	2426	4.60	3.42	0	20	Average number of hours of water supply and electricity supply daily in individual houses - Interval data
Electricity supply squared	2426	32.76	45.81	0	400	
Water supply	2426	4.6	6.73	0	20	
Water supply squared	2426	66.39	130.01	0	400	
Private primary/nursery schools	2438	0.46	0.49	0	1	Dummy variable. 1=Private school 2=Public schools 3=Market 4=Highway 5=Airport 6=Industries 7=None. Choice of as many as applicable
Public primary schools	2438	0.57	0.49	0	1	
Market	2438	0.32	0.46	0	1	
Highway	2438	0.28	0.45	0	1	
Airport	2438	0.11	0.31	0	1	
Industries	2438	0.13	0.33	0	1	
Type of house	2376	2.29	1.32	1	5	Dummy variable. Types of houses (flat, bungalow, duplex, traditional house, and other). House provider (local authority, employer, private, other organisation, other individual(s), and not applicable) .
House provider	1441	4.40	1.22	1	6	
Flight path (if close to airport)	2333	2.47	0.65	1	3	Dummy variable. 1=Yes 2=No 3=Not Applicable

Table 5.2: Variable definition and descriptive statistics (Housing Attributes)

Other attributes related to both the property and the households obtained in the survey include: mode of travel to work; cost of commuting to work; commuting time; whether the respondent live close to close-relatives; otherwise the frequency of visit to relatives in an average week; and estimated of cost of commuting for each visit.

The most important piece of information about the household that is extracted from the questionnaire is the household's choice of residential location. Figure 5.1a shows a map of Kano, while Figure 5.1b shows a simplified map indicating how the city has been divided into six distinct areas, for the purposes of the residential location choice model of Chapter 7. The six areas are listed in Table 5.3, with the number of households in each area:

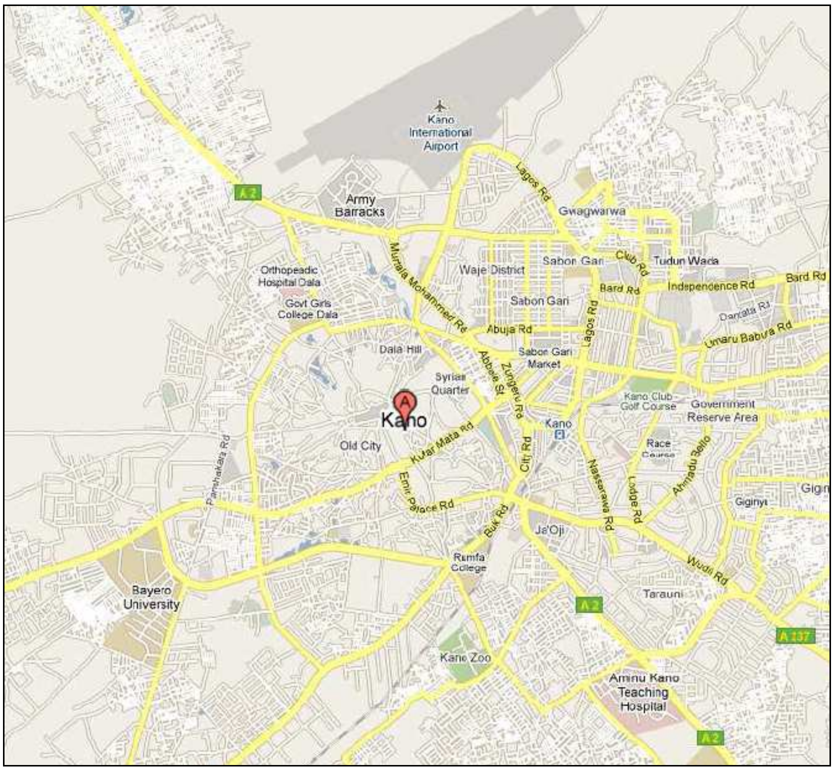

Figure 5.1a: Ariel Map of Kano City (Source: Google Maps)

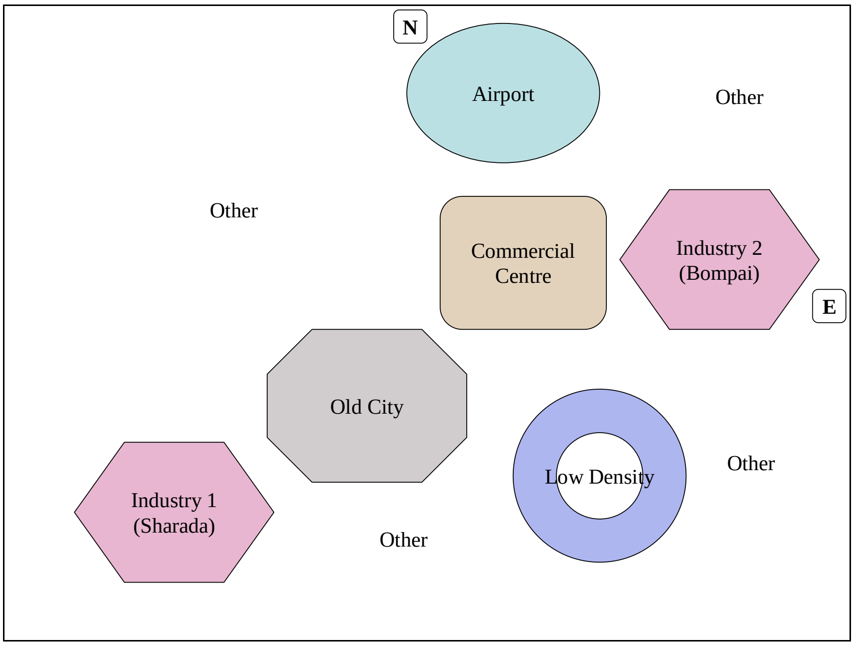

Figure 5.1b A sketch locations/alternatives in Kano City

Area Code	*Area*	*Number of households*	*%*
1	Old City	713	29.2
2	Low Density	278	11.4
3	Airport	232	9.5
4	Industrial Area 1 (Sharada)	71	2.9
5	Industrial Area 2 (Bompai)	289	11.9
6	Other	855	35.1
	Total	**2,438**	**100**

Table 5.3: Tabulation of sampled households between areas.

The next most important variable is rent paid by the household. Figure 5.2 shows average rent for the six areas. We see that rent is highest (unsurprisingly) in the "low density" area, and lowest in the "Airport" area. Two negative externalities would explain why the area close to the airport attracts low rent: airport noise; and heavy traffic congestion on the highway close to the airport. Also (see Figure 5.4 below), this area has the lowest average hours of both water and electricity supply.

Rent is also low in the "old city". There are many reasons for this. The old city is highly congested; the housing structure is old; most houses are built with mud (adobe) bricks; it predates modern state institutions, and therefore lacks proper town planning.

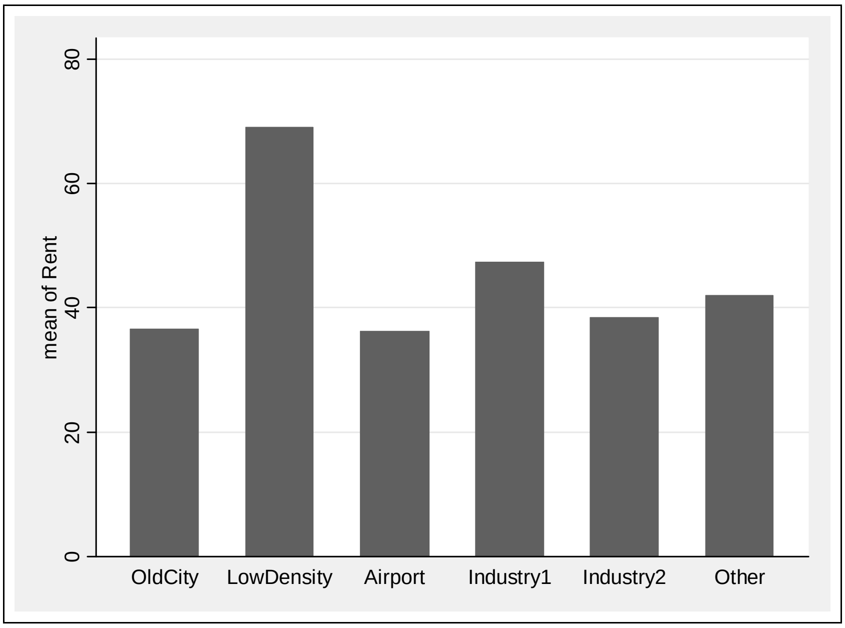

Fig 5.2: Average rent ('000 Naira per annum) in the six locations (mid-points of rent-intervals used in calculations)

Average annual income for each area is shown in Figure 5.3. As with rent, incomes are highest in "Low density" and lowest in "Airport".

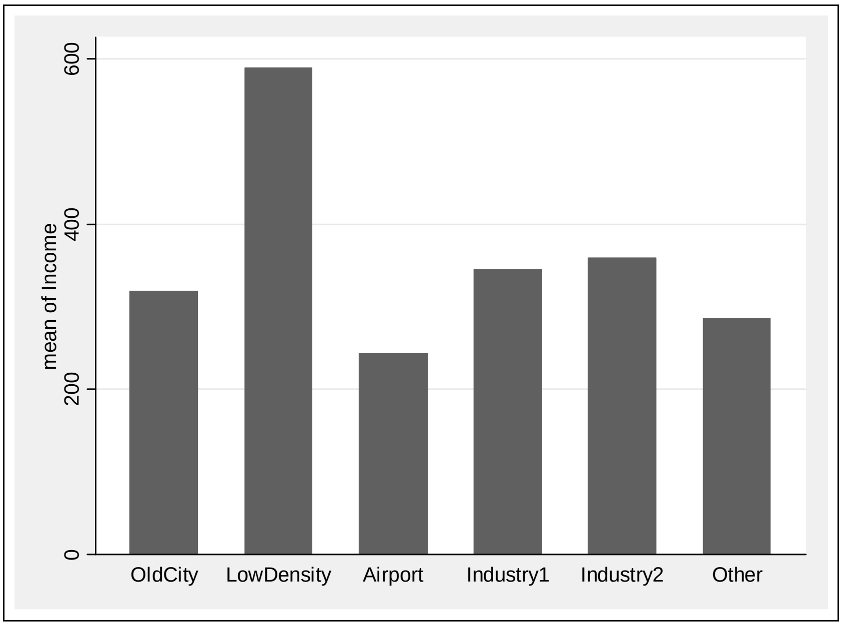

Fig 5.3: Average income ('000 Naira per annum) among household across the six locations (mid-points of income-intervals used in calculations)

We are also very interested in the reliability of water and electricity supply in the six locations, represented by the number of daily hours of supply. Averages of these are shown in Figure 5.4. Here, we see particularly high variation in water supply between areas, with Industrial Area 1 (Sharada) enjoying by far the highest supply, for the simple reason that this area is situated adjacent to the water treatment plant. Industrial Area 2 (Bompai) has the second lowest average of water supply, as a consequence of being far away from the water treatment plant.

"Old city", despite being close to the water treatment plant, has relatively poor water reliability. This is due to the negative effects of congestion and lack of town planning on water pressure.

The area with the most reliable electricity supply is the Low Density area. This is perhaps a consequence of the area being inhabited by government officials, with influence over the allocation of many amenities including electricity (although less so for water; being far from the treatment plant, they are not in a position to influence the water pressure in their locality).

“Old city” is again relatively disadvantaged in terms of electricity supply. Illegal electricity connections, refusal to pay electricity bills resulting from the difficulties of monitoring such a congested area, and the pressure on electrical transformers due to overload, leads to lower average electricity supply in this area.

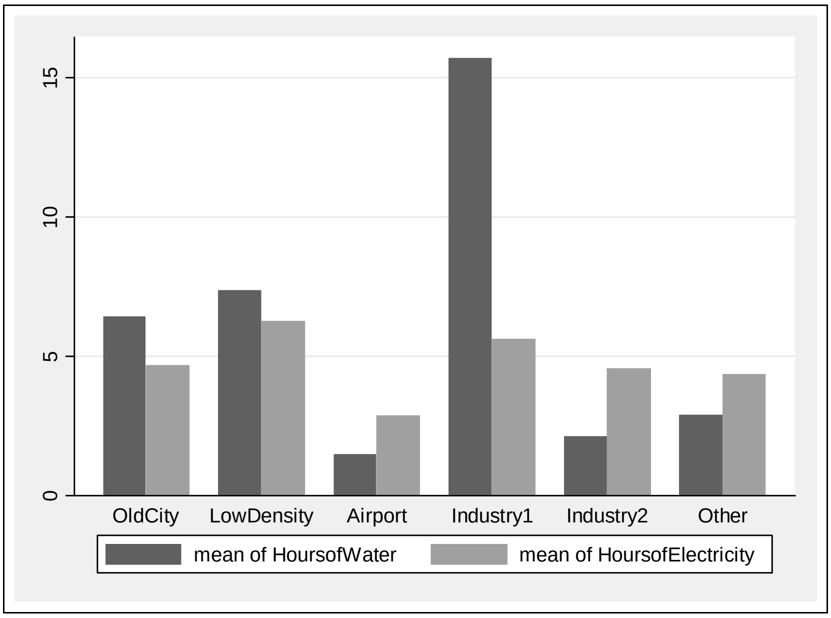

Fig 5.4: Average daily hours of water and electricity supply across the six locations (mid-points of intervals used in calculations)

Other attributes of the property included in the questionnaire are: number of bedrooms; number of bathrooms/toilets; size of the floor area in sq ft; size of garden/yard (if any) in sq ft; rent provider/house ownership; whether close (i.e. within 2 kilometre distance) to any the following: government approved private, nursery/primary school, public primary school, major market, highway/by-pass/express, airport, industrial estate; if close to airport, whether the house is on the runway/flight path, type of house e.g. bungalow, duplex, tradition etc.

Chapter Six: Valuing Utilities Provision Using Hedonic Price Model

6.1 Introduction

In this Chapter, we apply the Hedonic price modelling framework outlined in Chapter 4 to the data set described in Chapter 5. The principal objective is to obtain estimates of individuals' willingness-to-pay (WTP) for an additional daily hour of water supply and electricity supply.

Water is a necessity, and its deficit kills more than five million people each year in developing countries, this is ten times the number of people killed in civil wars (WHO/UNICEF, 2005). The poor provision of domestic water in developing countries is caused by lack of investment, rapid and unplanned expansion of cities and poor distribution/reticulation networks. The same applies to electricity although to a lesser extent because, alternative energy sources exist, even if using crude methods.

Economists and policy makers have been interested in household preferences and estimating willingness to pay (WTP) for public utilities in developing countries. We believe it is possible to use house price/rent data as revealed preference to estimate the households WTP for utilities. This is because where there is spatial variation in the supply of both utilities, we envisage that house price/rent differentials could be used to estimate the WTP for public water and electricity supply in the study area, Kano city, Nigeria. Kano is the second biggest commercial centre and third largest city in Nigeria.

This study is based on the primary data collected in Kano city that was described in Chapter 5. Number of hours of public water and electricity supply would be used as a proxy for a reliability index. It is expected that this would affect the rental value of houses in the study area. It is this rent differential that is used to estimate the WTP for water and electricity supply after controlling for other housing attributes. This research could not have been possible or would have been futile in a developed economy where the supply of public utilities is regular, for most of the time and in most places.

It is believed that preferences for housing related public utilities goods affect the way households form their decisions on where to live. Economic theory suggests that a rational economic agent would show a preference for a bundle of residential housing attributes that contain an optimal amount of physical/structural attributes, public utilities, public goods, and the least amount of negative externalities. This chapter treats housing as a differentiated good and estimates the WTP for some public utilities, residential housing physical/structural characteristics and neighbourhood attributes using rent data.

It is possible to use hedonic price to extract information on the consumer's valuation of utilities from the market for houses after controlling for other influences. Because the supply and quality varies across space within the study area, individuals would reveal their preference for public utilities and other neighbourhood attributes through their residential location choice. This is because residential housing price may include a premium for positive externalities such as, clean, accessible and quiet areas, and a discount for noisy, dirty and inaccessible areas. From this we could estimate the demand for and price of public utilities from the demand and the price differentials revealed in the housing market (Grafton et al, 2004).

Neoclassical, maximalist utility theory analyses how the household rationalises housing needs given income constraints. We could use the maximalist economic theory to analyse housing demand by households, who select from a menu of characteristics based on preferences in order to maximize their welfare, and housing supply by landlords, who produce houses with different characteristics and who, thanks to providence, inherit some neighbourhood characteristics, and price their property based on cost incurred and aim to make profit. In this context, the household is assumed to have an organised system of preferences, using considerable knowledge and skills to evaluate alternatives, then to selects the alternative which yields highest utility (Wong, 2002).

There are two basic neo-classical analyses of differentiated goods and associated attributes. The first approach is due to Lancaster (1966 and 1991) who proposed a theory of consumer utility based on characteristics rather than the goods themselves because, goods do not in themselves give utility to the consumer, it is characteristics/attributes which give utility. Lancaster argued that the consumer is not interested in goods, but in their attributes or characteristics. It is from these characteristics (most of which are consumed collectively) that the consumer derives utility. Individual consumers, subject to budget constraints, seek to maximise their utility by choosing goods that will give them the best combination of desired characteristics.

In order to explain the decision making process involving multiple goods, this approach assumes that, the consumer's utility function is separable. The consumer is expected to allocate resources between groups of goods, and attempt to optimise within each group by selecting the best combination of characteristics within the group. The individual consumer will allocate resources between groups, for example accommodation, leisure, food, transport etc; she will subsequently make a choice within a particular group, of the combination of characteristics which maximise her utility at least cost.

The Lancaster approach recognises a more complicated analysis, where goods have many attributes, and these attributes could be shared by more than one good and that, combined together, goods possess attributes different to their individual attributes (joint demand attribute). See Wong (2002) for some of these extensions.

The second approach uses the observed market clearing price (the interaction of consumers with heterogeneous taste for different combination of attributes, with the supply of goods with given attributes) and specific amounts of attributes associated with each to derive individual implicit or hedonic price. Hedonic price which is due to Rosen (1974) provides the functional relationship between the market clearing price of a good and its constituent attributes.

Hedonic price method of *valuing attributes*:

$$P(X) = f(x_1, x_2, ..., x_k) \tag{6.1}$$

$P(X)$ is the market clearing price; X - vector of housing attributes; and $x_1, x_2, ..., z_k$ – individual attributes of a good

In both cases, a rational consumer is expected to maximise her utility by consuming goods with given attributes subject to budget constraints. Consumer's WTP will depend on her income and taste which determines her preferences for given combination of characteristics. The solution to this optimisation problem would require that the marginal rate of substitution between characteristics and the price of the good must be equal. The consumer's WTP for an attribute must be equal to the implicit price of the attribute in the market.

6.2 Hedonic Pricing Model

The hedonic price model is premised on the fact that housing comprises of various characteristics which are not directly traded but that the implicit marginal price of the constituent characteristics can be derived by regression. In recent times, the most popular application of hedonic price model is in the housing market. The most extensive, although dated survey on hedonic price model with application to environmental economics, is provided by Cameron (1998).

In consuming housing goods, a rational consumer is expected to maximise her utility by selecting a given bundle of characteristics, subject to budget constraints, savings/investment decisions. Consumer's WTP will depend on her income and taste which determines her preference for given combination of characteristics.

It is expected that, the market, given current state of technology, will generate various combinations of house characteristics. The price-characteristic relationship is identified through the exchange between the buyers and sellers. The transaction resulting from supply and demand interactions could be used to generate data on prices and characteristics or the marginal value of a characteristic which is the partial derivative of the price equation with respect to a particular characteristic. Palmquist (1984)

In addition to producer generated structural characteristics; there are also spatial characteristics which are factors external to the producer's control, and those generated by public policy.

$$P_j = p(X_j) \qquad (6.2)$$

Equation 6.2 is the hedonic price function. The price of house j, P_j is a function of a vector of price of characteristics, structural, spatial, and neighbourhood characteristics. Differentiating P with respect to an element of X would yield the implicit price of a constituent attribute/characteristics. Hedonic pricing model therefore provides a link between a differentiated product and constituent characteristics through its price.

$$P_j = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon \qquad (6.3)$$

When the vector of house price is regressed against the vector of house characteristics, the coefficients, also called the hedonic weights, β_i, (the part of a product's overall price attributable to a given characteristic) are usually interpreted as the price of the corresponding characteristic. Day (2003), Hulten (2002).

The main objective of this chapter is to estimate, from residential house rent data, the value people attach to utilities, (water and electricity) but we have to control for other attributes namely, neighbourhood attributes (proximity to private primary/nursery schools, public primary schools, market, highway/express, airport and industries) and structural attributes (number of bedrooms, number of toilets, house floor area and garden/courtyard). The marginal price of these housing attributes could be treated as analogous to consumer's WTP. We intend to draw policy implications and offer suggestions.

We indicated in Chapter 4 that decision on functional form is an empirical one. It is possible to use a simple linear hedonic price function on rent/house price data to estimate inverse demand function - a good approximation of the marginal bid function - by regressing the price data on house attributes. The resulting coefficients would reveal the marginal price that households implicitly pay for each attribute. However, with non-linear hedonic price function and preferences, observed choices do not yield a well-behaved inverse demand function, this is made more complicated where there is substitutability or complementarity between attributes. When the hedonic price function is non-linear, the implicit price would be different for each property market, and as such, welfare estimates would be market-specific (Day 2001, Bartik 1988).

It is possible to conduct an ex-post analysis of the effect of a change in housing attributes. Let us consider two possible scenarios (see sections 4.7 and 4.8 for all possible scenarios). First, consider a marginal change in environmental quality, within a small area. If we add another assumption to this scenario, full information and a zero moving cost, this ensures that prices remain the same following adjustment. For the second, and more interesting case, consider a non-marginal change applying to a wide area, which would alter the supply and demand functions of housing attributes. We may impute the welfare impact of such a change by looking at the household demand for attributes after the change. This is because, we assume that households would choose to consume quantities of each housing attribute, up to the point where their demand curve for that attribute intersects its implicit price. At the household's optimal choice, the household's WTP for an additional unit of the attribute is equal to the implicit price of the attribute.

6.3 Relevant Studies

There is an enormous amount of literature in both theory and empirical work of hedonic pricing in housing/urban studies. Espey and Lopez (2000) estimate the relationship between residential property values and airport noise and proximity to the airport in Nevada and find proximity to airport has negative impacts on property values. This is in contrast to Tomkins et al (1998) whose study of Manchester airport finds the airport to be an amenity. Pennington et al. (1990) argues that, although airport noise affects property value, not all parts of the airport produce noise. Jud and Winkler (2006), examine the influence of the announcement of a new airport hub on house prices near the airport. Their results indicate that housing property prices in the neighbourhood of the Greensboro/High Point/Winston Salem metropolitan airport (North Carolina, U.S.A.) declined in the post-announcement period. Other airport and noise related hedonic price studies include Mieszkowski and Saper (1978), Cohen and Coughlin (2006), McMillan et al. (1980) Nelson (2003) and van Praag and Baarsma (2005).

There are numerous air quality hedonic price studies. Examples are, Chattopadhyay (1999), Trijonis et al. (1985), Batalhone et al. (2002), and Ridker and Henning (1967). It is known that air pollution affects health, irritates the eyes, nose and throat, and corrodes metal and stone, discolour buildings and dirty neighbourhoods. Ridker and Henning (1967) provide empirical evidence to show how air pollution affects property values and how it affects household's location decisions.

Several studies have estimated the impact of school characteristics on house prices. Downes and Zabel (1997) recognise, the difficultly for individuals to decide which school characteristics to consider when deciding where to reside. On whole they report from their study in Chicago that, schools test score have significant impacts on house values. Similar results were reported by Haurin and Brasinton (1996) Cheshire and Sheppard (2002) and Jud and Watts (1981).

More recently, Bayer et al (2007), develops a framework for estimating household preferences for school and neighbourhood attributes in the presence of sorting, addressing the endogeneity of school and neighbourhood characteristics. Cao and Hough (2007) estimates a hedonic price model to determine implicit price of proximity to bus routes and a negative impact of bus transit on apartment rent after controlling for other factors. They speculate that the negative relationship found could be due to spurious relationships from other causal factors and the nuisance effects of bus transit itself.

Hamilton (2007) examines the average price of accommodation in the coastal districts of Schleswig-Holstein, Germany using landscape and other characteristics of these districts. The analysis shows that an increase in the length of open coast results in an increase in the average price of accommodation. Williamson et al (forthcoming) derive economic values for housing relating to mitigating the effects of acid mine drainage using 21 years of housing sales data in West Virginia. The results indicate that, being located near an impaired acid mine drainage stream has an implicit marginal cost of $4,783 on housing.

We expect a significant impact of the provision/quality of public utilities on the value of housing property. This is because public utilities and residential property are complementary. Houses in a location where there is efficient provision of amenities/utilities/municipal services, all things being equal, are likely attract higher prices compared to areas where these facilities are poor or non-existent. Utilities like water supply, electricity, waste disposal, recreation centres, parking spaces, outdoor and street lighting.

If we consider the relationship between "public goods" such as neighbourhood and environmental quality, we expect locations with negative externalities to command lower prices. Bhattarai et al (2005) estimate the demand for public goods in Ohio housing market and reported positive impact of public goods on the housing market. They also found from cross-elasticity estimates that, school quality are substitute for environmental quality and neighbourhood safety.

Using data from a sample of rural households in one region of the Philippines, North and Griffin (1993) estimate the determinants of the rental value of dwellings using the bid-rent approach to the hedonic price model. The main objective is to obtain the relative valuation these households place on owning a private source of water and distance to a public or communal source. The results indicate that low-middle and high-income households value an in-house piped water source highly relative to other characteristics of their homes. Middle-and high-income households value deep well or piped water in the yard, although at a substantially lower level than piped water in the house.

To the best of our knowledge, ours is the first hedonic pricing study of electricity supply. We have not come across any study in the literature that estimates the valuation or WTP for electricity supply.

Yusuf and Koundouri, (2005) is the only previous hedonic pricing method study on domestic water supply valuation that we have come across. Using imputed monthly rent in a study of Indonesian housing market, by comparing rural and urban areas, with water-related characteristics of the house as the "focus variables", the study concludes that households value access to safe and improved domestic water sources. The study report estimates of WTP for having piped water, pumped water, and well water as 14,053, 5,548, and 748 respectively, in local currency.

Epp and Al-Ani (1979) examine the effect of water quality and value of non-farm residential property adjacent to small rivers and streams in rural Pennsylvania. They conclude that water quality significantly affects the value of adjacent residential houses in the study area. Using a mixture of hedonic pricing and cost benefit analysis, Coelli et al (1991) study agricultural and domestic water supply in Western Australia, especially the benefits of a “comprehensive water supply scheme”. Their conclusion is that the benefits of the water scheme are considerably less than the costs.

6.4 Econometric Model Specification

For the analysis of our data we choose interval regression because our dependent variable (annual rent) is grouped into intervals. Table 6.1 shows the distribution of rent between intervals. The dependent variable could only be obtained as intervals, because of elicitation problems. The extra information provided by interval regression allows more efficient estimation of coefficients and identifies the variance of the error term. This method has been used by van Doorslaer and Jones (2003), self-reported health condition; Piekkola (2004), wages and collective bargaining; and Shen (2008), WTP for eco-labeled products.

Interval	***Number of households***	**%**
< 30,000	44	1.8
31,000 – 39,000	1,218	50.4
40,000 – 49,000	440	18.2
50,000 – 59,000	241	10.0
60,000 – 69,000	157	6.5
70,000 – 79,000	100	4.1
80,000 – 99,000	58	2.4
100,000-149,000	49	2.0
150,000-199,000	68	2.8
> 200,000	23	1.0
NA	17	0.7
Total	2,415	100

Table 6.1: Distribution of households between rent intervals (Naira per annum)

When data is collected in group form, a range of two extreme values are created. Analyses of this type require a generalization of censored regression known as interval regression. The extreme values of the categories on either end of the range are either left-censored or right-censored. It is possible to use mid-point in a linear regression but this is an inferior estimation method (Stewart, 1983). Some studies have used OLS regression on the midpoints of the intervals. But there is a danger that the results would not reflect the uncertainty concerning the nature of the exact values within each interval nor would it deal adequately with the left-and right-censoring issues in the tails. In short, OLS has limitations on the amount of information used in the data analysis.

It would also possible to use ordered probit/logit regression. This sort of model is often applied to attitudinal data, for which the outcomes are ordered. One feature of such models is that the "cut-points" (i.e. the values separating different outcomes) are assumed to be unknown parameters requiring estimation. However, as explained by Daykin and Moffatt (2002), it is not appropriate to apply the ordered probit model in situations in which the cut-points are known in advance. As is clear from the left-hand column of Table 6.1, the cut-points are known in this case. For this reason, we do not use ordered probit/logit. Interval regression is undoubtedly the natural approach in the presence of known cut-points.

We intend to conduct detailed diagnostic tests and obtain detailed statistics such as the R^2 which is not available in STATA with interval regression. We have overcome that problem by calculating a suitable alternative.

6.5 The Data

As mentioned earlier, residential housing is a special type of differentiated good. Different types of hedonic pricing studies could be undertaken from the huge amount of information on the characteristics of a residential housing in any location. The data used in this study is cross section data which was obtained from a questionnaire administered at the study area, Kano city in northern Nigeria in November 2006.

Unlike in other places, for example Scotland, where it is mandatory to make public, data on house prices (Lake et al, 2000), data on house prices is not available in the study area. A solution would have been to observe the market directly but, not all houses are offered for sale and the houses in the market are offered at different times.

Valuation by estate agents is not reliable because hedonic pricing model requires individual consumer's willingness to pay for a particular property.

Rent data is used in this research not only because house price data is not available, but because we believe that rent data would be a better variable, compared to house price data, to capture the WTP for housing attributes, because mobility amongst household living in rental properties is higher, therefore the price for rent is more competitive.

From the questionnaire we obtained information on rent, structural attributes and neighbourhood attributes of each individual property. A related dummy variable was used to denote proximity of individual houses to each of the selected neighbourhood attributes. Proximity is defined as residing within two kilometre radius. Average number of hours per day of water and electricity supply grouped in five categories was asked (none; 1-2; 3-6; 7-12; 13-15; and 16-24 hours of water and electricity supply). There could be a small "noise" in the water supply data, because at present, households do leave their taps on, using tanks at night or undertake a "vigil", to collect water when the supply comes back. The rent data is the reported annual rent of each individual property.

Given that the data is cross-section, routine diagnostic checks were carried out. These include basic statistics, (descriptive statistics of the variables are presented in Table 1) and lowess, a semi parametric test, to establish the nature of the relationship between the dependent variable and the explanatory variables of interest, water supply and electricity supply (Figs 6.1a – 6.2b). Lowess is used to obtain a graph from a locally weighted regression of rent and log-rent on water supply and electricity supply. Lowess is mainly used in fitting models to localized subsets of the data to generate a function that describes the deterministic part of the variation in the data, point by point without the need to specify a general function of any form to fit a model to the data, only to fit segments of the data (Cleveland, 1979).

Lowess is reliable because it combines the simplicity of linear least squares regression with the flexibility of nonlinear regression. Apart from data exploration, it is also used in diagnostic checking of parametric models, and providing a nonparametric regression surface (Hawker et al, 2007), (NIST/SEMATECH, no date), (Cleveland and Devlin, 1988), (Cleveland, 1979).

Rent fitted against water supply in its linear and log forms exhibits a quadratic/inverted U shape relationship, this is consistent with theory, figure 6.1a and figure 6.1b. The impact of water supply on rent is positive up to a certain point, after which it starts to decline. This is mostly because people demand water during the day; the peak period of consumption is in the afternoon and early evening and partly because water is a good that has a satiation point. Lastly, unlike electricity, it can be stored for future use. Our results below would show that satiation is reached at twelve hours per day.

There is a positive relationship between electricity supply and rent. Moreover, the function has a convex shape. This reflects the fact that electricity is constantly demanded and rent is an increasing function of electricity supply, people a willing to pay a premium for 2h-hour electricity supply. Electricity consumption in Nigeria has risen in the last three decades due to increase in population without commensurate increase in infrastructure, increase in use of domestic electrical appliances, including air cooling devices, which are used day and night. It is also required to power tools and machinery by both small and medium scale businesses.

The satiation/turning point for hours of water supply is estimated using the following formula.

$$R = \alpha + \beta_1 W + \beta_2 W^2 \tag{6.4}$$

$$W^* = -\frac{\beta_1}{2\beta_2} \tag{6.5}$$

$R = rent$

$W = Water\ Supply$

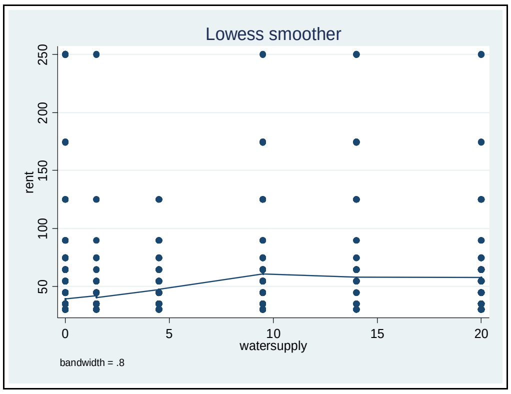

Figure 6.1a

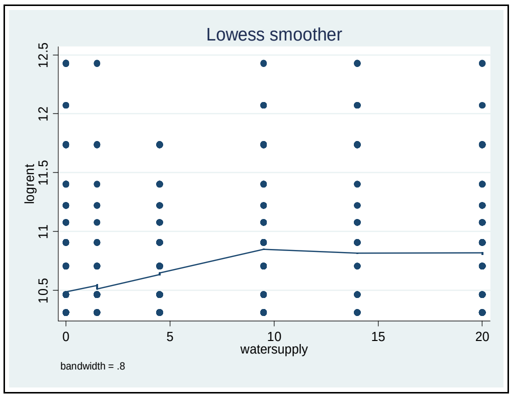

Figure 6.1b

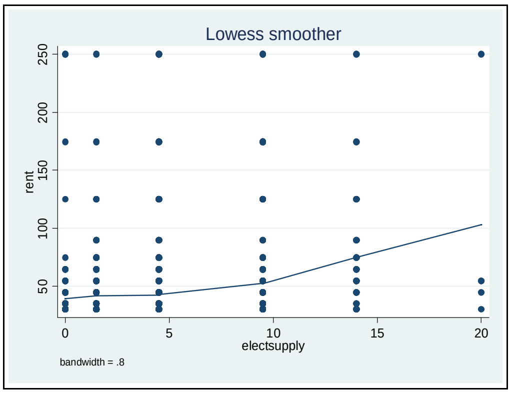

Figure 6.2a

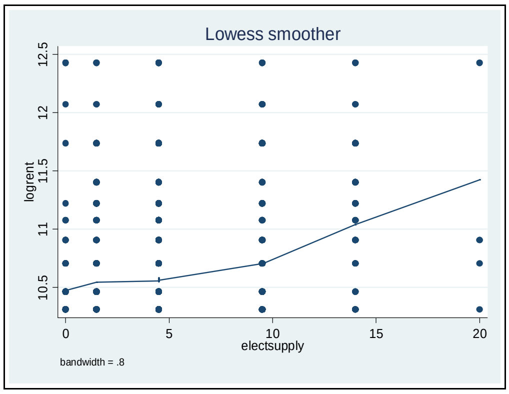

Figure 6.2b

6.6 Econometric Results

We expect a positive impact of utilities on rent. This is because public utilities and residential housing are joint-demand goods. Houses in location where there is efficient provision of utilities, all things being equal, are likely attract higher rent compared to areas where these facilities are poor or non-existent, utilities like water supply, electricity, waste disposal, recreation centres, parking spaces, and street lighting. We expect locations with higher negative externalities, *ceteris paribus*, to command lower rent. In this case we include neighbourhood attributes such as schools, market, highway, industries and airport and the environmental “cost” associated with living close to these attributes such as quietness, noise and industrial pollution.

We undertake a formal (interval) regression analysis to estimate (among other effects) the roles of water and electricity supply in the determination of rent.

Results from spatial analyses, such as this one, are often affected by spatial correlation. Proximity and adjacency of the dependent variable usually affects and are sometimes missed in econometric analysis. There is a rich literature on this problem in economic geography in general, especially in gravity models. Porojan (2001), Pandit and Laband (2007)

There are several formal tests for spatial autocorrelation, the most popular tests are Moran I; Geary's C; Ripley's K; and Join Count Analysis. Moran I, the most popular test, is a weighted correlation coefficient that is used to detect departures from spatial randomness. It is applied to locations with continuous variables associated with them, computes and compare the value of the variable at any one location with the value at all other locations. Another way to test for spatial autocorrelation is a Hausman type test, i.e. to estimate the model with and without cluster -robust standard errors (with clusters defined at the level of location) and compare.

If spatial autocorrelation is detected, the standard errors are wrong and should be adjusted by estimating cluster robust standard errors. This procedure is useful because this adjustment affects the level of significance/confidence of a particular variable in the hedonic pricing model (but it does not affect our estimated coefficient, WTP and the welfare change).

We therefore estimate two of our models twice, with and without cluster robust standard errors in order to solve for spatial autocorrelation which, as pointed out above, is common for most cross-section studies. Other two hedonic pricing models could not be estimated with (cluster) robust standard errors because they contain "area" dummies and the same "area" definition is used for the standard error correction.

Initially the most of the standard error disappeared when we estimate the model with cluster standard errors. We had to use bootstrap cluster standard errors, using 50 replications to properly estimate our models. This method is supported in theory and there is precedence in empirical panel data studies (Guan, 2003; Cameron et al, 2008).

Results from unadjusted models are presented in Tables 6.2 and 6.3 for models adjusted for spatial autocorrelation using cluster robust standard errors. Full computer output of these results and welfare change analysis “do file” are presented in Appendix I.

Our model selection is guided by existing literature and economic theory, in terms of the included variables – structural characteristics, public utilities and neighborhood attributes, and *a priori* results. As pointed out above, we include a quadratic water supply variable to capture the diminishing marginal utility, and the possibility of a satiation point, with respect to this attribute.

Although we estimated six different models - Linear and log-lin models with and without area dummies and two models without area dummies but adjusted for spatial autocorrelation - all reference to results in this chapter, unless clearly otherwise clearly stated, refers to models 1 and 2, summarised in Table 6.3. These are interval regression results from the linear and log models. We chose these models because they are adjusted for spatial autocorrelation.

The magnitude of the results for neighborhood attributes, water and electricity supply are basically the same for all the models.

Out of the four physical/structural attributes included in the linear model, three are strongly significant. This means that these residential housing attributes exert positive influence on rent in the study area. The size of a garden/courtyard is positively related to rent but its coefficient is not statistically significant.

	Housing Characteristics	Model 1 Linear Dependent Variable	Model 2 Log Dependent Variable	Model 3 Linear Dependent Variable	Model 4 Log Dependent Variable
Intercept		-142.819 (11.695)	0.718 (0.174)	-125.557 (11.227)	0.9305 (0.1678)
Physical/structural Characteristics	Number of bedrooms	29.042 (3.390)***	0.481 (0.051)***	27.366 (3.250)***	0.4570 (0.0485)***
	Number of bedrooms squared	-3.718 (0.468)***	-0.061 (0.007)***	-3.358 (0.448)***	-0.0563 (0.0067)***
	Number of toilets	15.094 (1.666)***	0.241 (0.025)***	12.812 (1.600)***	0.2091 (0.0239)***
	House floor area	0.030 (0.004)***	0.001 (0.000)***	0.022 (0.004)***	0.0004 (0.0001)***
	Garden/Courtyard	0.005 (0.004)	0.000 (0.000)	0.004 (0.004)	0.0000 (0.0001)
Public Utilities	Electricity supply	1.019 (0.267)***	0.015 (0.004)***	0.626 (0.259)**	0.0094 (0.0039)**
	Water supply	5.817 (0.479)***	0.097 (0.007)***	5.746 (0.458)***	0.0961 (0.0068)***
	Water supply squared	-0.236 (0.025)***	-0.004 (0.000)***	-0.237 (0.024)***	-0.0040 (0.0004)***
Neighbourhood Characteristics	Private primary/nursery schools	10.896 (1.898)***	0.171 (0.028)***	9.256 (1.857)***	0.1398 (0.0277)***
	Public primary schools	-0.953 (1.884)	-0.014 (0.028)	2.729 (1.872)	0.0445 (0.0279)
	Market	-3.510 (1.989)*	-0.043 (0.030)	-1.086 (1.935)	-0.0141 (0.0289)
	Highway	2.421 (2.104)	0.058 (0.031)*	-2.544 (2.099)	-0.0214 (0.0313)
	Airport	-8.920 (3.058)**	-0.143 (0.046)***	-	-
	Industries	-7.911 (2.854)**	-0.118 (0.043)***	-	-
Area Dummy ("City" is the base area)	Low Density	-	-	36.205 (3.051)***	0.5350 (0.0456)***
	Industries 1	-	-	7.809 (5.439)	0.0553 (0.0769)
	Industries 2	-	-	10.767 (3.306)***	0.1549 (0.0476)***
	Airport	-	-	2.952 (3.387)	0.0511 (0.0506)
	Other	-	-	12.955 (2.311)***	0.2104 (0.0345)***
Log likelihood		-3535.267	-3240.596	-3466.478	-3176.619
McKelvey and Zavoina R^2		0.184	0.205	0.218	0.236
LR Chi^2		787.223 (14)	891.57 (14)	924.80 (18)	1019.52 (17)
Number of Observations		2272	2272	2272	2272
Ancillary Statistic/lnsigma		3.5730 (0.024)	-0.622 (0.024)	3.522 (0.024)	0.511 (0.0120)

Table 6.2 Interval regression results (with non-robust standard errors) – Dependent Variable Annual Rent, or Log of Annual Rent

See Table 6.1 for the rent intervals in Naira per annum
Non-robust standard errors in brackets (except for LR Chi^2 - df in brackets)
* Mildly Significant (ρ <0.10) ** Significant (ρ <0.05) *** Strongly Significant (ρ <0.01)

	Housing Characteristics	Model 1 Linear Dependent Variable ♠	Model 2 Log Dependent Variable ♠
Intercept	**-**	-142.819 (29.299)	0.718 (0.428)
Physical/structural Characteristics	Number of bedrooms	29.042 (8.575)***	0.481 (0.126)***
	Number of bedrooms squared	-3.718 (1.277)***	-0.061 (0.020)***
	Number of toilets	15.094 (2.149)***	0.241 (0.038)***
	House floor area	0.030 (0.007)***	0.001 (0.000)***
	Garden/Courtyard	0.005 (0.004)	0.000 (0.000)
Public Utilities	Electricity supply	1.019 (0.298)***	0.015 (0.004)***
	Water supply	5.817 (1.323)***	0.097 (0.022)***
	Water supply squared	-0.236 (0.060)***	-0.004 (0.001)***
Neighbourhood Characteristics	Private primary/nursery schools	10.896 (4.107)***	0.171 (0.053)***
	Public primary schools	-0.953 (3.081)	-0.014 (0.058)
	Market	-3.510 (3.967)	-0.043 (0.064)
	Highway	2.421 (3.693)	0.058 (0.048)
	Airport	-8.920 (5.648)	-0.143 (0.072)**
	Industries	-7.911 (5.467)	-0.118 (0.052)**
Area Dummy ("City" is the base area)	Low Density	-	-
	Industries 1	-	-
	Industries 2	-	-
	Airport	-	-
	Other	-	-
Log likelihood		-3535.267	-3240.596
McKelvey and Zavoina R^2		0.184	0.205
LR Chi^2		787.22 (14)	891.57(14)
Number of Observations		2272	2272
Ancillary Statistic/lnsigma		3.5730 (0.184)	-0.622 (0.115)

Table 6.3 Interval Regression Results – Dependent Variable Annual Rent, or Log of Annual Rent - Spatial Autocorrelation Adjusted Standard Errors

See Table 6.1 for the rent intervals in Naira per annum
Robust Standard errors in brackets (except for LR Chi^2 - df in brackets)
♠ Robust/spatial correlation adjusted standard errors.
* Mildly Significant (ρ <0.10) ** Significant (ρ <0.05) *** Strongly Significant (ρ <0.01)

We also included six neighborhood attributes. Out of these attributes, only proximity to private primary/nursery school is statistically significant. Proximity to public schools, market, and highway, all appear to have the expected signs but are not significant. While living close to industries, airport and market are found to be negatively related to rent, in highly congested developing country city, the tenth most populated city in Africa, living close to highway has a positive coefficient. This means highway provides easy access to other parts of the city and serves as a gateway to other parts of the country. However, it is not statistically significant.

The results indicate that areas close to private schools attract at least 17 percent more rent than other areas in the city. Of course, we cannot be sure about the direction of causality in respect of proximity to private schools. It is not clear whether households are attracted to areas with private schools which increase demand and rent for these areas, or there is some kind of "sorting" going on according to income, where private schools are established in areas where individuals could afford to pay for them.

What does the area dummies tell us? Looking at models 3 and 4 (Table 6.3), our results indicate that, relative to the "old-city", households are willing to pay extra to reside in all but two other parts of the city. "Low density area" is clearly the most desirable, followed by the second industrial area (Bompai estate) and "other". To estimate models with area dummies, we had to drop (proximity to) airport and industries, as a neighborhood attribute, in order to avoid perfect collinearity.

From our dataset (see figs 6.1 and 6.2), and from the regression results, we may infer that electricity supply has a convex effect, while water supply has a concave effect. The results seem to imply that, while welfare is maximized with 24 hours of electricity supply, welfare reaches a maximum when water provision is only around twelve hours. This means that there is a decreasing marginal benefit from hourly increase in water supply.

In all our six models, water supply is strongly significant in the both linear and quadratic variables. As expected, the water supply linear variable has a positive coefficient, while that of the quadratic term (water supply squared) is negative. Electricity supply has a positive and significant effect at 99 % level of confidence.

When it came to electricity supply, we tried to capture the convex shape seen in Figures 6.2a and 6.2b by including electricity supply squared. Unfortunately, this inclusion had the effect of distorting the results, and we exclude it from our final models. As it happens, we are happy with a linear specification for electricity supply, for the simple reason that we do not expect there to be a satiation point for this attribute: there is demand for energy 24 hours of the day, partly due to the need for air conditioning through the night.

From our results in model 2, we establish that, in proportional terms 1 hour increase in water supply would attract 2.3 per cent increase in rent. In absolute terms, from our estimates in the linear model (model 1), the *upper bound* to the total welfare benefit resulting from an improvement in provision of extra hour of water supply (for a period of one year) is 3,459 in local currency. This is approximately four times the daily wage of a manual labourer. This figure is the point estimate of value of water. This analysis is based on the assumption that changes in individual WTP is a proxy for the utility/welfare improvement derived from increased supply of public utilities.

We computed the point estimate of the *upper bound* to the total welfare benefit of increased water supply by taking the partial differential of the hedonic pricing estimated equation with respect to watersupply.

$$P = \alpha + 5.817ws + 0.236ws^2 \qquad (6.6)$$

$$\frac{dP}{dw} = 5.817 + 2(0.236)ws \qquad (6.7)$$

Mean hours of water supply is roughly 5 hours.

$$\frac{dP}{dw} \mid mean = 5.817 + 0.472(5) \qquad (6.8)$$

We use the following Stata command:

$$nlcom\ \_b\left[hourswater\right] + 2 * \_b\left[hourswater2\right] * 5$$

Coefficient	3.4592 (0.6908)
Confidence Interval	2.105304 ⟷ 4.813039

Table 6.4 Point Estimate of Water Supply from Interval regression Results

We estimated this in Stata using the delta method so as to obtain estimates, standard errors and confidence interval.

For an hour of electricity supply, the estimated *upper bound* to the total welfare improvement resulting from an improvement in provision is 1019 - about 30 percent of the same for water. It therefore appears that water is valued more than electricity. This is broadly consistent with results from the discrete choice models estimated in Chapter 7.

STATA does not compute an R^2 for interval regression. Numerous alternatives are available in the post-estimation procedures. In Table 6.3, we report the McKelvey-Zavoina pseudo-R^2, which indicates the explanatory power of the predictor variables. The pseudo-R^2 is 0.18 and 0.21 for the linear and log models respectively. These are acceptable, given that we do not expect high R^2 in housing location studies because of the multitude of factors that affect both the rent, individuals' valuation of residential housing services and individual choice decisions. Veall and Zimmermann (1996) assert that, it is generally common to expect low R^2 in microdata-based studies.

6.7 Welfare Change Analysis

In Section 4.8, we reviewed the theoretical literature on welfare change analysis, focusing on how coefficients from hedonic pricing studies may be interpreted in terms of welfare change. In this section, we conduct a welfare change analysis based on the results of the hedonic pricing model estimated in this chapter.

Under present financial circumstances and high demand for public utilities, it is not possible to provide every household with twenty-four hours' of water supply. In order to achieve optimal utilisation of limited resources, we visualise two possible types of policy objective available to the policy maker in attempt to improve the water supply in Kano city. From these we could determine the best policy scenario, that is, the policy that generates the greatest increase in welfare subject to a cost constraint. As earlier pointed out, we will be treating changes in individual WTP as a proxy for the utility derived from increased water supply.

The first type of policy objective that we consider is to increase water supply for every household by a certain (absolute) number of hours per day. The second is to set a minimum acceptable number of hours per day, and then to ensure that every household is brought up at least to that level. We will suppose that both types of objective are feasible, but it is beyond the scope of this research to propose exactly how such policies might actually be implemented; that is a matter of engineering and design. We simply make a straightforward assumption about the costs of provision. Also, we are unable to specify a particular time of the day for the minimum number of hours; this is because our research is based on revealed preference data. This should come out of a stated preference survey in Kano.

For reasons that shall become clear, use the results from model 1 (see Tables 6.2 and 6.3), which represents a "pre-improvement hedonic" in which hours of water and hours of water squared both appeared as explanatory variables. Let w denote hours of water supply, and let R denote rent in thousands of Naira. The results from model 1, obtained directly from Tables 6.2 and 6.3, may be represented by:

$$R = k + 5.817w - 0.236w^2 \tag{6.9}$$

where k is a constant. Note that (6.6) implies that rent is predicted to increase until hours of water supply reaches 12.3 hours, and then starts to fall. This is a plausible result: households are unlikely to attach importance to water supply outside waking hours. Since it is illogical for increases in supply to generate a *fall* in welfare, in what follows we shall disregard the downward-sloping part of the function. That is, we shall assume that welfare remains constant when the number of hours of supply is greater than 12.3.

Consider a policy that results in hours of water (for a given household) increasing from w_1 to w_2. Then the predicted change in rent may be computed as:

$$\Delta R = Max\left[\left(5.817w_2 - 0.236w_2^2\right) - \left(5.817w_1 - 0.236w_1^2\right), 0\right] \tag{6.10}$$

Note that in (6.10) we are simply applying (6.6) to the situation before and after the policy, and taking the difference. Note also that the role of the *Max* function is to rule out reductions in welfare.

Having computed (6.10), we appeal to the analysis of Bartik (1988) summarized in section 4.8 of this thesis. There, we arrived at the rule that (6.10) may be interpreted as an *upper bound* to the change in welfare resulting from the increase from w_1 to w_2. Table 6.5 shows how such an upper of each additional hour of water supply is computed for each policy. The STATA code used to carry out these computations is included in Appendix I.

Policy	***Change in Hours of Supply (Per Household)***	***Change in Welfare (Per Household) (Upper Bound)***	***Valuation Per Additional Hour ('000 Naira) (Upper Bound)***
1 extra hour for all	1	3.887	3.887
minimum 1 hour	0.49	2.73	5.571
minimum 2 hours	1.04	5.52	5.308
minimum 3 hours	1.65	8.33	5.048
minimum 4 hours	2.26	10.87	4.810
minimum 5 hours	2.94	13.36	4.544
minimum 6 hours	3.69	15.75	4.268
minimum 7 hours	4.44	17.98	4.050
minimum 8 hours	5.18	19.49	3.763

Table 6.5: Computation of welfare increase per additional hour of water supply for a range of policies.

In Table 6.5, we see that the "minimum 1 hour" policy is the best in terms of the change in welfare per additional hour of supply. However, this is simply a consequence of the hedonic function (6.9) being quadratic; it steepest between 0 hours and 1 hour of supply.

In order to consider which of the policies considered in Table 6.5 should actually be implemented, we need to consider the costs of provision. Here, purely for convenience, we make the straightforward assumption of constant marginal cost. That is, the cost of providing one additional hour of supply for one household is the constant c. To apply the concepts of cost benefit analysis, we would then compare the welfare increase per additional hour, given in the final column of Table 6.5, against *c*. Remembering that the numbers in the final column of Table 6.5 represent upper bounds, all we may say for sure (based on this analysis) is that any policy for which the number in the final column is *less than c* is definitely infeasible.

For example if *c* were very high at 6.0 (thousand Naira), then none of the policies in Table 6.5 would be feasible. However, if *c* were 5.0, the three policies "minimum 1 hour", minimum 2 hours", and "minimum 3 hours" would all become possibly feasible. If *c* were even lower, at 4.0, only one of the policies listed in the table would be ruled out as being definitely infeasible.

Note that the "1 extra hour for all" policy appears to be less beneficial than most of the "minimum *x* hours" policies, even in terms of welfare change per additional hour. This is because some household already enjoy high hours of supply, and, in accordance with (6.9), these households benefit little, or not at all, from additional hours of water supply. The "1 extra hour for all" policy is less effective because it increases the supply of all households, including those whose welfare improves little or not at all.

While the cost benefit analysis presented in this section is fairly crude, and relies on some quite strong assumptions, it is nevertheless useful as an illustration of how results from a hedonic pricing study may be applied in the evaluation of policies intended to bring about improvements in utility provision.

6.8 Conclusion to Chapter Six

Our results confirm that people attach some priority to having a supply of public water. This is derived from evidence of (household's WTP for) higher rent in areas with longer hours of water supply in Kano, Nigeria and the relatively high proportional contribution of extra hour of water supply to rent paid i.e. high rent elasticity in areas with higher water supply.

In the medium-term, we recommend increased investment by the government in the provision of water. This could be done in several ways, either through government-private sector initiative or through direct loan from private sources. Water project would be viable because of the high WTP amongst consumers which means that the price of water is well below consumer's reservation price. Consumers have also shown higher preference for water supply in comparison to electricity supply. Government should strive to provide minimum of four hours for every household this should be feasible, given available resources (a big river is few kilometres away from the city) and the fact that the satiation level there is less than 24 hours.

We are unable to make any recommendation for the long-term because, it is anticipated that, the problem could be solved in the medium-term with increased investment, better planning and controlled expansion of the city. Although it is possible to study the long term welfare benefits of this possible improvement in water supply, we do not intend to pursue this issue further because, Scotchmer (1986) shows that, even if a population is homogeneous, hedonic pricing model is not appropriate for analysing long-term benefits of a large scale public projects.

Chapter Seven: Discrete Choice Modelling of the Residential Location Decision

7.1 Introduction

In this Chapter, we apply the discrete choice models discussed in Chapter 3 to our housing location choice data described in Chapter 5. The objective is the same as that of Chapter 6: to estimate WTP for water and electricity supply. However, the econometric modelling strategy is very different to that used in Chapter 6. It will therefore be interesting to see how close the WTP estimates are between the two methods.

It is important to analyse socioeconomic factors in discrete choice models because they implicitly reflect heterogeneity of preferences. Several studies have reported different types of sorting along income, racial and other socioeconomic characteristics in residential location choice. It is possible to estimate the residential location choice probabilities using discrete choice models.

The Multinomial logit (or Multinomial probit) model(s) could be used to explain individual choice decisions in terms of socioeconomics characteristics. Alternative specific conditional logit (or probit) model(s) could be used to explain choices in terms of attributes of the location rather than characteristics of the individual. Mixed logit (or probit) model(s) analyse(s) choice decision using both the individual characteristics and the location attributes.

By “Mixed probit” we mean a Multinomial probit model that contains both household characteristics and attributes of alternatives. It is possible to estimate this model in Stata 11 and therefore to avoid the limitations of both Conditional logit and Mixed logit models.

In chapter 2, we summarised theories of land use pattern and the structure of the city, supply and demand theories of the housing market, in this chapter we analyse factors that could possibly influence residential location choice decision. Because this study is a cross section study, we assume that supply of housing and housing services to be fixed. We intend to look at housing demand, in particular residential location decision with a view to estimate the WTP for public utilities, water and electricity supply.

7.2 Residential Location Choice Decisions

Several factors which determine individual decisions on residential house location have been identified. Some of these factors include: travel cost and proximity to work; accessibility to other parts of the city, shopping centres and schools; quality of neighbourhood - local interactions and amenities; availability of public services; cost – house prices, rent and taxes; housing attributes - the number of rooms, types of appliances, gardens, garage etc; household socio-economic characteristics, age, gender, race, marital status, family size, education, nature/type of job, income/class status, gender of the head of households, number of adults in employment in the household, etc. Feridhanusetyawan and Kilkenny (1996), summarized these factors into five categories: workplace location; local amenities or "quality of life"; life-cycle and other personal characteristics; return to human capital accumulation, and; real costs of living.

Some of these factors especially income and housing price affects housing locations decision in the onset, because housing represents inflexible consumption which cannot be readily altered in response to price change (Turnbull et al, 1991). It has been argued that, decision to move depends on change in income, rent, preference for housing relative to other goods, condition of dwelling and the neighbourhood, accessibility, change of workplace, expected search and moving costs, but search and moving costs do not on their own, affect moving decision (Wong, 2002). And because, individual households cannot, before-hand, determine where houses are build nor the combination of their attributes, households are likely to look for alternative which satisfies preferences if searching is costly in the presence of imperfect information. This is called “the utility satisfying model”.

Some of the difficult issues to settle in empirical study of housing location decisions is the nature of tenure. Renting and own-house location decision must be treated separately, as earlier pointed out in section 2.4, the two are sometimes considered as substitute. Secondly, for the own-house case, housing as a durable asset provides both consumption and investment services and usually purchased with mortgage, which add extra parameter to the decision equation (Arnott, 1987). It is easier to handle data from first time house owners/renters. Moving decision presents econometric difficulties for standard choice decision models because the household decision about moving is conditional on having previously preferred the original location, (Bartik et al, 1992)

In a similar way, the result of choice decision study is likely to be sensitive to framing of questions. It might be easier to estimate a model that provides the value to individuals, of remaining at their current location rather than being coerced or enticed to move out. This way, it possible to observe whether (aside moving costs), individuals would relocate if they are dissatisfied with their current location. Bartik et al (1992) reported that, low-income households are willing to pay about 8% of their annual income to avoid being forced out of their current dwelling and that these "psychological moving costs" are higher for older and longer tenure households.

The question that confront each individual household in the consumption of housing service i.e. decision of where to live (for new residents) or where to relocate to (for existing residents) is, what combination of attributes and where to chose? Households must decide on the size of property suitable for their need, not only bigger or smaller but a given mix of attributes, subject to financial constraints. Some households would also have to reconcile between a house close to work or school, to buy or to rent. There are also several other positive and negative externalities to consider.

The numbers of income earners within the households affects housing expenditure and preferences. It has been shown that, since the Seventies, increase in women income through increased labour participation have influenced family size, marital status and housing demand. It has also been recognised that women are more likely to consider safe location because they are more risk averse and because they care more (in relative terms) about the safety of the family in general, and especially children (Skaburskis, 1997).

The elderly would prefer to remain in their current location and less likely to move than younger adults. This is because older adults are more likely to be homeowners who are less likely to move than renters. They are also likely to have an attachment to their current location, being more likely to have lived in the location for a long time. The elderly are supposed to be richer, with more education and, in industrialised countries, reach old age in better health.

There are two cause of relocations for the elderly, both voluntary and involuntary locations. Push factors such as change in financial and job situation due to retirement, changing geography of cities, decline in health of self or spouse and loss of spouse. Pull factors include, to be close to extended family, close to amenities such as nicer whether, better recreational and health facilities and lower cost of living commensurate with their pension and retirement plans (Krout and Wethington, 2003).

In this research we do not intend to go into the debate of who takes the final decision within the family. We assume a unitary model of household decision, and expect a household to be simply guided by its overall budget constraint in all decision making process. We adopt the unitary model also called the “single-agent”, “common preference”, “consensus”, “altruistic”, or “benevolent dictator” models because, it would be difficult to distinguish the incomes or the consumptions of individual family members in cross-sectional study where household consumption consists of aggregates. Moreover, it is assumed that if household members (spouses, children and other dependents), love, copy, or annoy each other, then they care about their own consumptions and the consumptions of other members of the family (Bergstrom, 1997).

We also expect the household consumption bundle to be considered as a whole, which means we expect the following consumption goods and their cost to enter the optimisation function: place of residence; cost of commuting to work and to school for children (measured by time and money spent); cost of commuting to the market; cost of basic food items and basic household items; utilities; leisure time, which includes commuting to and time spent with friends and family.

Looking at the households consumption optimisation function, housing - place of residence, plays a central role because it coordinates all other functions. It is therefore very critical for the household to choose where to live taking into consideration this pivotal role. Below we discuss the factors that households consider before making this choice. Because the factors are numerous we expect choice selection based on certain criteria/prioritisation on the part of the household, the decision maker.

Walker and Li (2007) argued that lifestyle, defined as “deep-rooted and embedded, prevalent attitudes indicating preferences towards a particular way of living”, is a key driver of the decision of where to live. But the concept “lifestyle” is as vague and general as the concept of individual household. Solomon and Ben-Akiva (1983) outline three different roles of the individual: formation of a household; supply of labour; and consumption of leisure subject to constrained resources.

Individuals either sell their labour as a composite commodity or use their time to produce goods for sale to earn income. This has been the trend since subsistence is replaced by exchange, division of labour and specialisation became diffused in modern economies.

Structure of, and access to other parts of the city affects tenure decision and location choice because depending on the physical structure of the city, certain options are only available is some parts of the city. The type and network of transport system, landscape, location of employment centres, recreation facilities, flow of air/wind are some of physical attributes of the city that would influence households choice. Individuals would choose to live in place that provides easy and quick access to their places of work/business. It is important to consider the cost in money terms of commuting – transport fares, travel time, loss of working/business time due to delays – and available modes of transport within a particular city. Transport mode decision is also determined by age, income, weather conditions, and physical and health status of the individual.

Longer travel time directly affect utility by reducing leisure time and it may reduce income due to higher community cost, and loss of productivity, leaving less to spend on housing and other consumption goods.

Kraus (2006) argues that, it is this simple trade off between desire to have more space, minimise travel cost which produce reasonable prediction of location choice. However, we also expect high price near schools, park, coastlines and other amenities. Irrespective of whether a city is monocentric or polycentric, these amenities would generate local peak in the demand function.

The structure of the city makes some parts more desirable either because of serenity, quite, flow of fresh air, topology (highland and valleys), settlement density and provision of public goods. These attributes could be due to policy design of planning authorities or due to providence - natural structure of the city. Because these amenities are normal goods, they affect the value (price) of properties in a particular city and create what we termed the economic status/peer-group effect. Only the rich (defined in relative terms) can afford to pay for property that has more of these amenities. It is common to find a particular city segmented along class difference where the poor is more likely to be found in slumps and high-density areas. This further entrench the ethnic and racial composition of the city because some racial and ethnic groups are more successful than others (Oliver and Shapiro, 2006).

In addition to proximity to work place, households with school age children also consider living close to "good" schools. Living close to schools could reduce the cost of school-run for the household. In some countries public school places are determined by catchment area, where this applies "good" schools attracts households with children. However, there are cases where living very close to school is detested because of negative externalities that arise from unruly behaviour of students/pupils. We therefore expect exponential relationship between proximity to schools and individual household preference for this "amenity".

When looking for a residential house, households would not search the whole city, they are more likely to be guided by experience of, or assisted by, work colleagues, family members and friends. This arises because of heuristics on the part of the individual household, information and search costs. This is more likely to happen in the case of own-house location than renting which is relatively more temporary. This type of social network creates settlements along economic class, race, ethnic lines etc.

Because individual households are not islands, they interact with other households, especially relatives and close friends, we expect extended families to live close to each other.

Public “goods”, such as infrastructure/amenities, utilities, sanitation, cleanness, quiet, attract households who can afford them, while environmental “bads” (in some cases, socially defined) are expected to drive off the financially comfortable households. Although we expect all the aforementioned factors to affect location choice decision, we are interested in the impact of the provision public utilities in attracting households. Where there is spatial variation in supply of utilities, households are likely to vote with their feet by relocating, or be willing to pay a higher premium for improved services.

7.3 The Research Problem: Public Utilities Provision in Kano

We are interested in studying the impact of the level of provision of public utilities, mainly water and electricity supply in location choice decision. How does this relate to household’s decision on where to live? Because it is possible to view the supply of public utilities, in this case a reliability index, as observed choice attributes, we could use appropriate discrete choice econometric models to estimate the WTP for these utilities.

Another question (partly introduced in Chapter One) is what is the state of supply of public utilities in Kano? Kano state is endowed with natural sources of water supply. There are three major sources, a major river which pass very close the city, with its source in the Niger republic, natural reservoirs in valleys close to the city creating natural dams and numerous artificial dams built primarily for irrigation and domestic water supply. These three sources ensure all the year supply of water in spite of the fact that Kano is located in a semi-arid climatic region.

Pipe-borne domestic water is supplied by a government owned company. The company enjoys support from government but is suppose to operate as a commercial entity that is expected to recover its recurrent expenditure, while the government provides capital expenditure. National government provides indirect support through it’s of finance irrigation projects which are supposed to increase and improve the various sources of water supply. It is important to point out that given abundant fresh water supply sources and state of technology, at present, waste water recycling is not a considered option.

Although the water company has no supply problem there is a problem with water treatment and distribution within the city and other communities in the state. This could be attributed to two problems. First, massive unplanned/uncoordinated expansion of settlements and the illegal connection by households to the water supply network. This makes demand projection very difficult. Second is the problem of erratic public power supply which affects the pumping of domestic water within the city. It takes hours after power cut before pumping station could restore the tempo that would ensure supply to all parts of the city and that is difficult to achieve when the power supply is inconsistent. Related to these problems is the reluctance by households to pay their water bills and the problem of water revenue vs water supply becomes a “chicken and egg” parable.

Individual households seek for alternative sources of this vital commodity. These alternatives depend on economic status of the house provider, population density of a given area, and water table and other physical properties of the land on which a particular property is built. In theory, it is possible to have boreholes in rented houses especially in the low density area. This is very rare. Boreholes are more likely to be found in own-house accommodation because it is expensive to construct and needs regular servicing which could significantly add to rent. There is also the issue of feasibility and viability of the borehole.

The most common alternative for public water supply in Nigerian cities is the purchase from water vendors who collect water either from a public water supply in another area, from a traditional well or a motorised bore-hole, which could be free or pay-per-use. The obvious problem with this source of water is its quality and high cost. Water supplied by vendors is generally unhygienic and expensive relative to tap water. In addition to poor sanitation, it is estimated that, in Nigeria, households purchase water from water vendors at a cost of up to 12 times amount being paid by households with public water supply (Ariyo and Jerome 2004).

Some houses, depending on physical space, generosity of the house provider, have traditional well, which have varying levels of yields and water quality. Traditional wells generally have low yield during the dry season. Other problems of open-well are it takes space, danger of drowning for kids (and the mentally challenged), and fear of contamination with septic tanks. Only two of Nigerian cities currently have central sewage system.

These factors combine to raise the attractiveness of public water supply. Although in some places, due to illegal water pipe connection by some household and leakages in the pipe network, the quality of public water supply is compromised.

Electricity supply is very crucial especially because of the large percentage of people working in the small scale enterprise sector who require energy for their machines. It is also essential for domestic use such as cooling and refrigeration, given the hot temperature in the city. The relative humidity in Kano range between 23% in March to 83% in August and the annual average temperature is 25.9 and a range of ($\pm$) 15.5 degrees Celsius (Maconachie, 2007).

This high demand for energy is met through the use of electric generators. Nigeria is arguably one of the countries with highest per capita electric generators, with most households that could afford buying different types, ranging from low to high capacity, silent and noisy, branded and locally fabricated generators.

Public electricity is cheaper and cleaner source of energy. Generators are noisy and are known to supply irregular electricity voltage and frequent surge which could damage equipments. Another problem with personal electricity generators is the fuel crisis in Nigeria. Although Nigeria is leading producer of crude oil in Africa and the sixth largest OPEC producer, fuel is always scarce and expensive to buy due to low refining capacity in the country. There is always the risk of domestic fuel storage. Even when fuel is available it is sometime adulterated with other substance.

The manufacturers association of Nigeria recently lamented the cost of doing business in the country and reported that about half of Nigerian households own private generating set, spending about $13.35 annually on fuel. The manufacturer’s umbrella lamented that, because virtually all industrial and commercial enterprises in Nigeria own electric generators, electricity has therefore become the most important infrastructural problems for businesses in Nigeria today (Nigeria Vanguard 2009).

In summary, the alternative to public water and electricity supply is very limited. It also means, individual WTP for both public water and electricity supply are by per greater than government determined tariff.

The data we have is a proxy for reliability index, measuring hours of water and electricity supply. There is small problem with the water supply data because it is possible to store water when the tap is running. Individuals would choose to live in area where it is possible to obtain public water supply and then decide whether to collect water for future use or not. Our response to this is that, since this option is available to all households in a given area, it does not significantly affect our results.

7.4 Relevant Previous Studies

Discrete choice model is used under two circumstances. If the dependent variable is qualitative in nature, and when it is convenient to categorise a continuous dependent variable. Examples of discrete choice models for continuous dependent variables are round-off replies, with data collected in ranges/intervals rather than exactly. Researchers sometimes use the discrete choice framework to analyse a continuous dependent variable because it is plausible to argue that, in certain situations, categorical data may be more reliable than continuous data (Borsch-Supan, 1987).

Discrete choice models have been applied to the study of disaggregated models since the ground-breaking works of Muth, Alonso and McFadden. Discrete choice has been used to study individual choice behaviour. Most prominent area of discrete choice application is transport mode choice and the choice of itinerary, examples are the numerous work of McFadden, and several others (Asensio, 2002). Other areas include residential housing location choice (McFadden, 1978), Borsch-Supan, (1987), Gabriel and Rosenthal (1989); recreational demand, Hanley et al (2001); marketing, Franses and Montgomery (2002), Anderson et al (1992) Adda and Cooper (2000); portfolio choice, Ramaswamy (1997); and labour supply, Labeaga et al. (2005), van Soest, (1995), Kornstad and Thoresen (2007).

The application of discrete choice model requires some simplifying assumptions. Assumption about the decision maker and her socio-economics characteristics; the choice set – alternatives; attributes of alternatives; and the decision rule (Ben-Akiva and Bierlaire, 2003).

The decision maker is assumed to be an “individual” who is rational and has full knowledge of all feasible alternatives. We shall consider household as a unit, and we shall overlook the internal decision making process within the house. Although it has been argued that the context in which a decision is made is an important determinant of an outcome (Swait et al, 2002), we assume a smooth and fair process of arriving at decision within the household. We assume that, either the household is headed by a “benevolent dictator” or as Cai (1989) put it, husbands, wives and children are altruistically linked. Sugden (2000), Brewer and Gardner (1996), Bacharach (1999), Adamowicz et al (2005) and Basu (2006), have provided detailed analysis of the complex nature of group decisions.

The decision maker has finite set of alternatives which will be explicitly listed. In a residential house location study, it is anticipated that, in some cases, an initial decision would have been taken. The decision maker is therefore aware of what has been chosen and other alternatives that have not been chosen. Five locations, selected based on their attributes will be covered in the study area for this research. In arriving at these “reduce set” we know that, it is possible the decision maker is aware of the universal set and excluded alternatives which could affect her decision. We therefore decided to include sixth alternative (other), in order to cover all possible alternatives. The decision maker will be expected to evaluate contemporaneously, these six locations base on their attributes. Responses collected through a researcher administered questionnaire are used to estimate probabilities and WTP.

Each location (alternative) is characterised by a set of attributes. These includes, noise pollution, air pollution, municipal waste, road network/access to other parts of the city, social interaction, economies of scale, cost of housing, availability of municipal services - water and electricity, good schools and personal security. These are attributes that could, ***ceteris paribus***, affect individual’s residential location decision. We have tried to avoid generic attributes in delineating our alternative locations. We intend to clearly define these attributes to avoid uncertainty in modelling problems and the elicitation of responses from decision makers. For example noise pollution could be traffic noise or aeroplane noise and could be day and night or a day time phenomenon.

The last assumption is the decision rule, which will be based on random utility model. We intend to show how neo-classical utility theory is inadequate to explain choice decision in disaggregated models and in the presence of heterogeneity amongst decision makers. It is the socioeconomic characteristics of the "decision maker" that will help us to capture heterogeneity in decision amongst individuals (Ben-Akiva and Bierlaire, 2003).

In disaggregated models studies, the underlying theory of individual choice behaviour is characterised as descriptive – how individual behave rather than how they should behave; generalisable; and operational, derivable from real data (Antonini, 2005).

The main objective of this component of the thesis is to analyse the consumption behaviour of a heterogeneous consumer (the household) for a differentiated good (residential house). Therefore, the observational unit of this research is the individual household, who has chosen dwelling from six different locations based on household's perceived characteristics of the six locations. The six locations, which constitute the choice set are clearly defined areas of "the city" comprising of: the ancient part of the city; areas surrounding the airport; areas close to two industrial estates; a low-density suburb; and for completeness, all other parts of the city.

We use five discrete choice models (alternative specific conditional Logit, Mixed-Logit, nested logit, alternative specific multinomial probit and mixed probit models) to analyse the data. But, why is ordinary least squares (OLS) regression insufficient to estimate the utility model of a characteristics-based choice decision.

McFadden (2003) explains why conventional econometric analysis is inadequate when economic variable is discrete. He argues that, when economic behaviour is expressed as a continuous variable, regression model is often adequate to describe the impact of economic factors on behaviour, or to predict economic behaviour in altered circumstances. This is true even when the behavioural response is limited in range or integer-valued, provided these departures from an unrestricted continuous variable are not conspicuous in the data, so that round-off of the dependent variable to an integer is negligible relative to other random elements in the model. However, conventional regression analysis is not feasible when behaviour is expressed as discrete variable. Although, data need not to be an integer, for example the outcome of a toss of a fair coin, which could be either head or tail, for computational reasons, it is in most cases presented as numbers.

In addition to the structure of the data, there is a problem with theoretical assumptions. While the neoclassical economics assume utility maximisation, in a differentiated goods market, for example residential house location choice decision, each individual will have indirect utility function conditioned on location that gives payoff to choosing a particular location. This decision depends on prices and income by the individual but, it also contain factors such as taste and perceptions and unmeasured attributes which are unobserved and from the point view of the analyst are random but, which could be heterogeneous in taste and perception not captured by the conventional economic theory.

In the homogenous good case, it is possible to estimate demand parameters in the presence of unobserved factors using instrumental variables. However, in the case of differentiated goods both observed and unobserved product characteristics enter the demand equation in a non linear form. This makes it impossible to use the instrumental variables method (Berry, 1994).

If the outcome of an experiment produces a discrete random numbers, conventional parametric method of data analysis such as OLS can not be used to explain causal relationships.

MacFadden (1984) argues that, when the number of alternatives is large, response probability models may impose heavy burdens of data collection and computation. However, he also proves that, the structure of discrete models permits a reduction in problem scale by either aggregating alternatives or by analyzing a sample of the full alternative set.

We intend to estimate the discrete choice regression and using the results estimate WTP for public water and electricity supply.

We make following assumptions - similar to Aufhauser et al (1986) - about the housing market in the study area.

a. each house is a multi-attribute commodity characterised by physical, neighbourhood and locational attributes.
b. location decisions are made by the household within a typical “neo-classical unitary model” of the household which envisage smooth decision making process within the household.
c. “the household” is disaggregated according to socio-economic characteristics, age, income, size, school age children, occupation, education etc.
d. household residential location decision is determined by housing preference changes which is affected by several factors, namely cost of accommodation, level of pollution, law and order, local business activities, quality of roads network, cost of commuting to work, schools in the neighbourhood, social interaction, provision of public utilities/municipal services etc,
e. different types of houses exist in the residential housing market. Rental housing, public housing, employer provided housing, shared houses, privately rented, outright owned and mortgaged owned houses.
f. all houses in the city are rented, either by owners or in the form of private letting. For reasons explained in Chapter 5, our research is restricted to the sub-population living in privately rented housing.

Finally, the only discrete choice water demand study that we came across is Nauges and Strand (2007). They estimate non-tap water demand in three cities in El Salvador and one city in Honduras. Using a combination of Multinomial logit model and OLS, six discrete alterative sources of water for domestic consumption namely: private tap; public tap; private well; public well; truck; and other, they find non-tap water demand elasticity with respect to total water cost of between -0.4 to -0.7.

Our study is the first to report estimates from a combination of Alternative specific “Conditional” probit and “Mixed” probit models.

7.5 Model Specification

We will be using the alternative specific (conditional logit and probit), nested logit models and the mixed (logit and probit) models. The discrete choice model was discussed in detail in Chapter 3. The specification of the ASCPM and ASMPM is:

$$y_{is}^{*} = z_{is}'\alpha + u_{is} \quad s = 1, \cdots, S \;\; i = 1, \cdots, n \tag{7.1}$$

And the specification of the nested logit and mixed (logit and probit) model is:

$$y_{is}^{*} = x_{i}'\beta_{s} + z_{is}'\alpha + u_{is} \quad s = 1, \cdots, S \;\; i = 1, \cdots, n \tag{7.2}$$

We can be more specific. The three alternative-specific attributes that we use are: rent (r), water-supply (w), and electricity-supply (e). So the ASCPM/NLM/ ASMPM becomes:

$$y_{is}^{*} = \alpha_{1} r_{is} + \alpha_{2} w_{is} + \alpha_{3} e_{is} + u_{is} \quad s = 1, \cdots, 6 \quad i = 1, \cdots, n \tag{7.3}$$

And we expect $\alpha_1 < 0; \alpha_2 > 0; \alpha_3 > 0$. The mixed logit model is the same, except that it controls for characteristics of the individual (collected in x_i) such as income, age and years of education.

In Chapter 3, in the context of travel mode choice, we explained how the value-of-time could be deduced from the ASCPM/NLM/ASMPM estimates. Here, we use the same ideas to estimate the WTP of water and electricity supply. Consider the utility function defined over rent and water supply.

In Figure 7.1, we see that in order to obtain the WTP of water, we need to take the ratio of the water coefficient in the ASCPM, to the rent coefficient. That is:

$$WTP_{water} = -\frac{\alpha_2}{\alpha_1} \tag{7.4}$$

Similarly, the WTP for electricity is:

$$WTP_{electricity} = -\frac{\alpha_3}{\alpha_1} \tag{7.5}$$

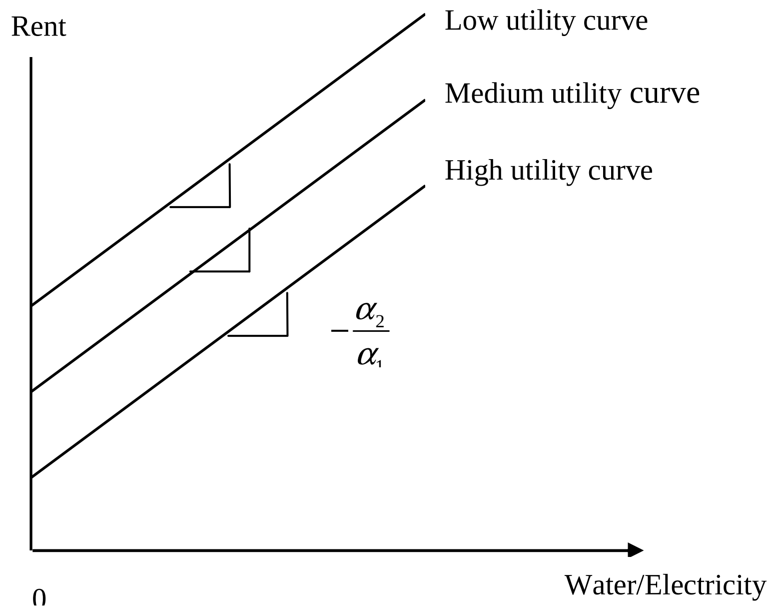

Figure 7.1: Indifference Curves over water/electricity and rent

When we have obtained estimates of the parameters of any of the three choice models, we simply apply formulae (7.4) and (7.5) to these estimates in order to deduce the required WTPs. Note that standard errors for the WTPs are obtained using the delta method (Greene, 2003) applied using the ***nlcom*** command in STATA.

7.6 Counterfactual Data

As explained in Chapter 3, in order to estimate ASCLM/NLM/ASCPM or mixed logit/probit model, we require attribute data (rent, water, electricity) on all areas. This presents a problem, since the questionnaire only obtains such data on the chosen area.

We need to estimate the "counter-factual" attributes, that is, the levels of attributes that each individual would experience in each of the 5 locations that they did not choose. For this purpose, we exploit sample information in the following way. Table 7.1 shows the average attribute levels for each of the six areas. Let us denote these as:

$$\bar{r}_s,\ \bar{w}_s,\ \bar{e}_s \quad s = 1, \cdots, 6 \tag{7.6}$$

Let us consider an in individual who has chosen area 1, so that their experienced attributes are:

$$r_1,\ w_1,\ e_1 \tag{7.7}$$

To obtain the counterfactuals for this individual, we simply apply the rules:

$$\tilde{r}_s = r_1 + (\overline{r}_s - \overline{r}_1)$$
$$\tilde{w}_s = w_1 + (\overline{w}_s - \overline{w}_1) \quad s = 2,\ldots,6$$
$$\tilde{e}_s = e_1 + (\overline{e}_s - \overline{e}_1) \qquad (7.8)$$

Area	***Description***	***Average Rent***	***Average Hourly Water Supply***	***Average Hourly Electricity Supply***
1	Old City and its fringes	36.6789	6.4276	4.6791
2	Low Density/Suburb	69.0876	7.3574	6.2798
3	Close to Airport	36.2879	1.4892	2.8750
4	Close to *Sharada* Industrial Estate	47.3732	15.7113	5.6143
5	Close to *Bompai* Industrial Estate	38.5243	2.1367	4.5761
6	Other	42.0071	2.9190	4.3528

Table 7.1: Average attribute levels of the six locations (alternatives)

7.7 The Sample Selection Problem

The sample selection problem arises in choice based samples when the sample is based on choices rather than exogenous characteristics of the decision makers. This is sometimes a problem in probabilistic choice models where it is the choices/alternatives that are used to design the sample and data collection. This problem leads to inconsistent and biased estimates of certain parameters (Heckman, 2008; Manski and Lerman, 1977; Nevo, 2003).

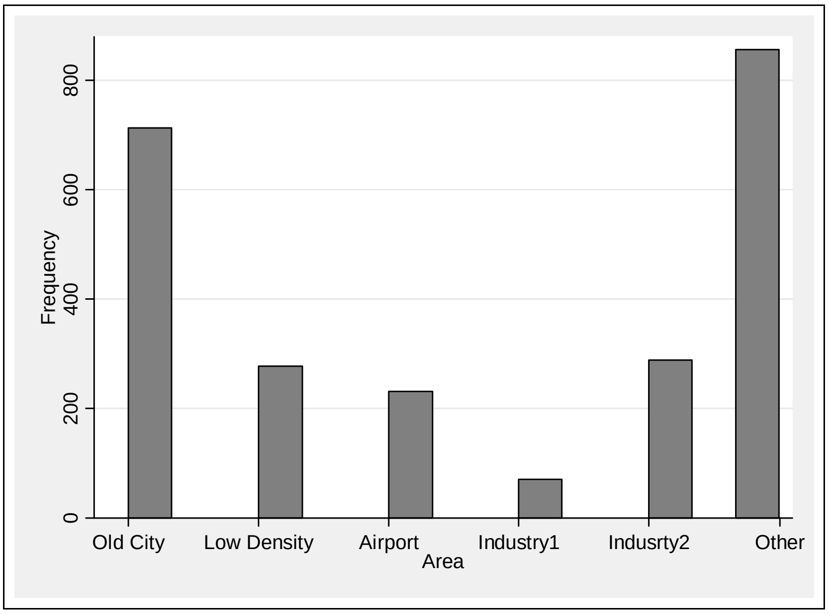

Figure 7.2: Observations in our sample classified according the six locations

Several remedies are available in a very extensive literature on sample bias and non-response problems. We considered adopting a procedure that is available in some software packages and discussed by Nevo (2003). It involves the use of additional sample, the "refreshment sample" to attach weight to each observation to ensure that parameter estimates are not affected by the sampling bias where the observation do have equal probability of being selected in the sample. In this procedure, the probability of inclusion in the sample and the weights (population of respondents from corresponding choice/alternative) are estimated jointly with the coefficients. Another possible method is simply to apply an appropriate adjustment to the constant term.

Fortunately, our data does not suffer from the sample selection problem. Although the data was collected at the local government level in near-equal proportions, our six alternatives cut across the eight local governments that are within the Kano metropolitan area, in such a way that the sample proportions seen in Figure 7.2 are roughly consistent with population proportions.

7.8 Results and Discussions

As mentioned earlier, we estimated five models, ASCLM, ASMPM, NLM, mixed logit and mixed probit models of housing location in Kano. We could not estimate probit models with quadratic terms in water and electricity supply variables because the models would not converge, because the models are not identified. Although it is possible to estimate logit based models with quadratic terms, we drop them in the ASCLM and mixed logit to maintain consistency. The results of all four models are summarized in Tables 7.2 to 7.6. Full computer output of our results (including models that would not converge) are presented in Appendix II.

Given our objective of estimating WTP for water and electricity supply, our emphasis is on the alternative specific models rather than the multinomial logit/probit model(s), which would be useful for the different objective of predicting choice probabilities for given types of individual.

Looking at the covariance matrix from the “mixed” probit model, there seem to be high correlations between the error terms of: Airport and Industrial location 2 (0.697); Airport and Industrial location 1 (0.63); Industrial location 1 and Industrial location 2 (0.54). This can be interpreted in terms of a nesting structure. It appears that Airport, Industrial location 1 and Industrial location 2 constitute a group (in the context of the nested logit model). Any individual who is likely to choose one of these three alternatives is also likely to choose the other two, because the alternatives are "similar". It also possible to just interpret the correlations in terms of similarity. The least similar pair is Low Density and Other, with correlation 0.27.

As explained in detail in Chapter 3, a major limitation of logit-based models is the IIA restriction. Although there are several tests for IIA assumption, none of them is completely satisfactory. Using simulation, Cheng and Long (2007) have demonstrated problems with the two most popular tests of IIA, the Hausman-McFadden and Small-Hsio tests. This is the basis of the next chapter, a meta-analysis study of the IIA assumption. One of the reasons we estimate (alternative specific) Conditional probit and “Mixed” probit models was to avoid this problem.

We started first by conducting a comprehensive IIA assumption test for all the logit-based models, clogit, asclogit and nested logit. For the clogit model we are able to conduct both the Hausman-McFadden and Small-Hsio test of the IIA. Unfortunately we could only conduct Hausman-McFadden test for the asclogit, we were unable to conduct the Small-Hsiao test for asclogit, in STATA 11. Postings on internet sites indicate that this is a common problem with STATA.

From the Hausman-McFadden test results, IIA is satisfied in all the results that were conclusive. We may conclude therefore that our model therefore broadly satisfies the IIA assumption. Note that some of the test statistics are negative. This is common, and Fry and Harris (1996), quoting McFadden (1983), asserts that negative test-statistics support the null hypothesis of IIA.

We decide to estimate the nested logit model after we looked at the estimate from the asmprobit and "mixed" probit models and the correlation and covariance matrices. (We could not estimate the variance covariance matrix of logit-based models in Stata 11).

From the nested logit results we obtain the tau test introduced in section 8.3. We test if the tau is equal to 1, because the model reduces to multinomial logit model, the nesting is unnecessary (Train 2003). We discuss this theoretical issue in detail in section 8.3, in the IIA meta-analysis chapter. From the tau test we further conclude that IIA assumption is satisfied because we could not the reject the hypothesis that tau is equal to/not different from 1.

We estimate and report results from five discrete choice models and estimate the WTP for both water and electricity supply from these estimates. Logit models indicate that electricity supply has a higher WTP compared to water supply.

From the logit based models households WTP for water supply is 4357; 6044; and 5004, while the WTP for electricity supply is 1,5268; 9024; and 9148 for ASCLM, mixed logit and NLM models respectively. Our detailed model results are reported summarized in table 7.2 and 7.3 and corresponding WTP estimates are reported in table 7.4.

Although looking the log-likelihood estimates they perform poorly, we are more inclined to accept the probit-based models, the alternative specific multinomial probit and the "mixed" probit. This is because, probit models have flexible formulation and do not suffer from the restrictions of the IIA assumption. The mixed probit results also correspond with the hedonic pricing model results. Our results from mixed probit and ASMPM are summarised in table 7.5 and WTP estimates in table 7.6.

The log-likelihood function for our models converge at -2002.76 and -1542.19 for the ASCMP and "Mixed" probit models respectively. Two of the three alternative specific attributes included, namely rent and water supply are highly statistically significant and of the theoretically expected sign in both the ASCMP and "Mixed" probit models. Higher rent is negatively related to household choice decisions while higher supply of water supply appears to be a positive amenity. Electricity supply has the expected sign in both models, is mildly significant in the ASCPM but not significant in the "Mixed" probit model.

We use the ***nlcom*** command (in Stata 11) to estimate the WTP because it gives both estimates, standard errors and the confidence intervals. This is useful because it is important to be able conduct a test of the significance of the estimate and to report a confidence interval for true WTP.

Variables	Alternative Specific Conditional Logit Model	Mixed Logit Model Area - Old City Base Choice				
Area		**Low Density**	**Airport**	**Industry 1**	**Industry 2**	**Other**
Constant	-	-0.6908 (0.7800)	-1.7559 (0.5548)	-4.6489 (1.1853)	-0.4741 (0.5857)	-0.0429 (0.4472)
Income/1000	-	0.0004 (0.0002)*	-0.0008 (0.0004)**	-0.0005 (0.0005)**	-0.0004 0.0003	-0.0005 (0.0002)**
Age	-	0.0402 (0.0156)**	0.0275 (0.0127)**	0.1059 (0.0561)*	0.0091 (0.0133)	0.0313 (0.0099)***
Years of Education	-	0.2126 (0.0420)***	-0.0345 (0.0243)	-4.6489 (1.1853)	0.0188 (0.0274)	0.0388 (0.0205)*
Rent	-0.0907 (0.0021)***	-0.1082 (0.0028)***				
Water Supply	0.0395 (0.0065)***	0.0654 (0.0131)***				
Electricity Supply	0.1385 (0.0190)***	0.0976 (0.0240)***				
Number of Observations	14448	14100				
LR/Wald Chi²	*LR* X^2(3) 4839.30	*Wald* X^2(18) 1539.73				
Log Likelihood	-1894.9068	-1168.9595				

Table 7.2: Discrete Choice Models (Alternative Specific Conditional Logit) Regression Results – Residential Location Choice Decision

Legend:
All models computed using Stata 11
Standard errors in brackets
***** Mildly Significant (ρ <0.10) ****** Significant (ρ <0.05) ******* Strongly Significant (ρ <0.01)

Variables	Nested Logit Model Nesting Structure: Old city (Base Choice), Low Density, Pollution (Airport, Industry1, Industry2), and Other		
Area	**Low Density**	**Pollution (Airport, Industry1, Industry2)**	**Other**
Income/1000	0.0005 (0.0002)**	-0.0004 (0.0002)*	-0.0005 (0.0002)**
Age	0.0247 (0.0107)**	-0.0077 (0.0051)	0.0273 (0.0052)***
Years of Education	0.1857 (0.0322)***	-0.0278 (0.0168)*	0.0369 (0.0171)**
Rent	-0.1031 (0.0029)***		
Water Supply	0.0516 (0.0095)***		
Electricity Supply	0.0943 (0.0228)***		
Number of Observations	14100		
Wald Chi^2	X^2(18) 1256.55		
Log Likelihood	-1220.2135		
LR test for IIA (tau = 1)	X^2(3) 21.13		

Table 7.3: Discrete Choice Models (Nested Logit) Regression Results – Residential Location Choice Decision

Legend:
All models computed using Stata 11
Standard errors in brackets
***** Mildly Significant (ρ <0.10) ****** Significant (ρ <0.05) ******* Strongly Significant (ρ <0.01)

Utilities	Alternative Specific Conditional Logit Model	Mixed Logit Model	Nested Logit Model
Water Supply	0.4357 (0.0692) 0.3001 ↔ 0.571)♣	0.6044 (0.1193) 0.3706 ↔ 0.8382♣	0.5004 (0.0901) 0.3237 ↔ 0.6770♣
Electricity Supply	1.5268 (0.2039) 1.1272 ↔ 1.9265♣	0.9024 (0.2189) 0.4733 ↔ 1.3315♣	0.9148 (0.2165) 0.4904 ↔ 1.3392♣

Table 7.4: WTP for water supply and electricity supply from Alternative Specific Conditional Logit and Nested Logit Models

Legend:
Standard errors in brackets
WTP estimated using the delta method, computed using Stata 11
WTP = - (Coefficients of Water Supply/Electricity Supply) divided by the Price (Rent) Coefficient
♣ Confidence Intervals - 95% level of confidence.

Variables	Alternative Specific Multinomial Probit Model	Mixed Probit Model Area - Old City Base Choice				
Area		**Low Density**	**Airport**	**Industry 1**	**Industry 2**	**Other**
Constant	-	-1.4918 (0.3481)	-0.3855 (0.1913)	-2.0632 (0.4412)	0.0388 (0.2018)	-0.1075 (0.1875)
Income/1000	-	0.0005*** (0.0001)	-0.0003*** (0.0001)	-0.0003 (0.0002)	-0.0002** (0.0001)	-0.0002** (0.0001)
Age	-	0.0184*** (0.0071)	0.0057 (0.0043)	0.0198** (0.0082)	0.0029 (0.0045)	0.0122*** (0.0041)
Years of Education	-	0.1019*** (0.0169)	-0.0115 (0.0082)	0.0442** (0.0187)	-0.0053 (0.0091)	0.0141* (0.0083)
Rent	-0.0273*** (0.0011)	-0.0361*** (0.0017)				
Water Supply	0.0124*** (0.0026)	0.0237*** (0.0049)				
Electricity Supply	0.0125* (0.0067)	0.0123 (0.0084)				
Number of Observations	14448	14100				
Wald Chi^2	X^2(3) 637.81	X^2(18) 483.62				
Log Likelihood	-2002.76	-1542.19				

Table 7.5: Discrete Choice Models (Probit) Regression Results – Residential Location Choice Decision

Legend:
All models computed using Stata 11
Standard errors in brackets
***** Mildly Significant (ρ <0.10) ****** Significant (ρ <0.05) ******* Strongly Significant (ρ <0.01)

Utilities	Alternative Specific Multinomial Probit Model	Mixed Probit Model
Water Supply	0.4519 (0.0893) 0.2768 ↔ 0.6270♠	0.6569 (0.1328) 0.3967 ↔ 0.9171♠
Electricity Supply	0.4580 (0.2410) -0.0144 ↔ 0.9304♠	0.3421 (0.2319) -0.1124 ↔ 0.7967♠

Table 7.6: WTP for water supply and electricity supply from probit models

Legend:
Standard errors in brackets
WTP estimated using the delta method, computed using Stata 11
WTP = - (Coefficients of Water Supply/Electricity Supply) divided by the Price (Rent) Coefficient
♣ Confidence Intervals - 95% level of confidence.

Our estimates of WTP water supply and electricity supply are both within a reasonable range (at 95% confidence interval) in both ASCPM and "Mixed" probit models. Results are reported in table 7.6 and detailed computer output in Appendix II.

The estimated WTPs from the ASCMP model are 0.4519 and 0.4580 for water and electricity supply respectively. The estimates for WTP for water supply are higher in the "Mixed" probit model, 0.6569 while the estimate for WTP for electricity supply is lower 0.3421 compared to the ASCLM. The mixed logit model

We are more inclined to accept the estimates from the "Mixed" probit model because the results roughly corresponds with our hedonic pricing model results and the ASCMP model includes only location attributes, while the "Mixed" probit model takes accounts of the household socioeconomic characteristics as well. In money terms, this implies, households are willing to pay ₦656.9 more rent (per annum) for one hour's additional daily supply of water and ₦342.1 for one hour's additional daily supply of electricity. Converted to pounds sterling, these valuations are around £2.50 and £1.30 respectively.

We could also analyse location choice decisions from the mixed probit model. From the estimated choice probabilities we see that years of education, income and age are positive and highly significant in the low density area. This means that the rich, more educated and older households are more likely to reside in this part of the city relative to the old city. Income is negative for areas close to airport, industries and Other. It is also statistically significant in all but industrial estate 2. This result is consistent with expectations because financially better off households are less likely to live in these areas.

7.9 Conclusion to Chapter Seven

The results from this Chapter support our previous results in the hedonic pricing model. There is a positive premium for living in areas with longer period of water (and electricity) supply in Kano. This is evidenced by the positive WTP we estimated from the attribute based location choice model.

Although our results show that consumers have absolute higher preference for water supply in comparison to electricity supply, both are vital for life. There is a strong relationship between these two utilities because sustained water supply reticulation depends on availability of electricity. On its own right, electricity is required for both domestic chores and air conditioning in a place where the day time temperature hovers around 35 degrees Celsius most of the year and the prevalence of small scale enterprises (most located in residences) that depend on public electricity which is relatively cheaper than other sources of energy.

We recommend serious effort to provide electricity in the short-run which would increase water supply through its multiplier effect. Public-private sector partnership could be used to source funds, technology and management for both electricity and water supply at the relevant government levels. Tripartite collaboration between the private sector, federal and state governments could be employed to provide these public utilities. This is because the private sector, if given the necessary conditions, would be interested for two reasons, the market potentials (large population) and high WTP for both water and electricity as shown in this chapter.

As the supply of electricity increases, Governments should encourage the use of energy saving devices. This would reduce the demand for electricity, the surplus capacity could be used to further increase water supply (and the provision of other public utilities such as street lighting and traffic lights). This would make it possible to archive 24 water supply. But lack of proper planning or uncontrolled expansion of city would aggravate the problem. We therefore recommend, in addition, a strict town planning control. This is important to monitor demand and how to provide for public utilities (and other municipal services).

Chapter Eight: IIA Meta-Regression Analysis

8.1: Introduction

The research reported in this Chapter is motivated partly by one of the conclusions reached in Chapter 7: that the assumption of Independence of Irrelevant Alternatives (IIA) is broadly accepted for residential location choice data that we were analysing there. This conclusion led us to ask if IIA is typically accepted in studies of residential location. Thinking more generally, we considered whether there might be a greater tendency for IIA to be accepted in some types of application than others. More fundamentally, what proportion of all IIA tests that are carried out, result in rejection? The literature is not informative on this sort of question. For example, Cramer (1991, p.48) writes:

> *"The IIA property is due to the blind indifference of the model to any similarity or dissimilarity of the S states, which are all treated on the same footing. This is a substantive assumption, and in many applications it is clearly inappropriate."*

How does he know it is inappropriate in many applications? In what sort of applications is it most inappropriate? These are questions that appear to be unanswered in the literature. Here we attempt to answer these and other related questions by collecting a large number of tests of the IIA assumption from the previous literature, and analysing these results in a meta-analytic framework.

8.2: IIA and the Multinomial Logistic Regression

In deriving our logistic regression model (in Chapter 3) we made some assumptions. We assume that the ε's are independently and identically distributed random variables (they have the same variance and zero covariance) and follow a Gumbel distribution. Related to this is the assumption of independence from irrelevant alternatives. That is, the odds ratio amongst alternatives is not affected by adding one or more alternatives. For example, the odds ratio of alternatives 1 and 2, do not depend on the presence of a third alternative:

$$\frac{\text{Prob}(1)}{\text{Prob}(2)} = \frac{\exp(x'\beta_1)}{\exp(x'\beta_2)} = \exp\left[x'(\beta_1 - \beta_2)\right] \qquad (8.1)$$

This ratio is independent of β_3. To put it differently, this assumption requires that the ranking between two bundles in a choice set is not affected by the content of the remaining bundles in the set.

The violation of the IIA assumption may occur for various reasons, specification of the model, the inclusion of close substitutes the in choice sets, structure of the unobserved factors in the error term, the existence of random taste variations, i.e. heterogeneous preferences (Salensminde, 2002), (Travisi and Nijkamp, 2004).

This problem is peculiar to logistic models - multinomial and conditional logit models. But logistic models have remained popular because of the ease of computation and interpretation of results. Other microeconometric models do not suffer from the same problem, for example Count models are free from the IIA assumption and, unlike logistic models, actually benefit from increased in the numbers of alternatives by adding degrees of freedom (Kim et al, 2008). More 'flexible' discrete choice models, although more computationally difficult, could be used as alternatives to solve for the violation of the IIA.

Some economists have questioned this 'obsession with' the IIA assumption by econometricians on the grounds that the elements of a choice, or feasibility set can convey information that affects one's choices and values (Basu, 2000), (Bateman et al, 2005). Using the *internal consistency argument*, Basu (2000) shows that it is plausible and perfectly rational for individual decision to be affected by an addition or subtraction of alternatives depending on the individual socioeconomic background. He supports his argument with an example of a Muslim choosing between three restaurants. She may decide not eat anything in the chosen restaurant after discovering that the restaurant also serve Pork. Hausman and McFadden (1984) acknowledge that, it is not only the elements of a choice that affect the IIA assumption, an alternative specification (as well as functional form) of a model might satisfy IIA.

Does this argument permanently condemn the logistic model or do we just ignore the IIA problem? Answer to this is beyond the scope of this study. However, there is a middle ground. It is recommended for applied research to test for the violation of this assumption and to proceed with logistic model if it is satisfied otherwise to use other models that do not rely on the IIA assumption.

Violation of the IIA assumption would result in inconsistent and biased estimates, and incorrect predictions (of e.g. market share) (Fry and Harris, 1998; Mazzanti 2003). The assumption and the cost it imposes have assumed great importance in empirical research ever since the estimation of superior choice models (NLM, MNP) first became feasible, due to better understanding of the underlying theory and availability of computer software.

8.3: IIA Tests

Several tests have been developed to test for the violation of the IIA assumption. The most popular tests are the choice partitioning tests, also called likelihood ratio tests. There is also a very wide body of literature that seek to interrogate the choice portioning IIA test procedures. Majority of these studies which are based on simulation, attempt to examine the various test procedure, their robustness and size property. This is because, the three choice partitioning tests that are commonly used to test for IIA frequently arrive at different conclusions (Long and Freese, 2003).

It has been shown that some test might work poorly in small samples while others are asymptotically biased. Cheng and Long (2007) uses series of Monte Carlo simulations to evaluate three tests of IIA. They show that the size properties of the three IIA tests depend on the data structure for the independent variables and that tests of the IIA assumption that are based on the estimation of a restricted choice set are unsatisfactory for applied work. Fry and Harris (1996) investigates the size and power properties of six tests for IIA in the multinomial Logit model. Their results show that the majority of tests based upon partitioning the choice set appear to have very poor size and power properties in small samples.

Choice Set Partitioning Tests

McFadden, Train, and Tye Test (MTT) test

The MTT test (McFadden, 1981), involves estimation of the full model with all the alternatives and then estimating a restricted model with a sub-set of alternatives. This is an approximate likelihood ratio test, with degrees of freedom equal to rows of the restricted model.

$$MTT = -2\left[\log L\left(\hat{\hat{\delta}}_C\right) - \log L\left(\hat{\delta}_D\right)\right] \tag{8.2}$$

Small and Hsiao (1985) have shown that the MTT is asymptotically biased towards accepting the IIA model structure. This is because of the use of overlapping estimation sample which produces values of MTT that tend to be small favouring the hyposthis of IIA (Fry and Harris, 1998).

Small and Hsiao (SH) Test

Small and Hsiao test (Small and Hsiao, 1985) is an improved version of MTT. In this test the unrestricted model is estimated on samples generated by randomly dividing the total sample into two equal parts. The model produces a weighted average of estimates $\hat{\theta}_f^{AB}$.

$$\hat{\delta}_C^{AB} = \left(1/\sqrt{2}\right)\hat{\delta}_C^{A} + \left(1 - 1/\sqrt{2}\right)\hat{\delta}_C^{B} \tag{8.3}$$

The model is estimated on a restricted sub-sample. The sub-sample is restricted by eliminating observations of a given alternative. The Small and Hsiao test has a χ^2 distribution with degrees of freedom equal to the number of explanatory variables plus one.

$$SH = -2\left[L_r^B\left(\hat{\theta}_f^{AB}\right) - L_r^B\left(\hat{\theta}_r^{B}\right)\right] \tag{8.4}$$

Hausman-McFadden (HM) Test

This is the most widely used IIA test. Hausman and McFadden (1984) propose a modified Hausman test where the full choice set and restricted discrete choice models are estimated, using the maximum log likelihood. If the IIA assumption holds, the restricted and unrestricted models should be consistent for the same parameters.

$$HM = \left(\hat{\theta}_r - \hat{\theta}_f\right)' \left[C\hat{o}v\left(\hat{\theta}_r\right) - C\hat{o}v\left(\hat{\theta}_f\right)\right]^{-1} \left(\hat{\theta}_r - \hat{\theta}_f\right) \qquad (8.5)$$

The HM test is asymptotically Chi-square distribution with degrees of freedom equal to the number of elements that is identifiable from the restricted set model (Ben-Akiva and Lerman, 1985).

There are several other tests, an example is the Nested logit-based IIA test. This test involves a test of equality of estimated coefficients $\beta's$. Hausman and McFadden (1984) proposed the nested logit test by looking at the scalar parameter of the choice model assuming that an individual forms a weighted average of the attributes of alternatives called the inclusive value, also called the log-sum.

From our nested logit equation in Chapter 3, ξ is the scalar parameter of the model. It account for similarity of the error term.

$$P_i = \frac{\exp(\gamma Z_i + \xi_i D_i)}{\sum \exp(\gamma Z_s + \xi_i D_s)} \qquad (8.6)$$

If $\xi = 1$, the model reduces to multinomial logistic model, meaning the nesting of alternatives is unnecessary because there is "no correlation among unobserved components of utility for alternatives with a nest" (Train 2003: pp 84). If 0 < ξ < 1, the model fails to satisfy the IIA assumption but consistent with the random utility model. For ξ > 1, interpretation of the choice model becomes problematic (Hausman and McFadden, 1984).

McFadden (1984) provides the generalised formula for the nested logit model-based IIA test, using the scalar parameter, when the alternatives are more than three.

$$W = \frac{(1-\xi)^2}{SE_{\lambda}^2} \qquad (8.7)$$

The Wald statistics for the null hypothesis that IIA is satisfied, which is χ^2 with 1 degree of freedom.

8.4: Motivation

Our first motivation is based on evidences from several empirical studies (most using Monte Carlo simulations) to test the efficacy of the popular IIA tests. Results from these studies have shown that there are flaws in the two popular IIA tests – Hausman-McFadden and Small-Hsiao tests. It was found that, the two test mostly produce conflicting results, Small-Hsiao is more likely to reject the IIA assumption and Hausman-McFadden is sensitive to the sample size. Long and Freese (2003), Cheng and Long (2007), Fry and Harris (1996)

The second motivation is from our discrete choice results, in the previous chapter, which indicate that the IIA assumption has not been violated. Several other residential location decision studies have also reported similar results. Tu and Goldfinch (1996), Cho, (1997) argue that the IIA axiom may not be violated in residential location studies because individuals should have made their housing choice after obtaining full market information. Cho (1997) argues that, because housing takes highest proportion of household income and moving cost, households would pay more attention to searching and therefore would commit less error in making their choice, it is very likely that the IIA assumption is satisfied is most housing location studies. This argument is based on the suggestion by Tu and Goldfinch (1996) that, the question of whether residential location decision (joint or sequential) has little effect on the final outcome if choices are made with full market information.

Accordingly, household's decision would not be influenced by the decision process and additional alternatives if it is made after obtaining complete information on the market. Since buying a dwelling is the biggest lifetime decision of most households, because of the high cost, source of finance – life savings and mortgage - it is reasonable to assume that households would be very careful when choosing a dwelling and would not buy a dwelling until they find something suitable (Tu and Goldfinch, 1996) (Colom and Molés, 2008) (Yates and Mackay, 2006).

Heuristics in decision making do not directly affect the IIA assumption but affect ε, the random component of utility. If, after house-search, households could obtain complete information on the market, they would choose the alternative that maximises their utility and the household's final choice would be influenced solely by their socioeconomic characteristics. This argument is stronger in the case of own-house location choice decision.

Colom and Molés' (2008) study of Spanish households tenure choice (own vs rent) and dwelling size indicates that households arrive at a decision without accounting for similarities among the six available alternatives. The multinomial logit model is found to be better suited to describe their behaviour than the other hieratical models which mean, intuitively, the IIA assumption is not a problem.

Dahlberg and Eklöf (2003) compared the predictions of three difference discrete choice models; the conditional logit model, the mixed logit model, and the multinomial probit model. They reported that in residential location studies, the conditional logit model leads to exactly the same conclusions with models that relax the IIA assumption as long as the model is not too parsimonious.

These empirical results notwithstanding, as we point out in Section 8.2, whatever the application, whether or not IIA holds also depends on the specification of the model (Hausman and McFadden, 1984). However, satisfying the IIA assumption is hardly the primary condition in model selection process.

The Relationship between Specification Error and IIA in Discrete Choice Models.

A pertinent question to ask is, why would specification error in discrete choice models lead to violation of the IIA assumption. There is a close relationship between IIA testing and specification testing. An apparent violation of IIA may be for the straightforward reason that the model is misspecified. To see this, consider the choice model estimated in Chapter 7 in which the choice between 6 locations was assumed to be determined by rent, water supply and electricity supply. Utility derived by individual i from choosing location s was assumed to be given by:

$$y_{is}^{*} = \alpha_1 r_{is} + \alpha_2 w_{is} + \alpha_3 e_{is} + u_{is} \quad s = 1, \cdots, 6 \quad i = 1, \cdots, n \qquad (8.8)$$

(8.8) is in fact the conditional logit model (CLM). For present purposes, let us assume that (8.8) is the true model, and also that the error terms are uncorrelated between alternatives, hence that IIA is satisfied when (8.8) is estimated.

Now let us imagine that data on electricity supply is not available, so that the model must be estimated without this variable. The model becomes:

$$y_{is}^{*} = \alpha_1 r_{is} + \alpha_2 w_{is} + v_{is} \quad s = 1, \cdots, 6 \quad i = 1, \cdots, n \tag{8.9}$$

Since electricity supply is unobserved, it becomes part of the error term in (8.9). The relationship between the error terms in the two models is, approximately:

$$v_{is} = \alpha_3 e_{is} + u_{is} \quad s = 1, \cdots, 6 \quad i = 1, \cdots, n \tag{8.10}$$

(8.10) tells us that, in model (8.9), we expect a positive correlation in the error terms between areas with similar levels of electricity provision. Hence it will appear that areas with high electricity provision are "similar" to each other (in the sense of nested choices), and that areas with low electricity provision are also "similar" to each other. This, of course, means that there is a violation of IIA, as a direct result of the exclusion of the electricity supply variable.

Because of the large volume of discrete choice studies (made possible by development in theory and computational feasibility) we systematically revisit these issues using a meta-analysis. We believe that, a meta-analysis which would compare results from different studies would contribute to the debate and provide further insight on the two popular tests of the IIA assumption. To our knowledge, this study is the first of its kind.

Our meta-analysis is unique because it is looking at a particular technique. All the other meta-analyses that we have come across look at the magnitudes of the effect of a particular policy such as employment and minimum wage effects (Card and Krueger, 1995) or price elasticity of demand (Dalhuisen et al., 2003).

8.5: Meta-Regression Analysis

The term meta-analysis, which originated in medical science, psychology and psychotherapy, is due to Gene Glass and the earliest reported meta-analysis study is a report by Karl Pearson in 1904. Schulze (2004) defines meta-analysis as a study of studies, a method for systematic literature review on certain substantive question of interest. One of the major advantages of the meta-analysis, as quantitative literature review, is that it can resolve differences between studies/systematic variation that are likely to occur due to difference in the type and format of the data, location of study, technique used in data analysis.

It is possible to analyse the extent to which any of these variables affect reported results. The broad objective of meta-analysis, in its general context is to synthesise current knowledge, reveal or prove cumulation of knowledge, clean-up or make sense of research literature, analyse effect size and determine moderator variables on a particular research problem (Littell et al, 2008), (Leandro, 2005), (Schulze, 2004).

The application of meta-analysis in economics has been relatively new. While it started in medical science as an attempt to synthesise conflicting results from clinical trails or "flood of conflicting scientific evidence", it was the "avalanche of information" and large volume of research that made qualitative literature review in economics unattractive. Something was needed that could provide a balanced and systematic literature review. This was happening at a time when the proponents of meta-analysis have won the support of statistician. It was discovered that, it is possible to conduct a detailed study on particular issue, similar to the conventional econometric analysis, by looking at a number of independent studies that have used different data set and methods, which could provide more "insight and greater explanatory power" than individual studies (Stanley, 2001, 2005).

It has been argued that Economic is going through "a renaissance" which has resulted an "avalanche" in the number of empirical researches. These huge numbers empirical studies have become difficult to comprehend, in spite of the ambiguity in the findings which has rendered qualitative literature review almost unfeasible (Stanley, 2005).

Meta-analysis regression is conducted in the same way as normal regression analysis. The dependent variable could be continues or discrete. The dependent variable is determined by the objective of the study, the explanatory variables (also called moderator variables) include the relevant attributes of the articles to be included in the study. Meta-analysis usually focuses on a research question (e.g. does the introduction of a minimum wage reduce employment?) in contrast, this study is a meta-analysis of a particular methodology (namely, IIA testing).

A meta-analysis starts with a data collection where all available research articles (both published and unpublished) is collected. The sampled articles and their "attributes" are synchronised to arrive at a comparable metric (Stanley, 2001). Both random and fixed effects models could be estimated depending on the nature of the dependent variable.

8.6: Our Meta-Analysis Data

There are two aspects in which our meta-analysis breaks new ground. First, combing results from different studies that are presented in different ways (ie reject/accept and/or exact P-value). We employ microeconometric technique, ordered probit using p-values which reflect extent of IIA acceptance/rejection. Second, because some studies report more than one IIA test result dealing with multiple observations problem, using panel data technique (random effects probit model). All previous meta-analysis studies that we come across analyse only one observation.

Our objective is to estimate two discrete choice models, ordered probit model with p-values as dependent variable to test for the extent of acceptance/rejection and binary probit model of IIA acceptance/rejection. We therefore collected all available discrete choice studies that reported IIA results. Some studies where excluded due to insufficient information.

We collected data using popular Internet search engines and journal publishers' web page search facility namely: google, journal publishers/archive search facility - ScienceDirect, Jstor, Ingenta and Springer. Both published and unpublished articles that report IIA test results are included. We obtain detailed statistics using electronic search in the relevant software in which the article is published (mostly adobe pdf and

Microsoft word documents). We fast-read articles were search facility is not available example where the pdf is from scanned pages and treated as a picture.

Our data set may be biased towards recent articles because most of pre-2000 pdf files are scanned pictures files and could not be search from the major internet search engines. However, this is not a major problem because, we could argue that it was about this time that Microeconometrics became very popular and its application became diffused. It was the 2000 Bank of Sweden/Nobel Prize in Economics science jointly awarded to James Heckman (theory and methods for analyzing sample selection) and Daniel McFadden (theory and methods for analyzing discrete choice models), the advancements in personal computer technology (computers with built-in math processors), and availability of computer software, that has made it possible to undertake empirical work in Microeconometrics in general and post-estimation tests in particular.

Unpublished works are included to avoid publication bias or the so called "file-drawer problem". This bias, which may not be deliberate, arises when journal editors and reviewers show a preference for statistically significant results and they are more likely to be published. While studies that find smaller and/or insignificant results and inconsistent with conventional view, are less likely to be published. Most researchers are reluctant to submit some articles for publication and they end up in their "file-drawer" because they believe that their findings do not meet certain expectations or do produce positive (or expected) results (Stanley, 2005).

A very simple way to correct for this bias is to collect all available studies, published and unpublished such as working papers, whether results are statistically significant or not. The problem with this method is that unpublished studies are hard to obtain relative to published studies. Another method is to use dummy for models used in each study. This is because another bias among editors and reviewers, a preference for certain (perhaps more complicated/elegant) model specification.

We acknowledge that our sample may be baised because many studies using models that relax or do not impose IIA, either because they want to avoid IIA or because of what they set out to examine, such as models with unobserved heterogeneity, may not report tests for it.

Our sample consists of 182 studies, of which 118 (64.84%) are peer reviewed and published in academic journals. Because the data is an unbalanced panel data, we have 374 observations, up to 8 observations per paper, and an average of 2.1 observations per paper.

IIA Test Result	***IIA Test Type***		***Total***
	Hausman-McFadden	Small-Hsiao	
Accept	215 (73%)	34 (67%)	249 (72%)
Reject	79 (27%)	17 (33%)	96 (28%)
Total	294 (100%)	51 (100%)	345 (100%)

Table 8.1: Summary of IIA Test Result by Test Type

In table 8.1 we see that almost three quarters (72%) of the IIA tests ever done have resulted in acceptance of IIA. We found this somewhat surprising given the obsession that there appears to be in the literature over the possibility of violations of IIA. We also see that 73 percent of Hausman-McFadden tests accept IIA, compared to only 67 percent of Small-Hsiao test. This is consistent with the power advantage of the Small-Hsiao test, that has been reported on the basis of monte carlo work (Fry and Harris, 1996).

Variables/Attributes of each study included our sample	***Definition***
Author(s) Name	Surname of Author(s)
Number of Authors	Dummy variable - 1 More than 1, 0 Single Author
Year	Year of publication or Working Paper
Post Nobel	Dummy Variable Post 2000 Daniel McFadden Joint Nobel Prize in Microeconometrics – Before and After
Published	Dummy variable 1 published, 0 unpublished
Journal	Name of the Journal
Journal Impact Factor	IDEAS/RePEc Impact Factors for Economics Journals
Choice Model	Dummy variable - 1 Multinomial logit, 2 Probit, 3 Nested Logit, 4 Mixed Logit, 5 Conditional logit, 6 Other
N	Sample
IIA result	Dummy variable - 1 reject, 0 accept
Test type	Dummy variable - 1 Hausman-McFadden test, 2 Small-Hsiao test, 3 Other
t/Chi^2 test	Reported statistics
df	Reported degrees of freedom
p-values	Reported statistics
Country studied	Country where the research was conducted
Per Capita	Per Capita Income of the Country where the research was conducted
Low, Lower and Upper Middle, High Income Countries	World Bank Classification of the country where the research was conducted
No of alternatives	Number of alternative/choices in the study
Nature of study	Dummy variable - 1 Transport mode choice, 2 Residential location choice, 3 Firm location choice, 4 Environmental valuation/pollution/utility/public goods studies, 5 Brand choice, finance and insurance, 6 Healthcare and medicare, 7 TV, telephone and Internet services, 8 Employment/labour studies, 9 School choice, 10 Voting, collation and political decisions, 11 Other

Table 8.2: Description of Variables

Specifically the following variables were created from information collected from the selected studies: names of authors, year of publication, published/unpublished, name of journal if published, journal impact factor, choice model used in the study (multinomial, conditional, mixed, nested logit and probit models), sample size used in the study, IIA test procedure (we collapsed them into three, Hausman-McFadden, Smal-Hsiao tests and other), IIA result (IIA assumption accepted or rejected), t/Chi^2

test statistic, degrees of freedom, p-values, country studied, country per capita income, country income status based on World Bank classification 1 to 4 (Low, Lower and Upper Middle, High Income Countries respectively), number of alternatives in the study, and nature of the study. Table 8.2 contains the definitions of all variables.

We use the following classification for the nature of study: transport mode, residential location, firm location, environmental and natural resource valuation, brand choice, finance product, health care and medicare, TV/telephone and internet, employment/labour, schools, voting/political decisions and other).

Nature of Study	*Number of studies*	*% of total*
Transport Mode	4	1.11
Residential Location	10	2.77
Firm Location	18	4.99
Environmental Valuation	91	25.21
Brand Choice	26	7.2
Financial Product Choice	34	9.42
Health Care/Medicare Choice	17	4.71
TV/Telephone/Internet Choice	60	16.62
Employment/Labour Choice	1	0.28
Voting/Political Choice	21	5.82
Other	79	21.88
Total	**361**	**100**

Table 8.3 Details on the nature of study

Table 8.3 shows a tabulation of studies by Nature of study. We see that the most common sort of study is environmental and natural resource valuation (a quarter of the total studies) and the most obscure is job/employment/labour choice (less that 0.5% of the total number of studies).

To remove outliers in some of the variables, we did some transformations. We generated the log of the sample size and per capita and trimmed the number of alternatives to maximum of eight. We also generated and created a ranking of p-values for the ordered probit regression model. Further details our model are provided below.

To give a feel for the type of data we have collected, in Table 8.4 we present details of two particular studies. What, briefly, do we see in these two examples? We see that Avalos and Hoyos (2008), in their model of Mexican merger decisions, accept IIA using the HM test, but reject using the SH test. The second example is Magnani's (2009) study of workers mobility choices, in which IIA is accepted using both tests. A detailed list of all studies used in the meta-analysis is provided in Appendix IV.

8.7: The Model

The information given relating to each test of IIA tends to vary between studies. Sometimes, complete information is given. Namely, IIA test results (accept/reject), the chi-squared test statistic is given, along with the degrees of freedom for the test, and the p-value.

In some other cases, only partial information is provided. For example, some give a test statistic with no degrees of freedom. Some simply indicate whether the test accepts or rejects. In order to perform a meta analysis using all of the available information extracted from the papers, we adopt the following approach.

We are particularly interested in the IIA test result, a binary decision of accept/reject, and the p-values which in some cases are not reported. Where the p-value is not reported chi-squared test statistics and degrees of freedom are reported we calculated the p-value in Microsoft Excel using the following Excel formula: CHIDIST (χ^2, df) Hensher et al (2005).

We need the p-values because it indicates whether IIA is accepted or rejected and the extent of acceptance/rejection. Because p-values range between 0 and 1, we can consider it as a ranking of the IIA test results. Statistics theory (a rule of the thumb) tells us that, lower p-values, ranging between 0.00 and 0.049 mean a rejection of the IIA assumption, while significantly higher p-values, between 0.5 and 1.0 indicate acceptance of the IIA assumption.

To estimate the ordered probit model, we set the p-values to one decimal point giving us 11 rankings of the IIA tests results.

Ordered and sequential models are applied to choice models where there is a natural ordering/ranking of alternatives. This topic was briefly introduced in Section 6.4. Using our notation in Chapter 3, in line with Cameron and Trivedi (2005), we can write the *m*-alternative model as follows:

$$y_i^* = x_i'\beta + u_i \qquad (8.11)$$

$$y_i = \begin{cases} 1 \; if \;\; y_i^* \le \mu_1 \\ 2 \; if \;\; \mu_1 < y_i^* \le \mu_2 \\ 3 \; if \;\; \mu_2 < y_i^* \le \mu_3 \\ \vdots \qquad\quad \vdots \\ m \; if \;\; \mu_{m-1} < y_i^* \end{cases}$$

$$y_i = j \qquad\qquad if \;\; \alpha_{j-1} < y^* \le \alpha_j$$

Where

$\alpha_0 = -\infty$ and $\alpha_m = \infty$

$$\Pr\left[y_i = j\right] = \Pr\left[\alpha_{j-1} < y^* \le \alpha_j\right] \qquad (8.12)$$

$$= \Pr\left[\alpha_{j-1} < x_i'\beta + u_i \le \alpha_j\right]$$

$$= \Pr\left[\alpha_{j-1} - x_i'\beta < u_i \le \alpha_j - x_i'\beta\right]$$

$$= F\left(\alpha_j - x_i'\beta\right) - F\left(\alpha_{j-1} - x_i'\beta\right) \qquad (8.13)$$

F is the conditional density function of u_i. As in other multinomial choice models (see Chapter 3) the estimation model would be determined by the assumption on the distribution of this cdf. If we assume logistic distribution, we arrive at ordered logit model, with all the attending limitations. The ordered probit model is premised on the assumption of normal distribution of the cdf. Cameron and Trivedi (2005)

Study	*Number of Observations*	*Number of Alternatives/Choices*	*Model(s)*	*IIA Test Result(s)*	*p-value, $\chi 2$ and df*
Avalos and Hoyos (2008), An Empirical Analysis of Mexican Merger Policy *Review of Industrial Organisation*	239	3 – possible decisions: a. Allowed; b. Conditioned; and c. Challenged	Multinomial and Ordered Logit	HM Test – accept SH Test - reject	Not reported Not reported
Magnani (2009), How Does Technological Innovation and Diffusion Affect Inter-Industry Workers Mobility (USA) *Structural Change and Economic Dynamics*	15004	4 mutually exclusive regimes of mobility: a. No mobility occurs; b. Mobility within 3-digit industries occurs (Intra3D mob); c. Mobility within 2-digit industries, but between3 − digit industries occurs (Inter3D mob); and d. Mobility between 2-digit sectors occurs (Inter2D mob)	Multinomial and Mixed Logit Models	HM Test – accept SH Test - accept	HM Test - $\chi 2$ 34.579(32) p-value 0.346 SH Test - $\chi 2$ 35.932(32) p-value 0.289

Table 8.4: Details of two examples of studies collected

If we assume normality for our model, the probability that y_i falls into s_{th} category is given by:

$$Prob(y_i = s) = \Phi(\mu_s - \beta' x_{ij}) - \Phi(\mu_{s+1} - \beta' x_{ij}) \qquad (8.14)$$

As with all the maximum likelihood models, the log-likelihood is the sum of individual probabilities.

$$LogL(\theta) = \sum_{s=1}^{N} \sum_{y_i = s} Log\left[\Phi(\mu_s - \beta' x) - \Phi(\mu_i - \beta' x)\right] \qquad (8.15)$$

Another way of performing this meta analysis is simply to model the binary variable representing acceptance or rejection of IIA, that is, to disregard the strength of evidence as represented by the p-values. To do this, we simply apply the random effects probit model to this binary variable. The random effects probit model is:

$$\Pr\left(reject_{ij} = 1 \,|\, x_{ij}\right) = \Phi\left(\beta' x_{ij} + u_i + \varepsilon_{ij}\right) \qquad (8.16)$$

Note that the explanatory variable vector x_{ij} has both an *i* subscript and a *j* subscript. This is because, while most of the explanatory variables in the model apply only to the article under analysis, there are a few explanatory variables that vary between the different tests within one article. For example, some articles report both a Hausman-McFadden test and a Small-Hsiao test, so any variable indicating which test has been used must vary between observations within such articles.

8.8: Results

As pointed out above, we estimate two models, random effects and ordered probit models. The summary of the random effects probit model result is provided in tables 8.5 and 8.6, while the ordered probit results is presented in table 8.7. Comprehensive results, the computer output are presented in Appendix III

Our main objective is to find out if there is a pattern in the outcome of IIA test results, by estimating the probability (and extent) of rejection (and otherwise) depending on certain 'attributes' of a particular study, such as sample size, country and location studied, type of study, whether published, (do publishers/reviewers have a tendency to reject articles based on IIA test outcome?). We included sample size in our estimation, but this study is not about asymptotic power of a particular (or generic) IIA test procedure, Monte Carlo study is most suitable, and that issue has been sufficiently handled by other studies (Fry and Harris, 1996)

Explanatory Variables	***Coeff. (Std Error)***
Constant	-14.014 (5.733)**
Published	0.634 (0.317)
Jimfact	-0.014 (0.080)
Log n	-0.063 (0.069)
Log per-capita income	1.067 (0.536)***
Number of alternatives	0.156 (0.090)*
Model mprobit	1.544 (0.674)**
Model nlogit	2.134 (0.492)***
Model mixed logit	1.977 (0.517)***
Model clogit	0.720 (0.358)**
Model "Other"	1.018 (0.555)*
Hausman-McFadden IIA Test	Base IIA Test
Small-Hsiao Test	0.945 (0.374)**
Low income country	3.326 (1.703)*
Low middle income country	2.204 (1.161)*
Upper middle income country	0.879 (0.861)
Study transport mode choice	-3.955 (2599.015)
Study residential location choice	1.055 (0.678)
Study firm location choice	0.062 (0.642)
Study environmental and natural resource valuation	0.872 (0.420)**
Study brand choice	0.480 (0.498)
Study health care/medicare choice	1.091 (0.484)**
Study tv telephone and Internet choice	2.616 (0.747)***
Study employment choice	0.789 (0.467)*
Study school choice	(omitted)
Study voting and political choice	-5.436 (1546.363)
Number of observations	293
Log likelihood	-112.955
Number of groups	137
Wald chi2(23)	41.80
Prob > chi2	0.0096
Observations per group:	
min	1
avg	2.8
max	8

Table 8.5: Random-Effects Probit Regression Results – All Variables

Dependent Variable: 0=Accept IIA; 1= Reject IIA

*Legend: * p < 0.05 ** p < 0.01 *** p < 0.001*

The first model estimates a binary probit IIA test estimation, using a dummy variable which include rejection and acceptance of the IIA assumption. In table 8.5 we provide results from a model that include all variables. This is to give an idea of the model selection process. We considered the statistical significance of these variables and arrived at the model with variables which are significant. This method has precedence (Card and Krueger, 1995).

Our random-effects probit model results (table 8.6) indicate that Hausman-McFadden (relative to Small-Hsiao) IIA test; TV, Internet and telephone; and environmental valuation choice models; number of alternatives are statistically significant. Hausman-McFadden and TV, Internet and telephone choice models are more likely to accept IIA. Studies with large number of alternatives are more likely to pass the IIA test. The two models, whole dataset and H-M test sub-sample converge at -148.89 and -117.51 respectively.

We could not establish publication bias because both journal impact factor and dummy for publication are not significant. Number of alternatives, dummies for residential location choice (and several other) models and Post-McFadden Nobel Prize are also not significant. We observe from the probit model, the number of alternatives is mildly significant in the Hausman-McFadden sub-sample. But this is not conclusive because most of the studies in this sample have three alternatives.

Our ordered probit model converges at -316. 78 log likelihood. Because most of the independent variables in meta-regression analysis are not causal variables, they are control variables, $Pseudo\ R^2$ from the meta-analysis regression is likely to be low. Low R^2 is also typical of microeconometrics models. The $Pseudo\ R^2$ from our model is 0.15. The detailed results are summarized in table 8.7.

The following variables are statistically significant. Sample size; log of per-capita income: the following choice models - mixed logit, nested logit, and multinomial probit models; Small-Hsio IIA (relative to Hausman-McFadden IIA) test; residential choice studies; environment and natural resource studies; health care and medicare choice studies; and employment choice.

Explanatory Variables	*Coef. (Std. Err.)*	
	Model 1 ***Whole Dataset***	**Model 2** ***H-M test Sub-Sample***
Constant	-5.2466 (2.4577)	-2.5676 (1.2708)
Log n	0.4762 (0.2760)	-0.1905 (0.1250)
Journal Impact Factor	0.0830 (0.1068)	0.0375 (0.1659)
TV, Telelephone and Internet Choice	3.9504 (1.3898)*	3.9673 (1.4244)**
Published	-	1.0064 (0.5985)
Residential Location Choice	0.9145 (1.1293)	0.2766 (1.0801)
Hausman-McFadden IIA test	4.8417 (2.5090)*	-
Small-Hsiao IIA Test (Base - normalised to zero)	-	-
Environment and Natural Resource Valuation Choice	-	1.5390 (0.6290)*
H-M IIA test-Sample size	-0.7038 (0.3116)*	-
Alternatives	-	0.3624 (0.1534)*
Post-McFadden Nobel Prize	0.4645 (0.6933)	-
Number of observations	323	273
Log likelihood	-148.89	-117.51
Number of groups	146	135
Obs per group:		
min	1	1
avg	2.2	2.0
max	13	13

Table 8.6: Random Effects Probit Results - Dependent Variable: 0=Accept IIA; 1= Reject IIA

*Legend: * p < 0.05 ** p < 0.01 *** p < 0.001*

Our results indicate that the larger the sample size the higher possibility of IIA being accepted. We must point out, once again, that this is not IIA power size test. Studies conducted in high income countries are more likely to reject the IIA assumption relative to low income countries. Three choice models are highly statically significant. Studies that use multinomial probit, nested logit, mixed ligit are more likely to report IIA rejection. This is not something new. Except for nested logit, where sequential decision may be the compelling reason for its usage, mixed logit and multinomial probit are largely employed to remedy the violation of the IIA assumption. Four categories of choice studies are more likely to reject IIA.

Environment and natural resource studies, health care and medicare choice studies, and employment choice are highly significant. Residential choice studies variable is mildly significant and other three choices more likely to reject the IIA assumption. This is contrary to what we expect. One of the motivations of this study is to empirically asses what appears to be a consensus in the literature that residential choice studies do not suffer from the problem of IIA assumption.

Explanatory Variables	***Coeff. (Std Error)***
Published	0.071 (0.257)
Jimfact	0.132 (0.077)
Log n	0.140 (0.060)**
Log per-capita income	-0.985 (0.432)**
Number of Alternatives	-0.034 (0.089)
Model mprobit	-1.806 (0.636)***
Model nlogit	-1.939 (0.294)***
Model mixed logit	-1.147 (0.343)***
Model clogit	-0.300 (0.292)
Model “other”	0.263 (0.344)
Hausman-McFadden IIA Test	Base IIA Test
Small-Hsiao IIA Test	-1.258 (0.245)***
Low income country	-2.350 (1.474)
Low middle income country	-1.403 (0.941)
Upper middle income country	-0.279 (0.687)
Study transport mode choice	(Omitted)
Study residential location choice	-1.186 (0.675)*
Study firm location choice	0.358 (0.550)
Study environmental and natural resource valuation	-0.960 (0.295)***
Study brand choice	-0.074 (0.400)
Study health care/medicare choice	-1.215 (0.398)***
Study tv telephone and Internet choice	-0.727 (0.622)
Study employment choice	-1.339 (0.355)***
Study school choice	(Omitted)
Study voting and political choice	-0.702 (0.613)
Number of observations	197
LR $Chi^2(22)$	-316.777
Log likelihood	114.02
Prob > Chi^2	0.000
Pseudo R^2	0.153

Table 8.7: Ordered Probit Regression of IIA P-Values

*Legend: * p < 0.05 ** p < 0.01 *** p < 0.001*

In line with previous Monte-Carlo studies on the two most popular IIA tests, our result indicates that Small-Hsio IIA test is highly significant. Studies that use Small-Hsio test are more likely to reject IIA relative to the Hausman-McFadden IIA test. This is also consistent with our random-effects probit regression of accept/reject.

8.9: Publication Bias and Probability of Publishing

Using dummy to adjusting for the bias in the model or including study type such OLS, GLS, cross-section and panel data. It is anticipated that panel data studies are more likely to be published relative to cross-section studies and OLS is less likely to be published relative to other more 'sophisticated' models. This method is problematic because other models have different IIA profile, we there opted to use a dummy for publication and IDEAS/RePEc 2007 Journal Impact factor, a universal ranking of economics journals.

In addition to correcting for this bias, we also estimate the probability of publishing a discrete choice study from our data set. It would be interesting to estimate the probability of publication because publication is not a causal variable, it does not affect the outcome of the IIA test, it is an endogenous variable.

In the probit model in which we estimate the probability of publishing a study, the dummy for recent studies (awaiting publication), Hausman-McFadden IIA test, multinomial probit and conditional logit models, firm location studies, environment and natural resource valuation, number of alternatives, multiple authors, are statistically significant. The log likelihood in our two models are -186.83 and -191.8227, while the Pseudo R^2, are 0.11 and 0.18 respectively. Both Pseudo R^2 are very low, typical of discrete choice models. Detailed results are provided in table 8.8.

Studies that estimate multinomial probit (a better or more superior model) are more likely to be published. Studies by more than one author, using large sample, using the Hausman-McFadden test, and providing more information, are more likely to be published. Also, studies with a smaller number of alternatives in the choice set are more likely to be published, and those with exactly three alternatives (the minimum number) particularly so. Of the study types, firm location choice studies are the most likely to be published, while studies of environment and natural resource valuation are the least likely to be published. Multiple authors increase the chance of publication, confirming the benefits of many minds working together.

Explanatory Variables	*Coef. (Std. Err.)*	
	Model 1	**Model 2**
Constant	-2.5102 (0.4881)	-1.1075 (0.4347)
Multinomial Probit model	1.7835 (0.4922)***	1.4299 (0.4978)**
Nested logit model	0.1047 (0.2519)	-0.0009 (0.2347)
Mixed logit model	-0.0559 (0.3195)	0.1314 (0.3034)
Conditional logit model	-0.6621 (0.2764)*	-0.6149 (0.2703)*
Recent Studies (Waiting Publication)	0.6972 (0.1715)***	-
Number of Alternatives	-	-0.1633 (0.0663)*
3 Alternatives	0.5964 (0.1714)**	-
Logn	0.1065 (0.0464)*	0.1183 (0.0456)**
Multiple Authors	0.4123 (0.1837)*	-
Hausman-McFadden IIA test	0.7827 (0.2602)**	0.7354 (0.2411)**
Residential Location Choice	0.8234 (0.5135)	1.0988 (0.5297)
Firm Location Choice	0.9444 (0.3768)*	1.0302 (0.4216)*
Environment and Natural Resource Valuation Choice	-0.5209 (0.2451)*	-0.5620 (0.2285)*
Brand Choice	-0.1521 (0.3388)	-0.2358(0.3304)
TV, Telelephone and Internet	0.1108 (0.4036)	0.4890 (0.3867)
Voting and Political decision	-04235 (0.4347)	-0.2076 (0.4216)
Employment and Job Choice	-0.2848 (0.2480)	-0.4619 (0.2433)*
Number of observations	327	324
Log likelihood	-173.53	-186.83
Pseudo R^2	0.18	0.11

Table 8.8: Publication Probit regression Results – 0 Published; 1 Unpublished

*Legend: * p < 0.05 ** p < 0.01 *** p < 0.001*

8.10: Conclusions to Chapter Eight

We point out above that one of our meta-analysis models, the randon effect model employs a panel data regression because we have multiple observations per study. Florax (2002) introduces this issue, indirectly, in an attempt to explain the impact of between-study and within-study autocorrelation on meta-analysis results. Our meta-analysis is therefore unique because, previous meta-analysis empirical studies have not dealt this problem.

Although we have established that Small-Hsiao test is more likely (relative to Hausman-McFadden test) to reject IIA (previous studies have reported similar results) we are unable to say whether this is a good or a bad thing. We can not conclude on the basis of this analysis, which test is better.

However, it would be sensible to conduct both tests and reject IIA if it is rejected by the Hausman-McFadden test and accept IIA if it accepted by the Small-Hsiao test. Perhaps it is time we scrutinise available (and explore more) alternative IIA test procedures.

Because we had to discard several studies for lack of sufficient information (this applies to both published and unpublished studies) we recommend authors to provide detailed statistics to provide in each study.

Chapter Nine: Summary, conclusion and recommendations

9.1 Summary and Conclusions

This research consist of nine chapters. The first chapter is the introduction, and the current chapter, is the conclusion. Chapters two, three and four are essentially literature review chapters. Chapter two is a theoretical review of the dynamics of housing market - demand and supply of residential housing; chapter three a theoretical and mathematical derivation of probabilistic choice models; and chapter four, a review of the theory and estimation procedure of the hedonic pricing model.

Two of our three empirical studies are based on primary data on housing location decisions in Kano Nigeria. Chapter five introduces the housing location choice primary data collection and provides a summary statistics of some of the variables. Chapters six and seven are the residential housing location empirical studies.

The data comprise both household and location attributes which enable us to estimate two complementary econometric models, hedonic pricing and discrete choice models. The main objective of this component is to estimate WTP for two utilities public water and electricity supply. The third empirical chapter (chapter eight) is a meta-analysis of the IIA assumption in discrete choice studies.

We collected primary data in Kano in November 2006. About 3,000 questionnaires were administered on heads of households living in rented housing. Because the data is interval data (because of sensitive nature of some the questions), we estimate the hedonic pricing model using the interval regression method, this is both consistent with theory and precedence. We report households WTP for water and electricity supply. Our results confirm that household's WTP for both water and electricity supply reflected in higher rent in areas with longer hours of supply of these two utilities in the study area. We are treating changes in individual WTP as a proxy for the utility/welfare improvements from derived increase in hours of water and electricity supply.

To estimate WTP for utilities in discrete choice model we to have use the location attributes and a price variable. The alternative specific conditional logit, mixed logit, nested-logit, alternative specific multinomial probit and mixed logit, mixed probit models which are suited for this type of analysis have been estimated.

When we allow for the satiation point, households seems to attach more priority to water supply. This is because, although there are alternatives to both public water and electricity supply, without taking account for satiation, electricity is valued more because, it has no satiation point, it is demanded at all time, for domestic use and by (both small and large scale) businesses, and most important, because it cannot be stored relative to water. We acknowledge that, in theory it is possible to use batteries/accumulators to store energy on a small scale.

	Hedonic Pricing Model (model 1) *(Estimated Welfare improvement)*	***Alternative Specific Multinomial Probit Model*** *(Estimated WTP)*	***Mixed Probit Model*** *(Estimated WTP)*
Water Supply	*3,459*	*451.90*	*656.90*
Electricity Supply	*1019*	*458.00*	*396.70*

Table 9.1: Estimated WTP/Welfare Improvement (in Naira) for one additional daily hour of Water and Electricity Supply, for a period of one year.

Table 9.1 summarises the results of WTP of interest from various models estimated in this thesis. Although the hedonic pricing model and the mixed probit approaches are very different, the estimates of WTP obtained from the choice models and the estimates of total welfare improvement from the hedonic pricing model appear roughly consistent. We would expect the estimates from the hedonic pricing model to be higher because Bartik (1988) argues that hedonic pricing model usually gives an *upper bound* of the benefits from welfare improvements for housing attributes. See section 4.8 for details. Our estimates from the hedonic pricing model are always higher than the discrete choice models, which is consistent with them being the *upper bound*.

The WTP for both water supply and electricity supply from the Mixed probit model appears to be the highest of all the estimated choice models. This could be due to adjustment for household characteristics given the importance of these utilities for life.

Estimates from the hedonic pricing model are higher than the estimates from the discrete choice models. The WTP for extra one hour supply water and electricity from the hedonic pricing model is 3,459 Naira and 1, 019 Naira respectively. While the estimated WTP from the Mixed logit is 656.90 Naira for water and 396.70 Naira for electricity.

We consider these estimates of WTP to be reasonable. Overall they provide convincing evidence of the importance attached to water supply and electricity supply by urban residents in a developing country. It is clear from these results that any improvements in these services have the potential to increase overall welfare considerably.

We also undertook a cost-benefit analysis (welfare change benefits) of increase in water supply using hypothetical policy scenarios and their relative costs, from our dataset.

The first type of policy is to increase water supply for every household by a certain (absolute) number of hours per day. The second is to set a minimum acceptable number of hours per day, and then to ensure that every household is brought up at least to that level. We make the straightforward assumption of constant marginal cost.

The first policy appears to be less beneficial than most of the “minimum x hours” policies, even in terms of welfare change per additional hour. This is because some household already enjoy high hours of supply, and these households benefit little, or not at all, from additional hours of water supply. The policy is ineffective because it increases the supply of all households, including those whose welfare improves little or not at all.

We are able to conduct this research because the city under study is located in a developing country. Water and electricity supply is taken for granted in developed countries. In a developed country WTP for utilities can be sourced by stated preference study. We avoided stated preference by targeting actual decisions because, as pointed out before, the approach has two major problems. Individuals may not know the value of these utilities and there is incentive to report wrong valuation in an attempt to influence certain policy outcome.

Chapter eight, the third and final empirical chapter is a meta-analysis of the IIA assumption in discrete choice models. The theory says that the assumption is required otherwise the predictive power of the model is compromised. This implies that in some cases the IIA is problematic. We therefore attempt to find a pattern and the probability of accepting or rejecting the IIA assumption. For this purpose we collected 181 published and unpublished (in order to control for publication bias) discrete choice studies which report IIA test results and other variables of interest.

We use reported p-values (between zero and one) and estimate ordered probit model of the p-values. This model is complemented by a simple binary probit model (acceptance/rejection) of the IIA assumption. In both models (ordered probit and random effects probit models) we use the attributes of the IIA studies in the sample as explanatory variables.

Our results indicate that it is more likely accept IIA if the Hausman-McFadden test is used for the IIA test relative to Small-Hsiao test for IIA. Our results are similar to results reported from Monte Carlo studies of the IIA assumption. However, we are unable to say whether this it is a good or a bad thing because we did not test for power and size properties of the two most popular IIA tests.

Finally, using the IIA meta-analysis dataset, we estimate the probability of publishing a discrete choice study. This is a simple binary probit model of published and unpublished, with attributes of studies as the right hand side variables.

9.2 Recommendation

We make three specific recommendations from our results and data collection experience in this thesis.

9.2.1 Data Collection and Publication

Some of the articles we collected lack of basic information about their research. This is common to both published and unpublished studies. It is a disturbing if applied econometric study could be published in academic journal without details on the data and basic statistics. We strongly recommend authors and editors to ensure that detailed statistics is provided in each study.

We spent long time cleaning and had to throw away several observation in our household residential location data. Researchers should be more careful and closely monitor data collection process when administering questionnaire to reduce the incidence of "curbstoning" and improve the quality of research.

A possible practical solution to the potential problem of data quality would be to collect a "reasonable" but representative sample, which would be easier to monitor, and use Monte Carlo methods to check if the results are sensitive to sample size.

9.2.2 Investment in the Provision of Utilities in Kano

Results from both discrete choice and hedonic pricing model show that there is positive relationship between rent and number of hours of water and electricity supply. After adjusting for other factors that affect rent, the results indicate positive WTP for daily increase in hours of water and electricity supply.

This is good news for the government who could generate more investment, increase water rates and electricity price to improve supply and standard. It is also possible, given these results, to attract private sector investment by the government in the provision of public utilities either through collaboration or full privatisation of the supply of public utilities.

We strongly recommend the public-private sector initiative in the provision of water supply. We believe that, complete privatisation of water supply in Kano is not likely to work because of the importance of water for life and other socio-cultural factors. The private sector, would be interested in investing in the provision of public utilities given huge the market potentials and high WTP for both water and electricity. Appropriate incentives such as tax holiday, loan guarantee, import duty rebate/waiver, should be provided to the private sector in a tripartite collaboration between the private sector, federal and state governments.

Immediate improvement in the provision of electricity or the use of alternative energy sources for the water supply company would increase water supply by increasing consistency, wider coverage and supply pressure.

Above all government should embark on campaign to make households pay for utilities and strictly implement its town planning control policy. This is important to forecast demand for public utilities (and other municipal services).

9.2.3 IIA Assumption and Discrete Choice Models

Although we can not conclude on the basis of this study, which IIA test is better, we recommend that it would be sensible to conduct both Hausman-McFadden and Small-Hsiao IIA test and reject the IIA assumption if it is rejected by the Hausman-McFadden test and accept IIA if it accepted by the Small-Hsiao test.

We are aware that a number of alternative IIA test procedures exit, although they have not been critically scrutinized. These IIA test procedures should be critically examined to see if they are better than the most widely used test procedures and be made available in the major econometric software packages.

A longer term solution would be to come up with simpler and more flexible choice models. This is because, all other available alternative choice models (relative to the multinomial logit and conditional models) are either computationally difficult (probit model with more than four alternatives), based on complicated theoretical structures (random parameter – mixed logit model) or based on hypothetical sequential choice structure (nested model).

Appendices

Appendix I: Hedonic Pricing Model Results - Detailed Results: Computer Output

Model 1 Hedonic Pricing Model - Linear Model

```
intreg lrent urent nubdrooms nubdrooms2 nubtoilets housefloorarea gardenyard hourselectricity
hourswater hourswater2 privateschools publicschools market highway airport industries

Fitting constant-only model:
Iteration 0:   log likelihood = -4365.9505
Iteration 1:   log likelihood = -3949.5903
Iteration 2:   log likelihood = -3928.8873
Iteration 3:   log likelihood = -3928.8784
Iteration 4:   log likelihood = -3928.8784

Fitting full model:
Iteration 0:   log likelihood = -4042.2322
Iteration 1:   log likelihood = -3549.4562
Iteration 2:   log likelihood = -3535.3154
Iteration 3:   log likelihood = -3535.2672
Iteration 4:   log likelihood = -3535.2672

Interval regression                               Number of obs   =       2272
                                                  LR chi2(14)     =     787.22
Log likelihood = -3535.2672                       Prob > chi2     =     0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   nubdrooms |   29.04235   3.389503     8.57   0.000     22.39905    35.68566
  nubdrooms2 |  -3.717588   .4683703    -7.94   0.000    -4.635576   -2.799599
  nubtoilets |   15.09394   1.665728     9.06   0.000     11.82917     18.3587
housefloor~a |   .0304861   .0040548     7.52   0.000     .0225389    .0384334
  gardenyard |   .0051649   .0035008     1.48   0.140    -.0016966    .0120265
hourselect~y |   1.019021    .267018     3.82   0.000     .4956749    1.542366
  hourswater |   5.817288   .4787975    12.15   0.000     4.878862    6.755714
 hourswater2 |  -.2358117   .0245883    -9.59   0.000    -.2840038   -.1876195
privatesch~s |   10.89598   1.897533     5.74   0.000     7.176889    14.61508
publicscho~s |  -.9530223   1.884105    -0.51   0.613    -4.645801    2.739756
      market |  -3.509759   1.988535    -1.76   0.078    -7.407217    .3876989
     highway |   2.420858   2.104472     1.15   0.250    -1.703832    6.545548
     airport |  -8.919855   3.057703    -2.92   0.004    -14.91284   -2.926868
  industries |  -7.910916   2.853916    -2.77   0.006    -13.50449   -2.317344
       _cons |  -142.8197   11.69482   -12.21   0.000    -165.7411   -119.8983
-------------+----------------------------------------------------------------
    /lnsigma |    3.57302   .0236852   150.85   0.000     3.526597    3.619442
-------------+----------------------------------------------------------------
       sigma |     35.624   .8437626                      34.00805    37.31673
------------------------------------------------------------------------------
  Observation summary:      1170  left-censored observations
                               0     uncensored observations
                              17 right-censored observations
                            1085       interval observations

fitstat
Measures of Fit for intreg of lrent urent
Log-Lik Intercept Only:    -3928.878     Log-Lik Full Model:        -3535.267
D(2256):                    7070.534     LR(14):                      787.223
                                         Prob > LR:                     0.000
McFadden's R2:                 0.100     McFadden's Adj R2:             0.096
Maximum Likelihood R2:         0.293     Cragg & Uhler's R2:            0.293
McKelvey and Zavoina's R2:     0.184
Variance of y*:             1555.505     Variance of error:          1269.069
AIC:                           3.126     AIC*n:                      7102.534
BIC:                      -10364.772     BIC':                       -679.025
```

Model 2 Hedonic Pricing Model - Log Model

```
intreg loglrent logurent nubdrooms nubdrooms2 nubtoilets housefloorarea gardenyard
hourselectricity hourswater hourswater2 privateschools publicschools market highway airport
industries

Fitting constant-only model:
Iteration 0:   log likelihood = -4204.3052
Iteration 1:   log likelihood = -3726.8809
Iteration 2:   log likelihood = -3686.4687
Iteration 3:   log likelihood = -3686.3788
Iteration 4:   log likelihood = -3686.3788

Fitting full model:
Iteration 0:   log likelihood = -3816.3979
Iteration 1:   log likelihood = -3266.6891
Iteration 2:   log likelihood = -3240.6377
Iteration 3:   log likelihood = -3240.5963
Iteration 4:   log likelihood = -3240.5963

Interval regression                               Number of obs   =       2272
                                                  LR chi2(14)     =     891.57
Log likelihood = -3240.5963                       Prob > chi2     =     0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   nubdrooms |   .4807823   .0505575     9.51   0.000     .3816915    .5798732
  nubdrooms2 |  -.0613082   .0069889    -8.77   0.000    -.0750061   -.0476103
  nubtoilets |   .2408877   .0248812     9.68   0.000     .1921214     .289654
housefloor~a |   .0004722   .0000606     7.79   0.000     .0003534    .0005911
  gardenyard |   .0000725   .0000525     1.38   0.167    -.0000303    .0001754
hourselect~y |   .0151144   .0039931     3.79   0.000     .0072881    .0229408
  hourswater |   .0965738   .0071418    13.52   0.000     .0825761    .1105715
 hourswater2 |  -.0038998    .000367   -10.62   0.000    -.0046191   -.0031804
privatesch~s |    .170573   .0283297     6.02   0.000     .1150478    .2260982
publicscho~s |  -.0135254   .0281543    -0.48   0.631    -.0687068     .041656
      market |  -.0435947    .029718    -1.47   0.142    -.1018409    .0146515
     highway |   .0584425   .0314174     1.86   0.063    -.0031345    .1200194
     airport |   -.143048   .0455894    -3.14   0.002    -.2324016   -.0536944
  industries |  -.1183926   .0426132    -2.78   0.005     -.201913   -.0348723
       _cons |   .7176091   .1743993     4.11   0.000     .3757927    1.059425
-------------+----------------------------------------------------------------
    /lnsigma |  -.6222364   .0236162   -26.35   0.000    -.6685234   -.5759495
-------------+----------------------------------------------------------------
       sigma |   .5367427   .0126758                      .5124647    .5621709
------------------------------------------------------------------------------
  Observation summary:      1170  left-censored observations
                               0     uncensored observations
                              17 right-censored observations
                            1085       interval observations

fitstat

Measures of Fit for intreg of loglrent logurent
Log-Lik Intercept Only:    -3686.379     Log-Lik Full Model:        -3240.596
D(2256):                    6481.193     LR(14):                      891.565
                                         Prob > LR:                     0.000
McFadden's R2:                 0.121     McFadden's Adj R2:             0.117
Maximum Likelihood R2:         0.325     Cragg & Uhler's R2:            0.325
McKelvey and Zavoina's R2:     0.205
Variance of y*:                0.363     Variance of error:             0.288
AIC:                           2.867     AIC*n:                      6513.193
BIC:                      -10954.113     BIC':                       -783.367
```

Model 3 Hedonic Pricing Model - Linear Model with Area Dummy Variables

```
intreg lrent urent nubdrooms nubdrooms2 nubtoilets housefloorarea gardenyard hourselectricity
hourswater hourswater2 privateschools publicschools market highway industries Lowden Airport2
Industry1 Industry2 Other

Fitting constant-only model:
Iteration 0:   log likelihood = -4365.9505
Iteration 1:   log likelihood = -3949.5903
Iteration 2:   log likelihood = -3928.8873
Iteration 3:   log likelihood = -3928.8784
Iteration 4:   log likelihood = -3928.8784
```

```
Fitting full model:
Iteration 0:   log likelihood = -3977.5868
Iteration 1:   log likelihood = -3480.3937
Iteration 2:   log likelihood = -3466.5217
Iteration 3:   log likelihood = -3466.4782
Iteration 4:   log likelihood = -3466.4782

Interval regression                               Number of obs   =       2272
                                                  LR chi2(18)     =     924.80
Log likelihood = -3466.4782                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   nubdrooms |   27.36644   3.250006     8.42   0.000     20.99654    33.73633
  nubdrooms2 |  -3.357759   .4484147    -7.49   0.000    -4.236635   -2.478882
  nubtoilets |   12.81197   1.597981     8.02   0.000     9.679981    15.94395
housefloor~a |    .022045   .0038657     5.70   0.000     .0144684    .0296215
  gardenyard |   .0039174     .00337     1.16   0.245    -.0026877    .0105224
hourselect~y |   .6260267   .2588702     2.42   0.016     .1186505    1.133403
  hourswater |   5.745548   .4584262    12.53   0.000     4.847049    6.644047
 hourswater2 |  -.2370152   .0238322    -9.95   0.000    -.2837255   -.1903049
privatesch~s |   9.257512   1.856905     4.99   0.000     5.618046    12.89698
publicscho~s |    2.72857   1.872244     1.46   0.145    -.9409618    6.398101
      market |  -1.086459   1.935233    -0.56   0.575    -4.879447    2.706529
     highway |  -2.544452   2.099577    -1.21   0.226    -6.659548    1.570643
  industries |  -7.363363   2.992238    -2.46   0.014    -13.22804   -1.498685
      Lowden |   36.20585   3.050543    11.87   0.000     30.22689     42.1848
    Airport2 |   2.951801   3.386513     0.87   0.383    -3.685643    9.589244
   Industry1 |   7.809896   5.438956     1.44   0.151    -2.850261    18.47005
   Industry2 |   10.76733   3.306075     3.26   0.001     4.287545    17.24712
       Other |   12.95492   2.311252     5.61   0.000      8.42495    17.48489
       _cons |  -125.5573   11.22717   -11.18   0.000    -147.5622   -103.5525
-------------+----------------------------------------------------------------
    /lnsigma |   3.521645   .0235865   149.31   0.000     3.475416    3.567873
-------------+----------------------------------------------------------------
       sigma |   33.84004   .7981681                      32.31127    35.44115
------------------------------------------------------------------------------
  Observation summary:      1170  left-censored observations
                               0     uncensored observations
                              17 right-censored observations
                            1085       interval observations

fitstat

Measures of Fit for intreg of lrent urent
Log-Lik Intercept Only:    -3928.878     Log-Lik Full Model:        -3466.478
D(2252):                    6932.956     LR(18):                      924.800
                                         Prob > LR:                     0.000
McFadden's R2:                 0.118     McFadden's Adj R2:             0.113
Maximum Likelihood R2:         0.334     Cragg & Uhler's R2:            0.334
McKelvey and Zavoina's R2:     0.218
Variance of y*:             1464.977     Variance of error:          1145.148
AIC:                           3.069     AIC*n:                      6972.956
BIC:                      -10471.436     BIC':                       -785.689
```

Model 4 Hedonic Pricing Model - Log Model with Area Dummy Variables

```
intreg loglrent logurent nubdrooms nubdrooms2 nubtoilets housefloorarea gardenyard
hourselectricity hourswater hourswater2 privateschools publicschools market highway Lowden
Airport2 Industry1 Industry2 Other

Fitting constant-only model:
Iteration 0:   log likelihood = -4204.3052
Iteration 1:   log likelihood = -3726.8809
Iteration 2:   log likelihood = -3686.4687
Iteration 3:   log likelihood = -3686.3788
Iteration 4:   log likelihood = -3686.3788

Fitting full model:
Iteration 0:   log likelihood = -3755.7843
Iteration 1:   log likelihood = -3202.1074
Iteration 2:   log likelihood =   -3176.66
Iteration 3:   log likelihood = -3176.6187
Iteration 4:   log likelihood = -3176.6187
```

```
Interval regression                               Number of obs   =       2272
                                                  LR chi2(17)     =    1019.52
Log likelihood = -3176.6187                       Prob > chi2     =     0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   nubdrooms |   .4570443   .0485793     9.41   0.000     .3618306    .5522581
  nubdrooms2 |  -.0562497   .0067027    -8.39   0.000    -.0693867   -.0431127
  nubtoilets |   .2091164   .0238908     8.75   0.000     .1622913    .2559415
housefloor~a |   .0003578   .0000579     6.18   0.000     .0002442    .0004714
  gardenyard |   .0000445   .0000505     0.88   0.378    -.0000544    .0001435
hourselect~y |   .0093613   .0038739     2.42   0.016     .0017686     .016954
  hourswater |   .0961199   .0068405    14.05   0.000     .0827128     .109527
 hourswater2 |  -.0039904   .0003541   -11.27   0.000    -.0046844   -.0032964
privatesch~s |   .1398349   .0276836     5.05   0.000     .0855761    .1940938
publicscho~s |   .0445454   .0279953     1.59   0.112    -.0103243    .0994151
      market |   -.014112   .0289007    -0.49   0.625    -.0707564    .0425325
     highway |   -.021445   .0313402    -0.68   0.494    -.0828708    .0399807
      Lowden |   .5350271   .0456143    11.73   0.000     .4456246    .6244296
    Airport2 |   .0510985   .0506004     1.01   0.313    -.0480764    .1502735
   Industry1 |   .0553313   .0769728     0.72   0.472    -.0955327    .2061953
   Industry2 |   .1549114   .0476247     3.25   0.001     .0615688     .248254
       Other |   .2103541   .0345083     6.10   0.000      .142719    .2779892
       _cons |   .9304553   .1677737     5.55   0.000     .6016249    1.259286
-------------+----------------------------------------------------------------
    /lnsigma |  -.6721475   .0235622   -28.53   0.000    -.7183286   -.6259663
-------------+----------------------------------------------------------------
       sigma |   .5106109   .0120311                      .4875665    .5347444
------------------------------------------------------------------------------
  Observation summary:      1170  left-censored observations
                               0     uncensored observations
                              17 right-censored observations
                            1085       interval observations

fitstat

Measures of Fit for intreg of loglrent logurent
Log-Lik Intercept Only:    -3686.379    Log-Lik Full Model:        -3176.619
D(2253):                    6353.237    LR(17):                     1019.520
                                        Prob > LR:                     0.000
McFadden's R2:                 0.138    McFadden's Adj R2:             0.133
Maximum Likelihood R2:         0.362    Cragg & Uhler's R2:            0.362
McKelvey and Zavoina's R2:     0.236
Variance of y*:                0.341    Variance of error:             0.261
AIC:                           2.813    AIC*n:                      6391.237
BIC:                      -11058.883    BIC':                       -888.137
```

Hedonic Pricing Model – Rejected Linear and Log Models

Electricity Supply and Electricity Squared both included, both coefficients with (a priori) wrong signs.

```
intreg lrent urent nubdrooms nubdrooms2 nubtoilets housefloorarea gardenyard  hourselectricity
hourselectricity2 hourswater hourswater2 privateschools publicschools  airport market highway
industries

Fitting constant-only model:
Iteration 0:   log likelihood = -4365.9505
Iteration 1:   log likelihood = -3949.5903
Iteration 2:   log likelihood = -3928.8873
Iteration 3:   log likelihood = -3928.8784
Iteration 4:   log likelihood = -3928.8784

Fitting full model:
Iteration 0:   log likelihood = -4035.2559
Iteration 1:   log likelihood = -3543.2491
Iteration 2:   log likelihood = -3529.1884
Iteration 3:   log likelihood = -3529.1412
Iteration 4:   log likelihood = -3529.1412

Interval regression                               Number of obs   =       2272
                                                  LR chi2(15)     =     799.47
Log likelihood = -3529.1412                       Prob > chi2     =     0.0000
```

```
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   nubdrooms |   28.74242   3.376353     8.51   0.000     22.12489    35.35995
  nubdrooms2 |  -3.694597   .4665618    -7.92   0.000    -4.609041   -2.780153
  nubtoilets |   14.84614   1.660753     8.94   0.000     11.59113    18.10116
housefloor~a |   .0305535   .0040437     7.56   0.000      .022628    .0384791
  gardenyard |   .0051106   .0034917     1.46   0.143     -.001733    .0119541
hourselect~y |  -1.375744   .7316947    -1.88   0.060    -2.809839    .0583513
hourselect~2 |   .1866887   .0531742     3.51   0.000     .0824691    .2909082
  hourswater |   5.909694   .4781831    12.36   0.000     4.972472    6.846916
 hourswater2 |  -.2397711   .0245367    -9.77   0.000    -.2878622   -.1916801
privatesch~s |   10.77961   1.890857     5.70   0.000     7.073598    14.48562
publicscho~s |  -.4969419   1.881307    -0.26   0.792    -4.184236    3.190352
     airport |  -9.749615   3.052439    -3.19   0.001    -15.73228   -3.766945
      market |  -2.916612   1.985979    -1.47   0.142     -6.80906    .9758361
     highway |   2.998971   2.102077     1.43   0.154    -1.121023    7.118966
  industries |  -7.186857   2.850508    -2.52   0.012    -12.77375   -1.599964
       _cons |  -137.7207    11.7188   -11.75   0.000    -160.6891   -114.7522
-------------+----------------------------------------------------------------
    /lnsigma |   3.568642   .0236867   150.66   0.000     3.522217    3.615067
-------------+----------------------------------------------------------------
       sigma |    35.4684   .8401293                      33.85942    37.15385
------------------------------------------------------------------------------

  Observation summary:      1170  left-censored observations
                               0     uncensored observations
                              17 right-censored observations
                            1085       interval observations
```

```
intreg loglrent logurent nubdrooms nubdrooms2 nubtoilets housefloorarea gardenyard
hourselectricity hourselectricity2 hourswater hourswater2 privateschools publicschools
airport market highway industries

Fitting constant-only model:
Iteration 0:   log likelihood = -4204.3052
Iteration 1:   log likelihood = -3726.8809
Iteration 2:   log likelihood = -3686.4687
Iteration 3:   log likelihood = -3686.3788

Fitting full model:
Iteration 0:   log likelihood = -3810.2863
Iteration 1:   log likelihood = -3261.0402
Iteration 2:   log likelihood = -3234.9213
Iteration 3:   log likelihood = -3234.8796

Interval regression                               Number of obs   =       2272
                                                  LR chi2(15)     =     903.00
Log likelihood = -3234.8796                       Prob > chi2     =     0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   nubdrooms |   .4766257   .0503809     9.46   0.000     .3778809    .5753704
  nubdrooms2 |  -.0610048   .0069646    -8.76   0.000    -.0746551   -.0473545
  nubtoilets |   .2373861   .0248162     9.57   0.000     .1887473     .286025
housefloor~a |   .0004734   .0000605     7.83   0.000     .0003549    .0005919
  gardenyard |   .0000718   .0000524     1.37   0.170    -.0000309    .0001744
hourselect~y |  -.0195603   .0109723    -1.78   0.075    -.0410657     .001945
hourselect~2 |   .0027077    .000799     3.39   0.001     .0011418    .0042736
  hourswater |   .0979537   .0071357    13.73   0.000      .083968    .1119394
 hourswater2 |  -.0039587   .0003664   -10.80   0.000    -.0046768   -.0032405
privatesch~s |   .1691277   .0282394     5.99   0.000     .1137795    .2244759
publicscho~s |  -.0070485   .0281217    -0.25   0.802     -.062166    .0480691
     airport |   -.155167   .0455371    -3.41   0.001     -.244418   -.0659159
      market |  -.0351631   .0296913    -1.18   0.236     -.093357    .0230307
     highway |   .0667463    .031395     2.13   0.034     .0052134    .1282793
  industries |  -.1080029     .04258    -2.54   0.011    -.1914581   -.0245477
       _cons |   .7903814   .1748351     4.52   0.000      .447711    1.133052
-------------+----------------------------------------------------------------
    /lnsigma |  -.6261639   .0236159   -26.51   0.000    -.6724502   -.5798776
-------------+----------------------------------------------------------------
       sigma |   .5346388    .012626                      .5104563    .5599669
------------------------------------------------------------------------------
  Observation summary:      1170  left-censored observations
                               0     uncensored observations
                              17 right-censored observations
                            1085       interval observations
```

Hedonic Pricing Model – Robust Standard Errors – Correcting for Spatial autocorrelation

Model 5 Hedonic Pricing Model - Linear Model

```
intreg lrent urent nubdrooms nubdrooms2 nubtoilets housefloorarea gardenyard hourselectricity
hourswater hourswater2 privateschools publicschools market highway airport industries,
vce(bootstrap, cluster(area))
(running intreg on estimation sample)

Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

Interval regression                              Number of obs      =      2272
                                                 Replications       =        50
                                                 Wald chi2(14)      =   1394.05
Log likelihood = -3535.2672                      Prob > chi2        =    0.0000

                                    (Replications based on 6 clusters in area)
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   nubdrooms |   29.04235   8.575136     3.39   0.001      12.2354    45.84931
  nubdrooms2 |  -3.717588   1.277392    -2.91   0.004     -6.22123   -1.213945
  nubtoilets |   15.09394   2.149316     7.02   0.000     10.88135    19.30652
housefloor~a |   .0304861   .0073147     4.17   0.000     .0161495    .0448228
  gardenyard |   .0051649   .0038748     1.33   0.183    -.0024295    .0127594
hourselect~y |   1.019021   .2976773     3.42   0.001     .4355839    1.602457
  hourswater |   5.817288   1.323426     4.40   0.000     3.223421    8.411155
 hourswater2 |  -.2358117   .0604123    -3.90   0.000    -.3542177   -.1174057
privatesch~s |   10.89598   4.106856     2.65   0.008     2.846695    18.94527
publicscho~s |  -.9530223   3.080988    -0.31   0.757    -6.991648    5.085603
      market |  -3.509759   3.967261    -0.88   0.376    -11.28545     4.26593
     highway |   2.420858    3.69375     0.66   0.512    -4.818759    9.660475
     airport |  -8.919855   5.647595    -1.58   0.114    -19.98894    2.149227
  industries |  -7.910916   5.467365    -1.45   0.148    -18.62676    2.804923
       _cons |  -142.8197   29.29948    -4.87   0.000    -200.2457   -85.39378
-------------+----------------------------------------------------------------
    /lnsigma |    3.57302    .184829    19.33   0.000     3.210761    3.935278
-------------+----------------------------------------------------------------
       sigma |     35.624   6.584349                      24.79796    51.17637
------------------------------------------------------------------------------

  Observation summary:      1170  left-censored observations
                               0     uncensored observations
                              17 right-censored observations
                            1085       interval observations

fitstat
Measures of Fit for intreg of lrent urent
Log-Lik Intercept Only:    -3928.878    Log-Lik Full Model:        -3535.267
D(2256):                    7070.534    LR(14):                      787.223
                                        Prob > LR:                     0.000
McFadden's R2:                 0.100    McFadden's Adj R2:             0.096
Maximum Likelihood R2:         0.293    Cragg & Uhler's R2:            0.293
McKelvey and Zavoina's R2:     0.184
Variance of y*:             1555.505    Variance of error:          1269.069
AIC:                           3.126    AIC*n:                      7102.534
BIC:                      -10364.772    BIC':                       -679.025
```

Model 6 Hedonic Pricing Model - Log Model

```
intreg loglrent logurent nubdrooms nubdrooms2 nubtoilets housefloorarea gardenyard
hourselectricity hourswater hourswater2 privateschools publicschools market highway airport
industries, vce(bootstrap, cluster(area))
(running intreg on estimation sample)


Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
```

```
Interval regression                               Number of obs      =      2272
                                                  Replications       =        50
                                                  Wald chi2(14)      =   3929.71
Log likelihood = -3240.5963                       Prob > chi2        =    0.0000

                                    (Replications based on 6 clusters in area)
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   nubdrooms |   .4807823   .1259511     3.82   0.000     .2339227     .727642
  nubdrooms2 |  -.0613082   .0195533    -3.14   0.002     -.099632   -.0229844
  nubtoilets |   .2408877   .0376183     6.40   0.000     .1671573    .3146181
housefloor~a |   .0004722   .0001171     4.03   0.000     .0002428    .0007017
  gardenyard |   .0000725   .0000737     0.98   0.325     -.000072    .0002171
hourselect~y |   .0151144    .003591     4.21   0.000     .0080762    .0221527
  hourswater |   .0965738   .0215588     4.48   0.000     .0543193    .1388283
 hourswater2 |  -.0038998   .0010886    -3.58   0.000    -.0060334   -.0017661
privatesch~s |    .170573   .0526915     3.24   0.001     .0672996    .2738464
publicscho~s |  -.0135254   .0583389    -0.23   0.817    -.1278675    .1008168
      market |  -.0435947   .0640933    -0.68   0.496    -.1692153    .0820259
     highway |   .0584425   .0477181     1.22   0.221    -.0350833    .1519683
     airport |   -.143048   .0718055    -1.99   0.046    -.2837843   -.0023117
  industries |  -.1183926    .051555    -2.30   0.022    -.2194386   -.0173466
       _cons |   .7176091   .4280942     1.68   0.094    -.1214402    1.556658
-------------+----------------------------------------------------------------
    /lnsigma |  -.6222364   .1147038    -5.42   0.000    -.8470516   -.3974212
-------------+----------------------------------------------------------------
       sigma |   .5367427   .0615664                       .428677    .6720509
------------------------------------------------------------------------------

  Observation summary:      1170  left-censored observations
                               0     uncensored observations
                              17 right-censored observations
                            1085       interval observations

fitstat
Measures of Fit for intreg of loglrent logurent
Log-Lik Intercept Only:    -3686.379     Log-Lik Full Model:        -3240.596
D(2256):                    6481.193     LR(14):                      891.565
                                         Prob > LR:                     0.000
McFadden's R2:                 0.121     McFadden's Adj R2:             0.117
Maximum Likelihood R2:         0.325     Cragg & Uhler's R2:            0.325
McKelvey and Zavoina's R2:     0.205
Variance of y*:                0.363     Variance of error:             0.288
AIC:                           2.867     AIC*n:                      6513.193
BIC:                      -10954.113     BIC':                       -783.367
```

WTP Point Estimated – Using Delta Method

```
nlcom _b[hourswater] + 2*_b[hourswater2]*5

       _nl_1:  _b[hourswater] + 2*_b[hourswater2]*5

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |   3.459171   .6907615     5.01   0.000     2.105304    4.813039
------------------------------------------------------------------------------
```

Welfare Change Estimation

```
*drop w wtp w0-w8 d_w0-d_w8 wtp0-wtp8 d_wtp0-d_wtp8

gen w=hourswater
gen w0=w+1
gen w1=max(1,w)
gen  w2=max(2,w)
gen w3=max(3,w)
gen w4=max(4,w)
gen w5=max(5,w)
gen  w6=max(6,w)
gen w7=max(7,w)
gen w8=max(8,w)
```

```
gen d_w0=w0-w
gen d_w1=w1-w
gen d_w2=w2-w
gen d_w3=w3-w
gen d_w4=w4-w
gen d_w5=w5-w
gen d_w6=w6-w
gen d_w7=w7-w
gen d_w8=w8-w

gen wtp=5.817*w-0.236*w^2
gen wtp0=5.817*w0-0.236*w0^2
gen wtp1=5.817*w1-0.236*w1^2
gen wtp2=5.817*w2-0.236*w2^2
gen wtp3=5.817*w3-0.236*w3^2
gen wtp4=5.817*w4-0.236*w4^2
gen wtp5=5.817*w5-0.236*w5^2
gen wtp6=5.817*w6-0.236*w6^2
gen wtp7=5.817*w7-0.236*w7^2
gen wtp8=5.817*w8-0.236*w8^2

gen d_wtp0=max(wtp0-wtp,0)
gen d_wtp1=max(wtp1-wtp,0)
gen d_wtp2=max(wtp2-wtp,0)
gen d_wtp3=max(wtp3-wtp,0)
gen d_wtp4=max(wtp4-wtp,0)
gen d_wtp5=max(wtp5-wtp,0)
gen d_wtp6=max(wtp6-wtp,0)
gen d_wtp7=max(wtp7-wtp,0)
gen d_wtp8=max(wtp8-wtp,0)

summ d_w0-d_w8

summ d_wtp0-d_wtp8
```

Appendix II: Discrete Choice Models Results - Detailed Results: Computer Output

Conditional Logit Model

```
clogit chosenarea rent hourswater hourselectricity, group(id)
note: 22 groups (90 obs) dropped because of all positive or
      all negative outcomes.

Iteration 0:   log likelihood = -1926.4932
Iteration 1:   log likelihood = -1895.1784
Iteration 2:   log likelihood = -1894.9069
Iteration 3:   log likelihood = -1894.9068

Conditional (fixed-effects) logistic regression   Number of obs   =      14448
                                                  LR chi2(3)      =    4839.30
                                                  Prob > chi2     =     0.0000
Log likelihood = -1894.9068                       Pseudo R2       =     0.5608

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0906848   .0020874   -43.44   0.000    -.0947761   -.0865935
  hourswater |   .0395158   .0064738     6.10   0.000     .0268274    .0522042
hourselect~y |    .138462   .0190026     7.29   0.000     .1012176    .1757064
------------------------------------------------------------------------------
```

Estimates of WTP for Water and Electricity

```
nlcom (val_water: -_b[hourswater]/_b[rent]) (val_elec: -_b[hourselectricity]/_b[rent])

   val_water:  -_b[hourswater]/_b[rent]
    val_elec:  -_b[hourselectricity]/_b[rent]

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   val_water |   .4357493   .0691955     6.30   0.000     .3001286      .57137
    val_elec |   1.526849   .2039104     7.49   0.000     1.127192    1.926506
------------------------------------------------------------------------------
```

Alternative Specific Conditional Logit Model

```
asclogit chosenarea rent hourswater hourselectricity, case(id) alternatives(area) noconstant
Iteration 0:   log likelihood = -1926.4932
Iteration 1:   log likelihood = -1895.1784
Iteration 2:   log likelihood = -1894.9069
Iteration 3:   log likelihood = -1894.9068

Alternative-specific conditional logit         Number of obs      =      14448
Case variable: id                              Number of cases    =       2408

Alternative variable: area                     Alts per case: min =          6
                                                              avg =        6.0
                                                              max =          6

                                                  Wald chi2(3)    =    1949.55
Log likelihood = -1894.9068                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0906848   .0020874   -43.44   0.000    -.0947761   -.0865935
  hourswater |   .0395158   .0064738     6.10   0.000     .0268274    .0522042
hourselect~y |    .138462   .0190026     7.29   0.000     .1012176    .1757064
------------------------------------------------------------------------------
```

```
. nlcom (val_water: -_b[hourswater]/_b[rent]) (val_elec: -_b[hourselectricity]/_b[rent])

   val_water:  -_b[hourswater]/_b[rent]
    val_elec:  -_b[hourselectricity]/_b[rent]

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   val_water |   .4357493   .0691955     6.30   0.000     .3001286      .57137
    val_elec |   1.526849   .2039104     7.49   0.000     1.127192    1.926506
------------------------------------------------------------------------------
```

(Alternative Specific Conditional) Mixed Logit Model

```
asclogit chosenarea rent hourswater hourselectricity, case(id) alternatives(area)
casevars(income age yearsofedu)
Iteration 0:   log likelihood = -1215.8448
Iteration 1:   log likelihood = -1174.0426
Iteration 2:   log likelihood = -1168.9807
Iteration 3:   log likelihood = -1168.9595
Iteration 4:   log likelihood = -1168.9595

Alternative-specific conditional logit         Number of obs      =     14100
Case variable: id                              Number of cases    =      2350

Alternative variable: area                     Alts per case: min =         6
                                                              avg =       6.0
                                                              max =         6

                                                  Wald chi2(18)   =   1539.73
Log likelihood = -1168.9595                       Prob > chi2     =    0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |   -.108208    .002815   -38.44   0.000    -.1137253   -.1026907
  hourswater |   .0654009   .0131131     4.99   0.000     .0396997    .0911022
hourselect~y |   .0976479   .0240414     4.06   0.000     .0505277    .1447681
-------------+----------------------------------------------------------------
1            |  (base alternative)
-------------+----------------------------------------------------------------
2            |
      income |   .0004068   .0002365     1.72   0.085    -.0000568    .0008704
         age |    .040176   .0156339     2.57   0.010     .0095342    .0708178
  yearsofedu |   .2125739   .0420192     5.06   0.000     .1302177      .29493
       _cons |  -.6908079   .7800153    -0.89   0.376     -2.21961     .837994
-------------+----------------------------------------------------------------
3            |
      income |  -.0008372   .0003766    -2.22   0.026    -.0015752   -.0000991
         age |   .0275534   .0127349     2.16   0.030     .0025936    .0525133
  yearsofedu |  -.0344758    .024307    -1.42   0.156    -.0821167     .013165
       _cons |  -1.755936   .5548171    -3.16   0.002    -2.843358   -.6685149
-------------+----------------------------------------------------------------
4            |
      income |  -.0005186   .0005256    -0.99   0.324    -.0015489    .0005116
         age |   .0484516   .0240185     2.02   0.044     .0013761    .0955271
  yearsofedu |   .1059339    .056107     1.89   0.059    -.0040338    .2159016
       _cons |  -4.648956   1.185294    -3.92   0.000    -6.972089   -2.325823
-------------+----------------------------------------------------------------
5            |
      income |  -.0003698   .0002734    -1.35   0.176    -.0009056     .000166
         age |   .0091326   .0132938     0.69   0.492    -.0169228     .035188
  yearsofedu |    .018762   .0274492     0.68   0.494    -.0350374    .0725614
       _cons |  -.4741326   .5857225    -0.81   0.418    -1.622128    .6738625
-------------+----------------------------------------------------------------
6            |
      income |  -.0005274   .0002112    -2.50   0.013    -.0009413   -.0001135
         age |   .0312944   .0099879     3.13   0.002     .0117186    .0508703
  yearsofedu |   .0387583   .0204992     1.89   0.059    -.0014194     .078936
       _cons |  -.0428549   .4471528    -0.10   0.924    -.9192582    .8335485
------------------------------------------------------------------------------
```

```
. nlcom (val_water: -_b[hourswater]/_b[rent]) (val_elec: -_b[hourselectricity]/_b[rent])

   val_water:  -_b[hourswater]/_b[rent]
    val_elec:  -_b[hourselectricity]/_b[rent]

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   val_water |   .6044002   .1192688     5.07   0.000     .3706377    .8381627
    val_elec |   .9024091   .2189338     4.12   0.000     .4733067    1.331511
------------------------------------------------------------------------------
```

Nested Logit Model Results

```
nlogitgen nlo = area (city: 1, lowden: 2, pollution: 3|4|5, other: 6)
new variable nlo is generated with 4 groups
label list lb_nlo
lb_nlo:
           1 city
           2 lowden
           3 pollution
           4 other

nlogit chosen rent  hourswater hourselectricity || nlo:  yearsofedu income age, base(city) ||
area:, noconstant case(id)
note: branch 1 of level 1 is degenerate and the associated dissimilarity parameter
([city_tau]_cons) is not defined; see help nlogit for details
note: branch 2 of level 1 is degenerate and the associated dissimilarity parameter
([lowden_tau]_cons) is not defined; see help nlogit for details
note: branch 4 of level 1 is degenerate and the associated dissimilarity parameter
([other_tau]_cons) is not defined; see help nlogit for details

tree structure specified for the nested logit model

 nlo        N       area  N    k
-----------------------------------
 city      2350 --- 1    2350  691
 lowden    2350 --- 2    2350  262
 pollution 7050 --- 3    2350  230
                 |- 4    2350   69
                 +- 5    2350  288
 other     2350 --- 6    2350  810
-----------------------------------
                 total  14100 2350

k = number of times alternative is chosen
N = number of observations at each level

Iteration 0:   log likelihood = -1821.9234
Iteration 1:   log likelihood = -1674.1495  (backed up)
Iteration 2:   log likelihood = -1663.8825  (backed up)
Iteration 3:   log likelihood = -1659.4088  (backed up)
Iteration 4:   log likelihood = -1526.6336  (backed up)
Iteration 5:   log likelihood = -1470.7959  (backed up)
Iteration 6:   log likelihood = -1466.7509  (backed up)
Iteration 7:   log likelihood = -1465.3476  (backed up)
Iteration 8:   log likelihood =  -1452.756  (backed up)
Iteration 9:   log likelihood = -1450.5822  (backed up)
Iteration 10:  log likelihood = -1449.6617  (backed up)
Iteration 11:  log likelihood = -1449.0232  (backed up)
Iteration 12:  log likelihood =  -1447.179  (backed up)
Iteration 13:  log likelihood =  -1443.285  (backed up)
Iteration 14:  log likelihood = -1313.5257
Iteration 15:  log likelihood = -1231.4045
Iteration 16:  log likelihood = -1221.8158
Iteration 17:  log likelihood = -1220.4108
Iteration 18:  log likelihood = -1220.2568
Iteration 19:  log likelihood = -1220.2171
Iteration 20:  log likelihood = -1220.2137
Iteration 21:  log likelihood = -1220.2135
Iteration 22:  log likelihood = -1220.2135
```

```
RUM-consistent nested logit regression         Number of obs      =      14100
Case variable: id                              Number of cases    =       2350

Alternative variable: area                     Alts per case: min =          6
                                                              avg =        6.0
                                                              max =          6

                                                  Wald chi2(12)   =    1256.55
Log likelihood = -1220.2135                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.1030771   .0029906   -34.47   0.000    -.1089384   -.0972157
  hourswater |   .0515747   .0094725     5.44   0.000      .033009    .0701405
hourselect~y |   .0942946   .0227599     4.14   0.000     .0496859    .1389033
------------------------------------------------------------------------------
nlo equations
------------------------------------------------------------------------------
city         |
  yearsofedu |     (base)
      income |     (base)
         age |     (base)
-------------+----------------------------------------------------------------
lowden       |
  yearsofedu |   .1856877   .0322076     5.77   0.000     .1225619    .2488136
      income |    .000479    .000225     2.13   0.033     .0000379     .00092
         age |   .0246836   .0107002     2.31   0.021     .0037116    .0456556
-------------+----------------------------------------------------------------
pollution    |
  yearsofedu |  -.0277628    .016751    -1.66   0.097    -.0605942    .0050686
      income |  -.0003507    .000209    -1.68   0.093    -.0007603     .000059
         age |  -.0076987   .0050978    -1.51   0.131    -.0176902    .0022928
-------------+----------------------------------------------------------------
other        |
  yearsofedu |   .0369857   .0171371     2.16   0.031     .0033977    .0705738
      income |  -.0005008   .0002064    -2.43   0.015    -.0009052   -.0000963
         age |   .0272518   .0052158     5.22   0.000     .0170291    .0374745
------------------------------------------------------------------------------
dissimilarity parameters
------------------------------------------------------------------------------
nlo          |
   /city_tau |          1   40786.74                     -79939.55    79941.55
 /lowden_tau |          1   198263.5                     -388588.4    388590.4
/pollution~u |   .7321211   .0501911                      .6337483    .8304939
  /other_tau |          1          .                             .           .
------------------------------------------------------------------------------
LR test for IIA (tau = 1):          chi2(3) =    21.13   Prob > chi2 = 0.0001
------------------------------------------------------------------------------

nlcom (val_water: -_b[hourswater]/_b[rent]) (val_elec: -_b[hourselectricity]/_b[rent])
   val_water:  -_b[hourswater]/_b[rent]
    val_elec:  -_b[hourselectricity]/_b[rent]

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   val_water |   .5003512   .0901415     5.55   0.000     .3236772    .6770252
    val_elec |   .9147971   .2165418     4.22   0.000      .490383    1.339211
------------------------------------------------------------------------------
```

Alternative Specific Probit Model

```
asmprobit chosenarea rent hourswater hourselectricity, case(id) alternatives(area) noconstant
intmethod(halton)
Iteration 0:   log simulated-likelihood = -3264.1275
Iteration 1:   log simulated-likelihood = -2677.4747  (backed up)
Iteration 2:   log simulated-likelihood = -2652.1637  (backed up)
Iteration 3:   log simulated-likelihood = -2495.8425  (backed up)
.
.
.
Iteration 35:  log simulated-likelihood = -2002.7582
Iteration 36:  log simulated-likelihood = -2002.7581
```

```
Iteration 37:  log simulated-likelihood = -2002.7581

Alternative-specific multinomial probit        Number of obs      =     14448
Case variable: id                              Number of cases    =      2408

Alternative variable: area                     Alts per case: min =         6
                                                              avg =       6.0
                                                              max =         6
Integration sequence:          Halton
Integration points:               300             Wald chi2(3)    =    637.81
Log simulated-likelihood = -2002.7581             Prob > chi2     =    0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |   -.027371    .001116   -24.53   0.000    -.0295583   -.0251837
  hourswater |   .0123694   .0025694     4.81   0.000     .0073336    .0174053
hourselect~y |   .0125349   .0067002     1.87   0.061    -.0005972    .0256669
-------------+----------------------------------------------------------------
     /lnl2_2 |  -.4731298   .0599158    -7.90   0.000    -.5905625   -.3556971
     /lnl3_3 |  -.6885425   .0574539   -11.98   0.000    -.8011501   -.5759348
     /lnl4_4 |  -.7393745   .0589692   -12.54   0.000    -.8549521    -.623797
     /lnl5_5 |  -.5626637   .1356907    -4.15   0.000    -.8286126   -.2967147
-------------+----------------------------------------------------------------
       /l2_1 |   .1819703   .0374869     4.85   0.000     .1084974    .2554432
       /l3_1 |   .2098528   .0411766     5.10   0.000     .1291482    .2905574
       /l4_1 |   .2067968   .0482621     4.28   0.000     .1122047    .3013888
       /l5_1 |  -.4859788   .0865288    -5.62   0.000    -.6555721   -.3163855
       /l3_2 |   .3960232   .0405786     9.76   0.000     .3164907    .4755557
       /l4_2 |   .4022693   .0410449     9.80   0.000     .3218227    .4827158
       /l5_2 |   .6735951   .0640955    10.51   0.000     .5479702      .79922
       /l4_3 |   .1876444   .0332772     5.64   0.000     .1224222    .2528665
       /l5_3 |   .3542718    .049398     7.17   0.000     .2574534    .4510902
       /l5_4 |   .2090518   .0508964     4.11   0.000     .1092967    .3088068
------------------------------------------------------------------------------
(area=1 is the alternative normalizing location)
(area=2 is the alternative normalizing scale)

. estat correlation

  +-------------------------------------------------------------------+
  |              |         2          3          4          5          6 |
  |--------------+----------------------------------------------------|
  |            2 |    1.0000                                           |
  |            3 |    0.2804     1.0000                                |
  |            4 |    0.3117     0.6521     1.0000                     |
  |            5 |    0.3024     0.6494     0.6450     1.0000          |
  |            6 |   -0.4467     0.4691     0.4679     0.4526     1.0000 |
  +-------------------------------------------------------------------+
Note: correlations are for alternatives differenced with 1

. estat covariance

  +-------------------------------------------------------------------+
  |              |         2          3          4          5          6 |
  |--------------+----------------------------------------------------|
  |            2 |         2                                           |
  |            3 |  .2573449   .4213035                                |
  |            4 |  .2967767   .2849289   .4531856                     |
  |            5 |  .2924548   .2882644     .29696   .4677185          |
  |            6 | -.6872778   .3312492   .3427287   .3367488   1.183663 |
  +-------------------------------------------------------------------+
Note: covariances are for alternatives differenced with 1

. nlcom (val_water: -_b[hourswater]/_b[rent]) (val_elec: -_b[hourselectricity]/_b[rent])

   val_water:  -_b[hourswater]/_b[rent]
    val_elec:  -_b[hourselectricity]/_b[rent]

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   val_water |    .451917   .0893336     5.06   0.000     .2768263    .6270077
    val_elec |   .4579607   .2410298     1.90   0.057    -.0144491    .9303704
------------------------------------------------------------------------------
```

(Alternative Specific) Mixed Probit Model

```
asmprobit chosenarea rent hourswater hourselectricity, case(id) alternatives(area)
casevars(income age yearsofedu)
Iteration 0:   log simulated-likelihood = -2283.9992
Iteration 1:   log simulated-likelihood = -2268.8366  (backed up)
Iteration 2:   log simulated-likelihood = -2261.8333  (backed up)
.
.
.
Iteration 55:  log simulated-likelihood = -1542.1925
Iteration 56:  log simulated-likelihood = -1542.1925
Iteration 57:  log simulated-likelihood = -1542.1925

Alternative-specific multinomial probit        Number of obs      =      14100
Case variable: id                              Number of cases    =       2350

Alternative variable: area                     Alts per case: min =          6
                                                              avg =        6.0
                                                              max =          6
Integration sequence:      Hammersley
Integration points:               300             Wald chi2(18)   =     483.62
Log simulated-likelihood = -1542.1925             Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0360907   .0017032   -21.19   0.000    -.0394289   -.0327526
  hourswater |   .0237085   .0048866     4.85   0.000     .0141309    .0332861
hourselect~y |   .0123481   .0084444     1.46   0.144    -.0042026    .0288988
-------------+----------------------------------------------------------------
1            |  (base alternative)
-------------+----------------------------------------------------------------
2            |
      income |   .0004519   .0001205     3.75   0.000     .0002157    .0006881
         age |   .0184212   .0070568     2.61   0.009     .0045901    .0322522
  yearsofedu |   .1019031   .0169249     6.02   0.000      .068731    .1350753
       _cons |   -1.49182   .3480694    -4.29   0.000    -2.174024   -.8096165
-------------+----------------------------------------------------------------
3            |
      income |  -.0003194   .0001203    -2.66   0.008    -.0005552   -.0000836
         age |   .0057001   .0043217     1.32   0.187    -.0027703    .0141706
  yearsofedu |  -.0114946   .0082151    -1.40   0.162    -.0275958    .0046066
       _cons |  -.3854566   .1912922    -2.02   0.044    -.7603823   -.0105308
-------------+----------------------------------------------------------------
4            |
      income |  -.0002701   .0001841    -1.47   0.142    -.0006309    .0000908
         age |   .0197841    .008218     2.41   0.016     .0036771    .0358911
  yearsofedu |    .044218   .0186644     2.37   0.018     .0076364    .0807995
       _cons |  -2.063203   .4412162    -4.68   0.000     -2.92797   -1.198435
-------------+----------------------------------------------------------------
5            |
      income |  -.0002207   .0001022    -2.16   0.031     -.000421   -.0000204
         age |   .0029009   .0044813     0.65   0.517    -.0058822    .0116841
  yearsofedu |  -.0052605   .0091441    -0.58   0.565    -.0231826    .0126617
       _cons |   .0387539   .2018365     0.19   0.848    -.3568384    .4343462
-------------+----------------------------------------------------------------
6            |
      income |  -.0002332   .0000964    -2.42   0.016    -.0004221   -.0000443
         age |   .0122184   .0041299     2.96   0.003      .004124    .0203128
  yearsofedu |   .0141426   .0083087     1.70   0.089    -.0021421    .0304273
       _cons |  -.1074489   .1875487    -0.57   0.567    -.4750376    .2601398
-------------+----------------------------------------------------------------
     /lnl2_2 |  -.4653058   .0781443    -5.95   0.000    -.6184657   -.3121458
     /lnl3_3 |  -.2140175   .0896538    -2.39   0.017    -.3897358   -.0382993
     /lnl4_4 |  -.5800966   .0708116    -8.19   0.000    -.7188848   -.4413085
     /lnl5_5 |  -.1507401   .0590327    -2.55   0.011     -.266442   -.0350381
-------------+----------------------------------------------------------------
       /l2_1 |   .3784239   .0489713     7.73   0.000      .282442    .4744058
       /l3_1 |    .324575   .1012572     3.21   0.001     .1261146    .5230355
       /l4_1 |   .3867887   .0508446     7.61   0.000     .2871351    .4864423
       /l5_1 |   .2798536   .0709304     3.95   0.000     .1408326    .4188745
       /l3_2 |   .5662735   .0957229     5.92   0.000       .37866    .7538869
       /l4_2 |   .4262696   .0544665     7.83   0.000     .3195173    .5330219
       /l5_2 |    .495055   .0675849     7.32   0.000      .362591    .6275191
```

```
      /l4_3 |   .1062284    .054969     1.93   0.053    -.0015087    .2139656
      /l5_3 |   .0495856   .0836597     0.59   0.553    -.1143843    .2135555
      /l5_4 |   .1706289   .0561052     3.04   0.002     .0606647     .280593
------------------------------------------------------------------------------
(area=1 is the alternative normalizing location)
(area=2 is the alternative normalizing scale)

. estat correlation
  +-------------------------------------------------------------------------+
  |              |         2          3          4          5          6 |
  |--------------+----------------------------------------------------------|
  |            2 |    1.0000                                                |
  |            3 |    0.5162     1.0000                                     |
  |            4 |    0.3126     0.6285     1.0000                          |
  |            5 |    0.4775     0.6973     0.5384     1.0000               |
  |            6 |    0.2675     0.5433     0.3786     0.4957     1.0000 |
  +-------------------------------------------------------------------------+
Note: correlations are for alternatives differenced with 1

. estat covariance
  +-------------------------------------------------------------------------+
  |              |         2          3          4          5          6 |
  |--------------+----------------------------------------------------------|
  |            2 |         2                                                |
  |            3 |  .5351723   .5375172                                     |
  |            4 |  .4590184   .4784145   1.077803                          |
  |            5 |  .5470018   .4140432    .452689   .6560214               |
  |            6 |  .3957727   .4167697   .4112022   .4200641   1.094693 |
  +-------------------------------------------------------------------------+
Note: covariances are for alternatives differenced with 1

. nlcom (val_water: -_b[hourswater]/_b[rent]) (val_elec: -_b[hourselectricity]/_b[rent])

   val_water:  -_b[hourswater]/_b[rent]
    val_elec:  -_b[hourselectricity]/_b[rent]

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   val_water |   .6569148   .1327536     4.95   0.000     .3967226    .9171071
    val_elec |     .34214   .2319204     1.48   0.140    -.1124156    .7966956
------------------------------------------------------------------------------
```

Sample of Rejected Models

These models were rejected because results are not consistent with theoretical expectations. The inclusion of quadratic variables distorts the whole model result.

Conditional Logit Model

```
clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, group(id)
note: 22 groups (90 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -741.81713
Iteration 1:   log likelihood =  -712.5834
Iteration 2:   log likelihood = -712.14954
Iteration 3:   log likelihood = -712.14782
Iteration 4:   log likelihood = -712.14782

Conditional (fixed-effects) logistic regression   Number of obs   =      14448
                                                  LR chi2(5)      =    7204.82
                                                  Prob > chi2     =     0.0000
Log likelihood = -712.14782                       Pseudo R2       =     0.8349

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |   -.086103   .0034773   -24.76   0.000    -.0929184   -.0792877
  hourswater |  -.6062201   .0427777   -14.17   0.000    -.6900629   -.5223774
 hourswater2 |   .0456944   .0031896    14.33   0.000     .0394429     .051946
hourselect~y |  -8.837766   .8291156   -10.66   0.000     -10.4628   -7.212729
hourselect~2 |   .8743273   .0778877    11.23   0.000     .7216702    1.026984
------------------------------------------------------------------------------
```

Alternative-specific Conditional Logit Model

```
asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, case(id)
alternatives(area) noconstant

Iteration 0:   log likelihood = -741.81713
Iteration 1:   log likelihood =  -712.5834
Iteration 2:   log likelihood = -712.14954
Iteration 3:   log likelihood = -712.14782
Iteration 4:   log likelihood = -712.14782

Alternative-specific conditional logit         Number of obs      =      14448
Case variable: id                              Number of cases    =       2408

Alternative variable: area                     Alts per case: min =          6
                                                              avg =        6.0
                                                              max =          6

                                                  Wald chi2(5)    =     780.66
Log likelihood = -712.14782                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |   -.086103   .0034773   -24.76   0.000    -.0929184   -.0792877
  hourswater |  -.6062201   .0427777   -14.17   0.000    -.6900629   -.5223774
 hourswater2 |   .0456944   .0031896    14.33   0.000     .0394429     .051946
hourselect~y |  -8.837766   .8291156   -10.66   0.000     -10.4628   -7.212729
hourselect~2 |   .8743273   .0778877    11.23   0.000     .7216702    1.026984
```

Alternative-specific Probit Model

```
asmprobit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, case(id)
alternatives(area) noconstant intmethod(halton)

Iteration 0:   log simulated-likelihood = -1159.5425
Iteration 1:   log simulated-likelihood = -931.26209  (backed up)
Iteration 2:   log simulated-likelihood = -910.53461  (backed up)
Iteration 3:   log simulated-likelihood = -895.94691  (backed up)
.
.
.
BFGS stepping has contracted, resetting BFGS Hessian
Iteration 205: log simulated-likelihood = -686.51607  (backed up)
Iteration 206: log simulated-likelihood = -686.51607  (backed up)
BFGS stepping has contracted, resetting BFGS Hessian
Iteration 207: log simulated-likelihood = -686.51607  (backed up)
Iteration 208: log simulated-likelihood = -686.51607  (backed up)
cannot compute an improvement -- flat region encountered

Convergence not achieved; you have estimated a maximum-likelihood model and Stata's
maximization procedure failed to converge to a solution; Check if the model is identified.
```

IIA Assumption Tests

Hausman Test of IIA - clogit

```
clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, group(id)
note: 22 groups (90 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951
Iteration 4:   log likelihood =  -939.5895

Conditional (fixed-effects) logistic regression   Number of obs   =      14448
                                                  LR chi2(5)      =    6749.93
                                                  Prob > chi2     =     0.0000
Log likelihood =  -939.5895                       Pseudo R2       =     0.7822

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------

est store all

clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 1, group(id)
note: 727 groups (3610 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -606.94929
Iteration 1:   log likelihood = -599.19493
Iteration 2:   log likelihood = -595.44262
Iteration 3:   log likelihood = -595.22134
Iteration 4:   log likelihood =  -595.2211
Iteration 5:   log likelihood =  -595.2211

Conditional (fixed-effects) logistic regression   Number of obs   =       8505
                                                  LR chi2(5)      =    4284.87
                                                  Prob > chi2     =     0.0000
Log likelihood =  -595.2211                       Pseudo R2       =     0.7826

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0714719   .0027159   -26.32   0.000     -.076795   -.0661489
  hourswater |   .8915462   .0522259    17.07   0.000     .7891852    .9939071
 hourswater2 |  -.0500489   .0029728   -16.84   0.000    -.0558753   -.0442224
hourselect~y |   1.323319   .0776111    17.05   0.000     1.171204    1.475434
hourselect~2 |  -.0711543    .005067   -14.04   0.000    -.0810855   -.0612231
------------------------------------------------------------------------------

est store partial

hausman partial all, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
        rent |   -.0714719    -.0816847        .0102128         .000966
  hourswater |    .8915462     .5863762          .30517        .0374989
 hourswater2 |   -.0500489    -.0309267       -.0191221        .0021423
hourselect~y |    1.323319     1.572483       -.2491639        .0357314
hourselect~2 |   -.0711543    -.0833422        .0121879        .0023315
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from clogit
           B = inconsistent under Ha, efficient under Ho; obtained from clogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          = -3504.09    chi2<0 ==> model fitted on these
                                        data fails to meet the asymptotic
                                        assumptions of the Hausman test;
                                        see suest for a generalized test

clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, group(id)
note: 22 groups (90 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951
Iteration 4:   log likelihood =  -939.5895

Conditional (fixed-effects) logistic regression   Number of obs   =      14448
                                                  LR chi2(5)      =    6749.93
                                                  Prob > chi2     =     0.0000
Log likelihood =  -939.5895                       Pseudo R2       =     0.7822

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------

est store all

clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 2, group(id)
note: 296 groups (1447 obs) dropped because of all positive or all negative outcomes.
```

```
Iteration 0:   log likelihood = -635.78348
Iteration 1:   log likelihood = -607.94059
Iteration 2:   log likelihood = -606.85315
Iteration 3:   log likelihood = -606.85223
Iteration 4:   log likelihood = -606.85223

Conditional (fixed-effects) logistic regression   Number of obs   =      10670
                                                  LR chi2(5)      =    5655.38
                                                  Prob > chi2     =     0.0000
Log likelihood = -606.85223                       Pseudo R2       =     0.8233

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0934666   .0035513   -26.32   0.000     -.100427   -.0865062
  hourswater |   .6696935   .0461291    14.52   0.000      .579282    .7601049
 hourswater2 |  -.0364653    .002633   -13.85   0.000    -.0416259   -.0313047
hourselect~y |   1.565323   .0881166    17.76   0.000     1.392618    1.738028
hourselect~2 |  -.0956024   .0068333   -13.99   0.000    -.1089954   -.0822093
------------------------------------------------------------------------------

est store partial

hausman partial all, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
        rent |   -.0934666    -.0816847       -.0117819        .0024837
  hourswater |    .6696935     .5863762        .0833173        .0283993
 hourswater2 |   -.0364653    -.0309267       -.0055385        .0016386
hourselect~y |    1.565323     1.572483       -.0071599        .0549344
hourselect~2 |   -.0956024    -.0833422       -.0122602        .0051435
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from clogit
           B = inconsistent under Ha, efficient under Ho; obtained from clogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      215.92
                Prob>chi2 =      0.0000

clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, group(id)
note: 22 groups (90 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951
Iteration 4:   log likelihood =  -939.5895

Conditional (fixed-effects) logistic regression   Number of obs   =      14448
                                                  LR chi2(5)      =    6749.93
                                                  Prob > chi2     =     0.0000
Log likelihood =  -939.5895                       Pseudo R2       =     0.7822

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------


est store all


clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 3, group(id)
note: 253 groups (1229 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -818.67763
Iteration 1:   log likelihood = -795.89201
Iteration 2:   log likelihood = -785.81341
Iteration 3:   log likelihood = -785.62236
Iteration 4:   log likelihood =  -785.6222
Iteration 5:   log likelihood =  -785.6222

Conditional (fixed-effects) logistic regression   Number of obs   =      10885
                                                  LR chi2(5)      =    5436.25
                                                  Prob > chi2     =     0.0000
Log likelihood =  -785.6222                       Pseudo R2       =     0.7758
```

```
------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0738061   .0025401   -29.06   0.000    -.0787846   -.0688276
  hourswater |   .3656079   .0396649     9.22   0.000     .2878661    .4433496
 hourswater2 |  -.0197418   .0021909    -9.01   0.000     -.024036   -.0154476
hourselect~y |   1.587966   .0731358    21.71   0.000     1.444623     1.73131
hourselect~2 |  -.0831061   .0046897   -17.72   0.000    -.0922977   -.0739145
------------------------------------------------------------------------------

est store partial

hausman partial all, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
        rent |   -.0738061    -.0816847        .0078786        .0000962
  hourswater |    .3656079     .5863762       -.2207683        .0158722
 hourswater2 |   -.0197418    -.0309267        .0111849        .0007433
hourselect~y |    1.587966     1.572483        .0154834        .0245378
hourselect~2 |   -.0831061    -.0833422        .0002361        .0013244
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from clogit
           B = inconsistent under Ha, efficient under Ho; obtained from clogit

    Test:  Ho:  difference in coefficients not systematic

                 chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                         =  -444.84    chi2<0 ==> model fitted on these
                                       data fails to meet the asymptotic
                                       assumptions of the Hausman test;
                                       see suest for a generalized test

clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, group(id)
note: 22 groups (90 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951
Iteration 4:   log likelihood =  -939.5895

Conditional (fixed-effects) logistic regression   Number of obs   =      14448
                                                  LR chi2(5)      =    6749.93
                                                  Prob > chi2     =     0.0000
Log likelihood =  -939.5895                       Pseudo R2       =     0.7822

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------

est store all

clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 4, group(id)
note: 90 groups (418 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood =  -733.3802
Iteration 1:   log likelihood = -720.61732
Iteration 2:   log likelihood = -720.01945
Iteration 3:   log likelihood = -720.01821
Iteration 4:   log likelihood = -720.01821

Conditional (fixed-effects) logistic regression   Number of obs   =      11695
                                                  LR chi2(5)      =    6088.91
                                                  Prob > chi2     =     0.0000
Log likelihood = -720.01821                       Pseudo R2       =     0.8087

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |    -.07537   .0025829   -29.18   0.000    -.0804324   -.0703076
  hourswater |   .6096924   .0387379    15.74   0.000     .5337674    .6856174
 hourswater2 |  -.0257066   .0023863   -10.77   0.000    -.0303836   -.0210297
hourselect~y |   1.518456   .0728777    20.84   0.000     1.375618    1.661293
hourselect~2 |  -.0804084   .0046666   -17.23   0.000    -.0895548   -.0712621
------------------------------------------------------------------------------

est store partial

hausman partial all, alleqs constant
```

```
                ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
        rent |      -.07537    -.0816847        .0063147        .0004781
  hourswater |     .6096924     .5863762        .0233162        .0133885
 hourswater2 |    -.0257066    -.0309267        .0052201        .0012027
hourselect~y |     1.518456     1.572483       -.0540272        .0237573
hourselect~2 |    -.0804084    -.0833422        .0029338        .0012402
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from clogit
           B = inconsistent under Ha, efficient under Ho; obtained from clogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =  -971.26    chi2<0 ==> model fitted on these
                                        data fails to meet the asymptotic
                                        assumptions of the Hausman test;
                                        see suest for a generalized test

clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, group(id)
note: 22 groups (90 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951
Iteration 4:   log likelihood =  -939.5895

Conditional (fixed-effects) logistic regression   Number of obs   =      14448
                                                  LR chi2(5)      =    6749.93
                                                  Prob > chi2     =     0.0000
Log likelihood =  -939.5895                       Pseudo R2       =     0.7822

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------

est store all

clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 5, group(id)
note: 310 groups (1514 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -890.48219
Iteration 1:   log likelihood = -864.97827
Iteration 2:   log likelihood = -860.98708
Iteration 3:   log likelihood =  -860.9604
Iteration 4:   log likelihood =  -860.9604

Conditional (fixed-effects) logistic regression   Number of obs   =      10600
                                                  LR chi2(5)      =    5102.10
                                                  Prob > chi2     =     0.0000
Log likelihood =  -860.9604                       Pseudo R2       =     0.7477

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0765629   .0025511   -30.01   0.000     -.081563   -.0715628
  hourswater |   .5061264   .0369192    13.71   0.000     .4337662    .5784867
 hourswater2 |  -.0268429   .0020651   -13.00   0.000    -.0308905   -.0227954
hourselect~y |   1.506549   .0698072    21.58   0.000     1.369729    1.643369
hourselect~2 |  -.0792909   .0045221   -17.53   0.000     -.088154   -.0704278
------------------------------------------------------------------------------

est store partial

hausman partial all, alleqs constant

                ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
        rent |    -.0765629    -.0816847        .0051218        .0002556
  hourswater |     .5061264     .5863762       -.0802497        .0064536
 hourswater2 |    -.0268429    -.0309267        .0040838        .0001297
hourselect~y |     1.506549     1.572483       -.0659337        .0112385
hourselect~2 |    -.0792909    -.0833422        .0040513        .0004585
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from clogit
           B = inconsistent under Ha, efficient under Ho; obtained from clogit
```

```
    Test:  Ho:  difference in coefficients not systematic

                  chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       75.36
                Prob>chi2 =      0.0000
                (V_b-V_B is not positive definite)

clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, group(id)
note: 22 groups (90 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951
Iteration 4:   log likelihood =  -939.5895

Conditional (fixed-effects) logistic regression   Number of obs   =      14448
                                                  LR chi2(5)      =    6749.93
                                                  Prob > chi2     =     0.0000
Log likelihood =  -939.5895                       Pseudo R2       =     0.7822

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------

est store all

clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 6, group(id)
note: 859 groups (4272 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -577.42021
Iteration 1:   log likelihood = -570.68496
Iteration 2:   log likelihood = -570.57189
Iteration 3:   log likelihood = -570.57182
Iteration 4:   log likelihood = -570.57182

Conditional (fixed-effects) logistic regression   Number of obs   =       7845
                                                  LR chi2(5)      =    3909.27
                                                  Prob > chi2     =     0.0000
Log likelihood = -570.57182                       Pseudo R2       =     0.7740

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0760328   .0030758   -24.72   0.000    -.0820613   -.0700043
  hourswater |   .4830981   .0437156    11.05   0.000     .3974172     .568779
 hourswater2 |  -.0240531    .002437    -9.87   0.000    -.0288295   -.0192768
hourselect~y |   1.525461   .0891237    17.12   0.000     1.350782    1.700141
hourselect~2 |  -.0762712   .0053816   -14.17   0.000    -.0868189   -.0657234
------------------------------------------------------------------------------

est store partial

hausman partial all, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
        rent |   -.0760328    -.0816847        .0056519        .0017372
  hourswater |    .4830981     .5863762       -.1032781        .0242832
 hourswater2 |   -.0240531    -.0309267        .0068736        .0013004
hourselect~y |    1.525461     1.572483       -.0470215        .0565358
hourselect~2 |   -.0762712    -.0833422         .007071        .0029534
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from clogit
           B = inconsistent under Ha, efficient under Ho; obtained from clogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      150.90
                Prob>chi2 =      0.0000
```

Small-Hsio Test of IIA - clogit

```
clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, group(id)
note: 22 groups (90 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
```

```
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951
Iteration 4:   log likelihood =  -939.5895

Conditional (fixed-effects) logistic regression   Number of obs   =      14448
                                                  LR chi2(5)      =    6749.93
                                                  Prob > chi2     =     0.0000
Log likelihood =  -939.5895                       Pseudo R2       =     0.7822

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------

estimates store all

clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 1, group(id)
note: 727 groups (3610 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -606.94929
Iteration 1:   log likelihood = -599.19493
Iteration 2:   log likelihood = -595.44262
Iteration 3:   log likelihood = -595.22134
Iteration 4:   log likelihood =  -595.2211
Iteration 5:   log likelihood =  -595.2211

Conditional (fixed-effects) logistic regression   Number of obs   =       8505
                                                  LR chi2(5)      =    4284.87
                                                  Prob > chi2     =     0.0000
Log likelihood =  -595.2211                       Pseudo R2       =     0.7826

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0714719   .0027159   -26.32   0.000     -.076795   -.0661489
  hourswater |   .8915462   .0522259    17.07   0.000     .7891852    .9939071
 hourswater2 |  -.0500489   .0029728   -16.84   0.000    -.0558753   -.0442224
hourselect~y |   1.323319   .0776111    17.05   0.000     1.171204    1.475434
hourselect~2 |  -.0711543    .005067   -14.04   0.000    -.0810855   -.0612231
------------------------------------------------------------------------------

estimates store one


suest all one

Simultaneous results for all, one
                                                  Number of obs   =      14448

                                (Std. Err. adjusted for 2408 clusters in id)
------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
all_chosen~a |
        rent |  -.0816847   .0048873   -16.71   0.000    -.0912637   -.0721057
  hourswater |   .5863762   .0337523    17.37   0.000      .520223    .6525294
 hourswater2 |  -.0309267   .0021546   -14.35   0.000    -.0351497   -.0267037
hourselect~y |   1.572483   .0912337    17.24   0.000     1.393668    1.751298
hourselect~2 |  -.0833422   .0071379   -11.68   0.000    -.0973322   -.0693521
-------------+----------------------------------------------------------------
one_chosen~a |
        rent |  -.0714719   .0049314   -14.49   0.000    -.0811374   -.0618065
  hourswater |   .8915462   .0574021    15.53   0.000     .7790401    1.004052
 hourswater2 |  -.0500489   .0035731   -14.01   0.000     -.057052   -.0430457
hourselect~y |   1.323319    .099905    13.25   0.000     1.127509    1.519129
hourselect~2 |  -.0711543   .0078016    -9.12   0.000    -.0864452   -.0558633
------------------------------------------------------------------------------

clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, group(id)
note: 22 groups (90 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951
Iteration 4:   log likelihood =  -939.5895

Conditional (fixed-effects) logistic regression   Number of obs   =      14448
                                                  LR chi2(5)      =    6749.93
                                                  Prob > chi2     =     0.0000
Log likelihood =  -939.5895                       Pseudo R2       =     0.7822
```

```
------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------

estimates store all

. clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 2, group(id)
note: 296 groups (1447 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -635.78348
Iteration 1:   log likelihood = -607.94059
Iteration 2:   log likelihood = -606.85315
Iteration 3:   log likelihood = -606.85223
Iteration 4:   log likelihood = -606.85223

Conditional (fixed-effects) logistic regression   Number of obs   =      10670
                                                  LR chi2(5)      =    5655.38
                                                  Prob > chi2     =     0.0000
Log likelihood = -606.85223                       Pseudo R2       =     0.8233

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0934666   .0035513   -26.32   0.000     -.100427   -.0865062
  hourswater |   .6696935   .0461291    14.52   0.000      .579282    .7601049
 hourswater2 |  -.0364653    .002633   -13.85   0.000    -.0416259   -.0313047
hourselect~y |   1.565323   .0881166    17.76   0.000     1.392618    1.738028
hourselect~2 |  -.0956024   .0068333   -13.99   0.000    -.1089954   -.0822093
------------------------------------------------------------------------------

estimates store two

suest all two

Simultaneous results for all, two
                                                  Number of obs   =      14448

                                (Std. Err. adjusted for 2408 clusters in id)
------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
all_chosen~a |
        rent |  -.0816847   .0048873   -16.71   0.000    -.0912637   -.0721057
  hourswater |   .5863762   .0337523    17.37   0.000      .520223    .6525294
 hourswater2 |  -.0309267   .0021546   -14.35   0.000    -.0351497   -.0267037
hourselect~y |   1.572483   .0912337    17.24   0.000     1.393668    1.751298
hourselect~2 |  -.0833422   .0071379   -11.68   0.000    -.0973322   -.0693521
-------------+----------------------------------------------------------------
two_chosen~a |
        rent |  -.0934666   .0076226   -12.26   0.000    -.1084066   -.0785266
  hourswater |   .6696935   .0423634    15.81   0.000     .5866628    .7527242
 hourswater2 |  -.0364653   .0026423   -13.80   0.000     -.041644   -.0312865
hourselect~y |   1.565323   .1374967    11.38   0.000     1.295834    1.834812
hourselect~2 |  -.0956024   .0135604    -7.05   0.000    -.1221802   -.0690245
------------------------------------------------------------------------------


clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, group(id)
note: 22 groups (90 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951
Iteration 4:   log likelihood =  -939.5895

Conditional (fixed-effects) logistic regression   Number of obs   =      14448
                                                  LR chi2(5)      =    6749.93
                                                  Prob > chi2     =     0.0000
Log likelihood =  -939.5895                       Pseudo R2       =     0.7822

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------
```

```
estimates store all

clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 3, group(id)
note: 253 groups (1229 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -818.67763
Iteration 1:   log likelihood = -795.89201
Iteration 2:   log likelihood = -785.81341
Iteration 3:   log likelihood = -785.62236
Iteration 4:   log likelihood =  -785.6222
Iteration 5:   log likelihood =  -785.6222

Conditional (fixed-effects) logistic regression   Number of obs   =      10885
                                                  LR chi2(5)      =    5436.25
                                                  Prob > chi2     =     0.0000
Log likelihood =  -785.6222                       Pseudo R2       =     0.7758

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0738061   .0025401   -29.06   0.000    -.0787846   -.0688276
  hourswater |   .3656079   .0396649     9.22   0.000     .2878661    .4433496
 hourswater2 |  -.0197418   .0021909    -9.01   0.000     -.024036   -.0154476
hourselect~y |   1.587966   .0731358    21.71   0.000     1.444623     1.73131
hourselect~2 |  -.0831061   .0046897   -17.72   0.000    -.0922977   -.0739145
------------------------------------------------------------------------------

estimates store three

suest all three

Simultaneous results for all, three
                                                  Number of obs   =      14448

                              (Std. Err. adjusted for 2408 clusters in id)
------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
all_chosen~a |
        rent |  -.0816847   .0048873   -16.71   0.000    -.0912637   -.0721057
  hourswater |   .5863762   .0337523    17.37   0.000      .520223    .6525294
 hourswater2 |  -.0309267   .0021546   -14.35   0.000    -.0351497   -.0267037
hourselect~y |   1.572483   .0912337    17.24   0.000     1.393668    1.751298
hourselect~2 |  -.0833422   .0071379   -11.68   0.000    -.0973322   -.0693521
-------------+----------------------------------------------------------------
three_chos~a |
        rent |  -.0738061   .0046572   -15.85   0.000    -.0829341   -.0646781
  hourswater |   .3656079   .0376539     9.71   0.000     .2918076    .4394081
 hourswater2 |  -.0197418    .002294    -8.61   0.000    -.0242379   -.0152457
hourselect~y |   1.587966   .0965696    16.44   0.000     1.398693    1.777239
hourselect~2 |  -.0831061   .0073834   -11.26   0.000    -.0975773    -.068635
------------------------------------------------------------------------------

clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, group(id)
note: 22 groups (90 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951
Iteration 4:   log likelihood =  -939.5895

Conditional (fixed-effects) logistic regression   Number of obs   =      14448
                                                  LR chi2(5)      =    6749.93
                                                  Prob > chi2     =     0.0000
Log likelihood =  -939.5895                       Pseudo R2       =     0.7822

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------

estimates store all

clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 4, group(id)
note: 90 groups (418 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood =  -733.3802
Iteration 1:   log likelihood = -720.61732
Iteration 2:   log likelihood = -720.01945
Iteration 3:   log likelihood = -720.01821
Iteration 4:   log likelihood = -720.01821
```

```
Conditional (fixed-effects) logistic regression   Number of obs   =      11695
                                                  LR chi2(5)      =    6088.91
                                                  Prob > chi2     =     0.0000
Log likelihood = -720.01821                       Pseudo R2       =     0.8087

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |    -.07537   .0025829   -29.18   0.000    -.0804324   -.0703076
  hourswater |   .6096924   .0387379    15.74   0.000     .5337674    .6856174
 hourswater2 |  -.0257066   .0023863   -10.77   0.000    -.0303836   -.0210297
hourselect~y |   1.518456   .0728777    20.84   0.000     1.375618    1.661293
hourselect~2 |  -.0804084   .0046666   -17.23   0.000    -.0895548   -.0712621
------------------------------------------------------------------------------

estimates store four

suest all four


Simultaneous results for all, four
                                                  Number of obs   =      14448

                                 (Std. Err. adjusted for 2408 clusters in id)
------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
all_chosen~a |
        rent |  -.0816847   .0048873   -16.71   0.000    -.0912637   -.0721057
  hourswater |   .5863762   .0337523    17.37   0.000      .520223    .6525294
 hourswater2 |  -.0309267   .0021546   -14.35   0.000    -.0351497   -.0267037
hourselect~y |   1.572483   .0912337    17.24   0.000     1.393668    1.751298
hourselect~2 |  -.0833422   .0071379   -11.68   0.000    -.0973322   -.0693521
-------------+----------------------------------------------------------------
four_chose~a |
        rent |    -.07537    .004627   -16.29   0.000    -.0844387   -.0663013
  hourswater |   .6096924   .0314024    19.42   0.000     .5481449      .67124
 hourswater2 |  -.0257066   .0022561   -11.39   0.000    -.0301285   -.0212848
hourselect~y |   1.518456   .0894016    16.98   0.000     1.343232    1.693679
hourselect~2 |  -.0804084   .0063572   -12.65   0.000    -.0928684   -.0679485
------------------------------------------------------------------------------


clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, group(id)
note: 22 groups (90 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951
Iteration 4:   log likelihood =  -939.5895

Conditional (fixed-effects) logistic regression   Number of obs   =      14448
                                                  LR chi2(5)      =    6749.93
                                                  Prob > chi2     =     0.0000
Log likelihood =  -939.5895                       Pseudo R2       =     0.7822

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------

estimates store all


clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 5, group(id)
note: 310 groups (1514 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -890.48219
Iteration 1:   log likelihood = -864.97827
Iteration 2:   log likelihood = -860.98708
Iteration 3:   log likelihood =  -860.9604
Iteration 4:   log likelihood =  -860.9604

Conditional (fixed-effects) logistic regression   Number of obs   =      10600
                                                  LR chi2(5)      =    5102.10
                                                  Prob > chi2     =     0.0000
Log likelihood =  -860.9604                       Pseudo R2       =     0.7477
```

```
------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0765629   .0025511   -30.01   0.000     -.081563   -.0715628
  hourswater |   .5061264   .0369192    13.71   0.000     .4337662    .5784867
 hourswater2 |  -.0268429   .0020651   -13.00   0.000    -.0308905   -.0227954
hourselect~y |   1.506549   .0698072    21.58   0.000     1.369729    1.643369
hourselect~2 |  -.0792909   .0045221   -17.53   0.000     -.088154   -.0704278
------------------------------------------------------------------------------

estimates store five

suest all five

Simultaneous results for all, five
                                                  Number of obs   =      14448

                                  (Std. Err. adjusted for 2408 clusters in id)
------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
all_chosen~a |
        rent |  -.0816847   .0048873   -16.71   0.000    -.0912637   -.0721057
  hourswater |   .5863762   .0337523    17.37   0.000      .520223    .6525294
 hourswater2 |  -.0309267   .0021546   -14.35   0.000    -.0351497   -.0267037
hourselect~y |   1.572483   .0912337    17.24   0.000     1.393668    1.751298
hourselect~2 |  -.0833422   .0071379   -11.68   0.000    -.0973322   -.0693521
-------------+----------------------------------------------------------------
five_chose~a |
        rent |  -.0765629    .004754   -16.10   0.000    -.0858806   -.0672451
  hourswater |   .5061264   .0346689    14.60   0.000     .4381766    .5740763
 hourswater2 |  -.0268429   .0021365   -12.56   0.000    -.0310303   -.0226555
hourselect~y |   1.506549    .090306    16.68   0.000     1.329553    1.683546
hourselect~2 |  -.0792909   .0069907   -11.34   0.000    -.0929923   -.0655894
------------------------------------------------------------------------------

clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, group(id)
note: 22 groups (90 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951
Iteration 4:   log likelihood =  -939.5895

Conditional (fixed-effects) logistic regression   Number of obs   =      14448
                                                  LR chi2(5)      =    6749.93
                                                  Prob > chi2     =     0.0000
Log likelihood =  -939.5895                       Pseudo R2       =     0.7822

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------

estimates store all

clogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 6, group(id)
note: 859 groups (4272 obs) dropped because of all positive or all negative outcomes.

Iteration 0:   log likelihood = -577.42021
Iteration 1:   log likelihood = -570.68496
Iteration 2:   log likelihood = -570.57189
Iteration 3:   log likelihood = -570.57182
Iteration 4:   log likelihood = -570.57182

Conditional (fixed-effects) logistic regression   Number of obs   =       7845
                                                  LR chi2(5)      =    3909.27
                                                  Prob > chi2     =     0.0000
Log likelihood = -570.57182                       Pseudo R2       =     0.7740

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rent |  -.0760328   .0030758   -24.72   0.000    -.0820613   -.0700043
  hourswater |   .4830981   .0437156    11.05   0.000     .3974172     .568779
 hourswater2 |  -.0240531    .002437    -9.87   0.000    -.0288295   -.0192768
hourselect~y |   1.525461   .0891237    17.12   0.000     1.350782    1.700141
hourselect~2 |  -.0762712   .0053816   -14.17   0.000    -.0868189   -.0657234
------------------------------------------------------------------------------

estimates store six
```

```
suest all six

Simultaneous results for all, six
                                                  Number of obs   =      14448

                                  (Std. Err. adjusted for 2408 clusters in id)
------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
all_chosen~a |
        rent |  -.0816847   .0048873   -16.71   0.000    -.0912637   -.0721057
  hourswater |   .5863762   .0337523    17.37   0.000      .520223    .6525294
 hourswater2 |  -.0309267   .0021546   -14.35   0.000    -.0351497   -.0267037
hourselect~y |   1.572483   .0912337    17.24   0.000     1.393668    1.751298
hourselect~2 |  -.0833422   .0071379   -11.68   0.000    -.0973322   -.0693521
-------------+----------------------------------------------------------------
six_chosen~a |
        rent |  -.0760328   .0058304   -13.04   0.000    -.0874601   -.0646055
  hourswater |   .4830981   .0383916    12.58   0.000     .4078519    .5583443
 hourswater2 |  -.0240531     .00241    -9.98   0.000    -.0287766   -.0193296
hourselect~y |   1.525461    .113299    13.46   0.000     1.303399    1.747523
hourselect~2 |  -.0762712   .0075516   -10.10   0.000    -.0910721   -.0614703
------------------------------------------------------------------------------
```

Hausman Test of IIA - asclogit

```
asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, case(id)
alternatives(area) noconstant

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951
Iteration 4:   log likelihood =  -939.5895

Alternative-specific conditional logit         Number of obs      =      14448
Case variable: id                              Number of cases    =       2408

Alternative variable: area                     Alts per case: min =          6
                                                              avg =        6.0
                                                              max =          6

                                                  Wald chi2(5)    =    1232.19
Log likelihood =  -939.5895                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------

est store all

asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 1, case(id)
alternatives(area) noconstant
note: 711 cases (3555 obs) dropped due to no positive outcome per case

Iteration 0:   log likelihood = -606.94929
Iteration 1:   log likelihood = -599.19493
Iteration 2:   log likelihood = -595.44262
Iteration 3:   log likelihood = -595.22134
Iteration 4:   log likelihood =  -595.2211
Iteration 5:   log likelihood =  -595.2211

Alternative-specific conditional logit         Number of obs      =       8505
Case variable: id                              Number of cases    =       1701

Alternative variable: area                     Alts per case: min =          5
                                                              avg =        5.0
                                                              max =          5

                                                  Wald chi2(5)    =     814.38
Log likelihood =  -595.2211                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0714719   .0027159   -26.32   0.000     -.076795   -.0661489
  hourswater |   .8915462   .0522259    17.07   0.000     .7891852    .9939071
 hourswater2 |  -.0500489   .0029728   -16.84   0.000    -.0558753   -.0442224
hourselect~y |   1.323319   .0776111    17.05   0.000     1.171204    1.475434
hourselect~2 |  -.0711543    .005067   -14.04   0.000    -.0810855   -.0612231
------------------------------------------------------------------------------

est store partial


hausman partial all, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
        rent |   -.0714719    -.0816847        .0102128         .000966
  hourswater |    .8915462     .5863762          .30517        .0374989
 hourswater2 |   -.0500489    -.0309267       -.0191221        .0021423
hourselect~y |    1.323319     1.572483       -.2491639        .0357314
hourselect~2 |   -.0711543    -.0833422        .0121879        .0023315
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from asclogit
            B = inconsistent under Ha, efficient under Ho; obtained from asclogit

    Test:  Ho:  difference in coefficients not systematic
```

```
                  chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          = -3504.09    chi2<0 ==> model fitted on these
                                        data fails to meet the asymptotic
                                        assumptions of the Hausman test;
                                        see suest for a generalized test

asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, case(id)
alternatives(area) noconstant

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951
Iteration 4:   log likelihood =  -939.5895

Alternative-specific conditional logit         Number of obs      =      14448
Case variable: id                              Number of cases    =       2408

Alternative variable: area                     Alts per case: min =          6
                                                              avg =        6.0
                                                              max =          6

                                                  Wald chi2(5)    =    1232.19
Log likelihood =  -939.5895                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------

est store all

asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 2, case(id)
alternatives(area) noconstant
note: 278 cases (1390 obs) dropped due to no positive outcome per case

Iteration 0:   log likelihood = -635.78348
Iteration 1:   log likelihood = -607.94059
Iteration 2:   log likelihood = -606.85315
Iteration 3:   log likelihood = -606.85223
Iteration 4:   log likelihood = -606.85223

Alternative-specific conditional logit         Number of obs      =      10670
Case variable: id                              Number of cases    =       2134

Alternative variable: area                     Alts per case: min =          5
                                                              avg =        5.0
                                                              max =          5

                                                  Wald chi2(5)    =     884.89
Log likelihood = -606.85223                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0934666   .0035513   -26.32   0.000     -.100427   -.0865062
  hourswater |   .6696935   .0461291    14.52   0.000      .579282    .7601049
 hourswater2 |  -.0364653    .002633   -13.85   0.000    -.0416259   -.0313047
hourselect~y |   1.565323   .0881166    17.76   0.000     1.392618    1.738028
hourselect~2 |  -.0956024   .0068333   -13.99   0.000    -.1089954   -.0822093
------------------------------------------------------------------------------

est store partial

hausman partial all, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
        rent |   -.0934666    -.0816847       -.0117819        .0024837
  hourswater |    .6696935     .5863762        .0833173        .0283993
 hourswater2 |   -.0364653    -.0309267       -.0055385        .0016386
hourselect~y |    1.565323     1.572483       -.0071599        .0549344
hourselect~2 |   -.0956024    -.0833422       -.0122602        .0051435
------------------------------------------------------------------------------
                        b = consistent under Ho and Ha; obtained from asclogit
         B = inconsistent under Ha, efficient under Ho; obtained from asclogit

    Test:  Ho:  difference in coefficients not systematic
```

```
                 chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                         =      215.92
               Prob>chi2 =      0.0000

asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, case(id)
alternatives(area) noconstant

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951
Iteration 4:   log likelihood =  -939.5895

Alternative-specific conditional logit         Number of obs      =      14448
Case variable: id                              Number of cases    =       2408

Alternative variable: area                     Alts per case: min =          6
                                                              avg =        6.0
                                                              max =          6

                                                  Wald chi2(5)    =    1232.19
Log likelihood =  -939.5895                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------

est store all

asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 3, case(id)
alternatives(area) noconstant
note: 232 cases (1160 obs) dropped due to no positive outcome per case

Iteration 0:   log likelihood = -818.67763
Iteration 1:   log likelihood = -795.89201
Iteration 2:   log likelihood = -785.81341
Iteration 3:   log likelihood = -785.62236
Iteration 4:   log likelihood =  -785.6222
Iteration 5:   log likelihood =  -785.6222

Alternative-specific conditional logit         Number of obs      =      10885
Case variable: id                              Number of cases    =       2177

Alternative variable: area                     Alts per case: min =          5
                                                              avg =        5.0
                                                              max =          5

                                                  Wald chi2(5)    =    1002.89
Log likelihood =  -785.6222                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0738061   .0025401   -29.06   0.000    -.0787846   -.0688276
  hourswater |   .3656079   .0396649     9.22   0.000     .2878661    .4433496
 hourswater2 |  -.0197418   .0021909    -9.01   0.000     -.024036   -.0154476
hourselect~y |   1.587966   .0731358    21.71   0.000     1.444623     1.73131
hourselect~2 |  -.0831061   .0046897   -17.72   0.000    -.0922977   -.0739145
------------------------------------------------------------------------------

est store partial

hausman partial all, alleqs constant

                ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
        rent |   -.0738061    -.0816847        .0078786        .0000962
  hourswater |    .3656079     .5863762       -.2207683        .0158722
 hourswater2 |   -.0197418    -.0309267        .0111849        .0007433
hourselect~y |    1.587966     1.572483        .0154834        .0245378
hourselect~2 |   -.0831061    -.0833422        .0002361        .0013244
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from asclogit
            B = inconsistent under Ha, efficient under Ho; obtained from asclogit

    Test:  Ho:  difference in coefficients not systematic

                 chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
```

```
                         =  -444.84    chi2<0 ==> model fitted on these
                                       data fails to meet the asymptotic
                                       assumptions of the Hausman test;
                                       see suest for a generalized test

asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, case(id)
alternatives(area) noconstant

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951
Iteration 4:   log likelihood =  -939.5895

Alternative-specific conditional logit         Number of obs      =      14448
Case variable: id                              Number of cases    =       2408

Alternative variable: area                     Alts per case: min =          6
                                                              avg =        6.0
                                                              max =          6

                                                  Wald chi2(5)    =    1232.19
Log likelihood =  -939.5895                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------

est store all

asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 4, case(id)
alternatives(area) noconstant
note: 70 cases (350 obs) dropped due to no positive outcome per case

Iteration 0:   log likelihood =  -733.3802
Iteration 1:   log likelihood = -720.61732
Iteration 2:   log likelihood = -720.01945
Iteration 3:   log likelihood = -720.01821
Iteration 4:   log likelihood = -720.01821

Alternative-specific conditional logit         Number of obs      =      11695
Case variable: id                              Number of cases    =       2339

Alternative variable: area                     Alts per case: min =          5
                                                              avg =        5.0
                                                              max =          5

                                                  Wald chi2(5)    =    1033.50
Log likelihood = -720.01821                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |    -.07537   .0025829   -29.18   0.000    -.0804324   -.0703076
  hourswater |   .6096924   .0387379    15.74   0.000     .5337674    .6856174
 hourswater2 |  -.0257066   .0023863   -10.77   0.000    -.0303836   -.0210297
hourselect~y |   1.518456   .0728777    20.84   0.000     1.375618    1.661293
hourselect~2 |  -.0804084   .0046666   -17.23   0.000    -.0895548   -.0712621
------------------------------------------------------------------------------

est store partial

hausman partial all, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
        rent |     -.07537    -.0816847        .0063147        .0004781
  hourswater |    .6096924     .5863762        .0233162        .0133885
 hourswater2 |   -.0257066    -.0309267        .0052201        .0012027
hourselect~y |    1.518456     1.572483       -.0540272        .0237573
hourselect~2 |   -.0804084    -.0833422        .0029338        .0012402
------------------------------------------------------------------------------
                        b = consistent under Ho and Ha; obtained from asclogit
         B = inconsistent under Ha, efficient under Ho; obtained from asclogit

    Test:  Ho:  difference in coefficients not systematic

                 chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
```

```
                          =  -971.26    chi2<0 ==> model fitted on these
                                        data fails to meet the asymptotic
                                        assumptions of the Hausman test;
                                        see suest for a generalized test

asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, case(id)
alternatives(area) noconstant

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951
Iteration 4:   log likelihood =  -939.5895

Alternative-specific conditional logit         Number of obs      =      14448
Case variable: id                              Number of cases    =       2408

Alternative variable: area                     Alts per case: min =          6
                                                              avg =        6.0
                                                              max =          6

                                                  Wald chi2(5)    =    1232.19
Log likelihood =  -939.5895                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------

est store all

asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 5, case(id)
alternatives(area) noconstant
note: 289 cases (1445 obs) dropped due to no positive outcome per case

Iteration 0:   log likelihood = -890.48219
Iteration 1:   log likelihood = -864.97827
Iteration 2:   log likelihood = -860.98708
Iteration 3:   log likelihood =  -860.9604

Alternative-specific conditional logit         Number of obs      =      10600
Case variable: id                              Number of cases    =       2120

Alternative variable: area                     Alts per case: min =          5
                                                              avg =        5.0
                                                              max =          5

                                                  Wald chi2(5)    =    1038.37
Log likelihood =  -860.9604                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0765629   .0025511   -30.01   0.000     -.081563   -.0715628
  hourswater |   .5061264   .0369192    13.71   0.000     .4337662    .5784867
 hourswater2 |  -.0268429   .0020651   -13.00   0.000    -.0308905   -.0227954
hourselect~y |   1.506549   .0698072    21.58   0.000     1.369729    1.643369
hourselect~2 |  -.0792909   .0045221   -17.53   0.000     -.088154   -.0704278
------------------------------------------------------------------------------

est store partial

hausman partial all, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
        rent |   -.0765629    -.0816847        .0051218        .0002556
  hourswater |    .5061264     .5863762       -.0802497        .0064536
 hourswater2 |   -.0268429    -.0309267        .0040838        .0001297
hourselect~y |    1.506549     1.572483       -.0659337        .0112385
hourselect~2 |   -.0792909    -.0833422        .0040513        .0004585
------------------------------------------------------------------------------
                        b = consistent under Ho and Ha; obtained from asclogit
         B = inconsistent under Ha, efficient under Ho; obtained from asclogit

    Test:  Ho:  difference in coefficients not systematic

                  chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       75.36
```

```
               Prob>chi2 =      0.0000
               (V_b-V_B is not positive definite)


asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, case(id)
alternatives(area) noconstant

Iteration 0:   log likelihood = -978.17354
Iteration 1:   log likelihood = -944.40738
Iteration 2:   log likelihood = -939.62837
Iteration 3:   log likelihood = -939.58951

Alternative-specific conditional logit         Number of obs      =      14448
Case variable: id                              Number of cases    =       2408

Alternative variable: area                     Alts per case: min =          6
                                                              avg =        6.0
                                                              max =          6

                                                  Wald chi2(5)    =    1232.19
Log likelihood =  -939.5895                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0816847   .0025383   -32.18   0.000    -.0866596   -.0767098
  hourswater |   .5863762   .0363507    16.13   0.000       .51513    .6576223
 hourswater2 |  -.0309267    .002061   -15.01   0.000    -.0349662   -.0268872
hourselect~y |   1.572483   .0688966    22.82   0.000     1.437448    1.707518
hourselect~2 |  -.0833422   .0044988   -18.53   0.000    -.0921596   -.0745247
------------------------------------------------------------------------------

est store all

asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 6, case(id)
alternatives(area) noconstant
note: 845 cases (4225 obs) dropped due to no positive outcome per case

Iteration 0:   log likelihood = -577.42021
Iteration 1:   log likelihood = -570.68496
Iteration 2:   log likelihood = -570.57189
Iteration 3:   log likelihood = -570.57182
Iteration 4:   log likelihood = -570.57182

Alternative-specific conditional logit         Number of obs      =       7845
Case variable: id                              Number of cases    =       1569

Alternative variable: area                     Alts per case: min =          5
                                                              avg =        5.0
                                                              max =          5

                                                  Wald chi2(5)    =     691.46
Log likelihood = -570.57182                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0760328   .0030758   -24.72   0.000    -.0820613   -.0700043
  hourswater |   .4830981   .0437156    11.05   0.000     .3974172     .568779
 hourswater2 |  -.0240531    .002437    -9.87   0.000    -.0288295   -.0192768
hourselect~y |   1.525461   .0891237    17.12   0.000     1.350782    1.700141
hourselect~2 |  -.0762712   .0053816   -14.17   0.000    -.0868189   -.0657234
------------------------------------------------------------------------------

est store partial

hausman partial all, alleqs constant

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
        rent |   -.0760328    -.0816847        .0056519        .0017372
  hourswater |    .4830981     .5863762       -.1032781        .0242832
 hourswater2 |   -.0240531    -.0309267        .0068736        .0013004
hourselect~y |    1.525461     1.572483       -.0470215        .0565358
hourselect~2 |   -.0762712    -.0833422         .007071        .0029534
------------------------------------------------------------------------------
                        b = consistent under Ho and Ha; obtained from asclogit
         B = inconsistent under Ha, efficient under Ho; obtained from asclogit

    Test:  Ho:  difference in coefficients not systematic

                 chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                         =      150.90
               Prob>chi2 =      0.0000
```

Hausman Test of IIA - asclogit (mixed logit)

```
asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, case(id)
alternatives(area) casevars(income age yearsofedu)

Iteration 0:   log likelihood = -588.85292
Iteration 1:   log likelihood =  -567.6415
Iteration 2:   log likelihood = -560.27514
Iteration 3:   log likelihood = -560.16378
Iteration 4:   log likelihood = -560.16368
Iteration 5:   log likelihood = -560.16368

Alternative-specific conditional logit         Number of obs      =      14100
Case variable: id                              Number of cases    =       2350

Alternative variable: area                     Alts per case: min =          6
                                                              avg =        6.0
                                                              max =          6

                                                  Wald chi2(20)   =     799.13
Log likelihood = -560.16368                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0884401    .003451   -25.63   0.000    -.0952039   -.0816763
  hourswater |    .971236   .0755669    12.85   0.000     .8231276    1.119345
 hourswater2 |  -.0391148     .00402    -9.73   0.000    -.0469939   -.0312357
hourselect~y |   1.481254   .0922275    16.06   0.000     1.300491    1.662016
hourselect~2 |  -.0776209   .0059152   -13.12   0.000    -.0892145   -.0660273
-------------+----------------------------------------------------------------
1            |   (base alternative)
-------------+----------------------------------------------------------------
2            |
      income |   .0003719   .0002573     1.45   0.148    -.0001323    .0008762
         age |   .0497745   .0177274     2.81   0.005     .0150294    .0845195
  yearsofedu |   .1517774   .0422894     3.59   0.000     .0688916    .2346631
       _cons |  -2.242399   .8603667    -2.61   0.009    -3.928686   -.5561109
-------------+----------------------------------------------------------------
3            |
      income |  -.0010744   .0007707    -1.39   0.163    -.0025849    .0004361
         age |   .0194763     .01788     1.09   0.276    -.0155678    .0545205
  yearsofedu |  -.0967144    .034396    -2.81   0.005    -.1641293   -.0292995
       _cons |   1.827903   .8049947     2.27   0.023     .2501421    3.405663
-------------+----------------------------------------------------------------
4            |
      income |  -.0003516   .0005207    -0.68   0.499    -.0013722     .000669
         age |   .0568159   .0284399     2.00   0.046     .0010747    .1125572
  yearsofedu |    .146708   .0704001     2.08   0.037     .0087263    .2846897
       _cons |  -6.942976   1.485985    -4.67   0.000    -9.855453   -4.030499
-------------+----------------------------------------------------------------
5            |
      income |  -.0005556   .0004731    -1.17   0.240    -.0014828    .0003715
         age |  -.0026325   .0235372    -0.11   0.911    -.0487645    .0434995
  yearsofedu |  -.0146094    .045428    -0.32   0.748    -.1036467    .0744279
       _cons |   1.949125   1.045932     1.86   0.062    -.1008649    3.999114
-------------+----------------------------------------------------------------
6            |
      income |  -.0004237    .000277    -1.53   0.126    -.0009666    .0001193
         age |    .027311   .0135744     2.01   0.044     .0007058    .0539163
  yearsofedu |  -.0176924   .0271537    -0.65   0.515    -.0709126    .0355278
       _cons |    1.90578   .6293261     3.03   0.002     .6723233    3.139236
------------------------------------------------------------------------------

. est store all

. asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 1, case(id)
alternatives(area) casevars(income age yearsofedu)
note: 692 cases (3460 obs) dropped due to no positive outcome per case

Iteration 0:   log likelihood = -329.99235
Iteration 1:   log likelihood = -275.44162
Iteration 2:   log likelihood = -268.65088
Iteration 3:   log likelihood = -268.42426
Iteration 4:   log likelihood = -268.42346
Iteration 5:   log likelihood = -268.42346

Alternative-specific conditional logit         Number of obs      =       8295
Case variable: id                              Number of cases    =       1659

Alternative variable: area                     Alts per case: min =          5
                                                              avg =        5.0
                                                              max =          5

                                                  Wald chi2(17)   =     432.07
Log likelihood = -268.42346                       Prob > chi2     =     0.0000
```

```
------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent | -.0715648    .003949   -18.12   0.000    -.0793047   -.0638248
  hourswater |  1.629801    .125277    13.01   0.000     1.384263    1.875339
 hourswater2 | -.0697872   .0064152   -10.88   0.000    -.0823607   -.0572137
hourselect~y |  1.355891   .1215777    11.15   0.000     1.117603    1.594179
hourselect~2 | -.0789218   .0079351    -9.95   0.000    -.0944744   -.0633692
-------------+----------------------------------------------------------------
2            |  (base alternative)
-------------+----------------------------------------------------------------
3            |
      income | -.0024579   .0011157    -2.20   0.028    -.0046446   -.0002712
         age | -.0285966   .0276021    -1.04   0.300    -.0826957    .0255025
  yearsofedu | -.2440917   .0564992    -4.32   0.000    -.3548281   -.1333554
       _cons |  6.193811    1.32116     4.69   0.000     3.604385    8.783237
-------------+----------------------------------------------------------------
4            |
      income | -.0017484   .0005375    -3.25   0.001    -.0028018   -.0006949
         age |  .0181149   .0290234     0.62   0.533    -.0387699    .0749996
  yearsofedu | -.0366802   .0676069    -0.54   0.587    -.1691872    .0958269
       _cons | -4.328699   1.511846    -2.86   0.004    -7.291862   -1.365535
-------------+----------------------------------------------------------------
5            |
      income | -.0013365   .0004682    -2.85   0.004    -.0022542   -.0004188
         age | -.0558372   .0307842    -1.81   0.070    -.1161731    .0044986
  yearsofedu | -.1885351   .0622983    -3.03   0.002    -.3106375   -.0664326
       _cons |  6.409064   1.463791     4.38   0.000     3.540086    9.278042
-------------+----------------------------------------------------------------
6            |
      income | -.0014659   .0003182    -4.61   0.000    -.0020896   -.0008423
         age | -.0344019   .0192457    -1.79   0.074    -.0721228     .003319
  yearsofedu | -.1893883   .0441662    -4.29   0.000    -.2759525    -.102824
       _cons |  6.310701   .9925926     6.36   0.000     4.365255    8.256146
------------------------------------------------------------------------------

. est store partial

. hausman partial all, alleqs constant

Note: the rank of the differenced variance matrix (19) does not equal the number of coefficients being
tested (21); be sure this is what you expect, or there may be problems computing the test.  Examine the
output of your estimators for anything unexpected and possibly consider scaling your variables so that the
coefficients are on a similar scale.

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
area         |
        rent |   -.0715648    -.0884401        .0168754        .0019198
  hourswater |    1.629801      .971236         .658565        .0999198
 hourswater2 |   -.0697872    -.0391148       -.0306724        .0049994
hourselect~y |    1.355891     1.481254       -.1253625        .0792163
hourselect~2 |   -.0789218    -.0776209       -.0013009        .0052893
-------------+----------------------------------------------------------------
3            |
      income |   -.0024579    -.0010744       -.0013836        .0008067
         age |   -.0285966     .0194763       -.0480729        .0210281
  yearsofedu |   -.2440917    -.0967144       -.1473774        .0448227
       _cons |    6.193811     1.827903        4.365908        1.047591
-------------+----------------------------------------------------------------
4            |
      income |   -.0017484    -.0003516       -.0013967        .0001331
         age |    .0181149     .0568159       -.0387011        .0057903
  yearsofedu |   -.0366802      .146708       -.1833882               .
       _cons |   -4.328699    -6.942976        2.614277        .2784354
-------------+----------------------------------------------------------------
5            |
      income |   -.0013365    -.0005556       -.0007808               .
         age |   -.0558372    -.0026325       -.0532047         .019841
  yearsofedu |   -.1885351    -.0146094       -.1739257        .0426307
       _cons |    6.409064     1.949125         4.45994        1.024066
-------------+----------------------------------------------------------------
6            |
      income |   -.0014659    -.0004237       -.0010423        .0001565
         age |   -.0344019      .027311        -.061713        .0136431
  yearsofedu |   -.1893883    -.0176924       -.1716959         .034833
       _cons |    6.310701      1.90578        4.404921        .7675863
------------------------------------------------------------------------------
                        b = consistent under Ho and Ha; obtained from asclogit
         B = inconsistent under Ha, efficient under Ho; obtained from asclogit

    Test:  Ho:  difference in coefficients not systematic

                 chi2(19) = (b-B)'[(V_b-V_B)^(-1)](b-B)
```

```
                  =        2.41
                Prob>chi2 =      1.0000
                (V_b-V_B is not positive definite)

asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, case(id)
alternatives(area) casevars(income age yearsofedu)

Iteration 0:   log likelihood = -588.85292
Iteration 1:   log likelihood =  -567.6415
Iteration 2:   log likelihood = -560.27514
Iteration 3:   log likelihood = -560.16378
Iteration 4:   log likelihood = -560.16368
Iteration 5:   log likelihood = -560.16368

Alternative-specific conditional logit         Number of obs      =     14100
Case variable: id                              Number of cases    =      2350

Alternative variable: area                     Alts per case: min =         6
                                                              avg =       6.0
                                                              max =         6

                                                  Wald chi2(20)   =    799.13
Log likelihood = -560.16368                       Prob > chi2     =    0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0884401    .003451   -25.63   0.000    -.0952039   -.0816763
  hourswater |    .971236   .0755669    12.85   0.000     .8231276    1.119345
 hourswater2 |  -.0391148     .00402    -9.73   0.000    -.0469939   -.0312357
hourselect~y |   1.481254   .0922275    16.06   0.000     1.300491    1.662016
hourselect~2 |  -.0776209   .0059152   -13.12   0.000    -.0892145   -.0660273
-------------+----------------------------------------------------------------
1            |  (base alternative)
-------------+----------------------------------------------------------------
2            |
      income |   .0003719   .0002573     1.45   0.148    -.0001323    .0008762
         age |   .0497745   .0177274     2.81   0.005     .0150294    .0845195
  yearsofedu |   .1517774   .0422894     3.59   0.000     .0688916    .2346631
       _cons |  -2.242399   .8603667    -2.61   0.009    -3.928686   -.5561109
-------------+----------------------------------------------------------------
3            |
      income |  -.0010744   .0007707    -1.39   0.163    -.0025849    .0004361
         age |   .0194763     .01788     1.09   0.276    -.0155678    .0545205
  yearsofedu |  -.0967144    .034396    -2.81   0.005    -.1641293   -.0292995
       _cons |   1.827903   .8049947     2.27   0.023     .2501421    3.405663
-------------+----------------------------------------------------------------
4            |
      income |  -.0003516   .0005207    -0.68   0.499    -.0013722     .000669
         age |   .0568159   .0284399     2.00   0.046     .0010747    .1125572
  yearsofedu |    .146708   .0704001     2.08   0.037     .0087263    .2846897
       _cons |  -6.942976   1.485985    -4.67   0.000    -9.855453   -4.030499
-------------+----------------------------------------------------------------
5            |
      income |  -.0005556   .0004731    -1.17   0.240    -.0014828    .0003715
         age |  -.0026325   .0235372    -0.11   0.911    -.0487645    .0434995
  yearsofedu |  -.0146094    .045428    -0.32   0.748    -.1036467    .0744279
       _cons |   1.949125   1.045932     1.86   0.062    -.1008649    3.999114
-------------+----------------------------------------------------------------
6            |
      income |  -.0004237    .000277    -1.53   0.126    -.0009666    .0001193
         age |    .027311   .0135744     2.01   0.044     .0007058    .0539163
  yearsofedu |  -.0176924   .0271537    -0.65   0.515    -.0709126    .0355278
       _cons |    1.90578   .6293261     3.03   0.002     .6723233    3.139236
------------------------------------------------------------------------------

. est store all

. asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 2, case(id)
alternatives(area) casevars(income age yearsofedu)
note: 266 cases (1330 obs) dropped due to no positive outcome per case

Iteration 0:   log likelihood = -402.14057
Iteration 1:   log likelihood = -385.17406
Iteration 2:   log likelihood =  -365.8791
Iteration 3:   log likelihood = -362.27615
Iteration 4:   log likelihood = -362.21636
Iteration 5:   log likelihood = -362.21634

Alternative-specific conditional logit         Number of obs      =     10440
Case variable: id                              Number of cases    =      2088

Alternative variable: area                     Alts per case: min =         5
                                                              avg =       5.0
                                                              max =         5

                                                  Wald chi2(17)   =    573.68
```

```
Log likelihood = -362.21634                     Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0936598   .0044939   -20.84   0.000    -.1024676   -.0848519
  hourswater |   1.218025   .1079169    11.29   0.000     1.006512    1.429538
 hourswater2 |  -.0524745   .0056615    -9.27   0.000    -.0635708   -.0413781
hourselect~y |   1.381448   .1102941    12.53   0.000     1.165275     1.59762
hourselect~2 |    -.08017   .0083541    -9.60   0.000    -.0965437   -.0637963
-------------+----------------------------------------------------------------
1            |  (base alternative)
-------------+----------------------------------------------------------------
3            |
      income |  -.0013379   .0007951    -1.68   0.092    -.0028962    .0002205
         age |   .0207966   .0207394     1.00   0.316    -.0198518     .061445
  yearsofedu |  -.1039881   .0399492    -2.60   0.009     -.182287   -.0256891
       _cons |   2.190268   .9437259     2.32   0.020     .3405988    4.039936
-------------+----------------------------------------------------------------
4            |
      income |  -.0009622   .0006018    -1.60   0.110    -.0021417    .0002173
         age |    .060282   .0301267     2.00   0.045     .0012347    .1193292
  yearsofedu |   .1464467   .0776934     1.88   0.059    -.0058295    .2987228
       _cons |  -6.748403   1.575769    -4.28   0.000    -9.836853   -3.659953
-------------+----------------------------------------------------------------
5            |
      income |  -.0008683   .0004711    -1.84   0.065    -.0017917     .000055
         age |  -.0047338   .0260908    -0.18   0.856    -.0558709    .0464033
  yearsofedu |  -.0219659   .0509686    -0.43   0.666    -.1218626    .0779308
       _cons |   2.560454   1.162964     2.20   0.028     .2810875    4.839821
-------------+----------------------------------------------------------------
6            |
      income |  -.0007988   .0003041    -2.63   0.009    -.0013948   -.0002029
         age |   .0283345   .0149021     1.90   0.057    -.0008732    .0575422
  yearsofedu |  -.0231906   .0302642    -0.77   0.444    -.0825074    .0361262
       _cons |    2.26326   .6953485     3.25   0.001     .9004018    3.626118
------------------------------------------------------------------------------

. est store partial

. hausman partial all, alleqs constant

Note: the rank of the differenced variance matrix (19) does not equal the number of coefficients being
tested (21); be sure this is what you expect, or there may be problems computing the test.  Examine the
output of your estimators for anything unexpected and possibly consider scaling your variables so that the
coefficients are on a similar scale.

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
area         |
        rent |   -.0936598    -.0884401       -.0052196        .0028785
  hourswater |    1.218025      .971236         .246789        .0770435
 hourswater2 |   -.0524745    -.0391148       -.0133597        .0039865
hourselect~y |    1.381448     1.481254        -.099806        .0604885
hourselect~2 |     -.08017    -.0776209       -.0025491        .0058992
-------------+----------------------------------------------------------------
3            |
      income |   -.0013379    -.0010744       -.0002635        .0001955
         age |    .0207966     .0194763        .0013203        .0105084
  yearsofedu |   -.1039881    -.0967144       -.0072737        .0203188
       _cons |    2.190268     1.827903        .3623648        .4925465
-------------+----------------------------------------------------------------
4            |
      income |   -.0009622    -.0003516       -.0006106        .0003016
         age |     .060282     .0568159         .003466        .0099393
  yearsofedu |    .1464467      .146708       -.0002613        .0328646
       _cons |   -6.748403    -6.942976        .1945732        .5243049
-------------+----------------------------------------------------------------
5            |
      income |   -.0008683    -.0005556       -.0003127               .
         age |   -.0047338    -.0026325       -.0021013        .0112576
  yearsofedu |   -.0219659    -.0146094       -.0073565        .0231105
       _cons |    2.560454     1.949125        .6113295        .5084387
-------------+----------------------------------------------------------------
6            |
      income |   -.0007988    -.0004237       -.0003752        .0001253
         age |    .0283345      .027311        .0010235        .0061491
  yearsofedu |   -.0231906    -.0176924       -.0054982        .0133643
       _cons |     2.26326      1.90578          .35748        .2957333
------------------------------------------------------------------------------
                        b = consistent under Ho and Ha; obtained from asclogit
         B = inconsistent under Ha, efficient under Ho; obtained from asclogit

    Test:  Ho:  difference in coefficients not systematic
```

```
                 chi2(19) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       72.47
                Prob>chi2 =      0.0000
                (V_b-V_B is not positive definite)

asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, case(id)
alternatives(area) casevars(income age yearsofedu)

Iteration 0:   log likelihood = -588.85292
Iteration 1:   log likelihood =  -567.6415
Iteration 2:   log likelihood = -560.27514
Iteration 3:   log likelihood = -560.16378
Iteration 4:   log likelihood = -560.16368
Iteration 5:   log likelihood = -560.16368

Alternative-specific conditional logit         Number of obs      =      14100
Case variable: id                              Number of cases    =       2350

Alternative variable: area                     Alts per case: min =          6
                                                              avg =        6.0
                                                              max =          6

                                                  Wald chi2(20)   =     799.13
Log likelihood = -560.16368                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0884401    .003451   -25.63   0.000    -.0952039   -.0816763
  hourswater |    .971236   .0755669    12.85   0.000     .8231276    1.119345
 hourswater2 |  -.0391148     .00402    -9.73   0.000    -.0469939   -.0312357
hourselect~y |   1.481254   .0922275    16.06   0.000     1.300491    1.662016
hourselect~2 |  -.0776209   .0059152   -13.12   0.000    -.0892145   -.0660273
-------------+----------------------------------------------------------------
1            |  (base alternative)
-------------+----------------------------------------------------------------
2            |
      income |   .0003719   .0002573     1.45   0.148    -.0001323    .0008762
         age |   .0497745   .0177274     2.81   0.005     .0150294    .0845195
  yearsofedu |   .1517774   .0422894     3.59   0.000     .0688916    .2346631
       _cons |  -2.242399   .8603667    -2.61   0.009    -3.928686   -.5561109
-------------+----------------------------------------------------------------
3            |
      income |  -.0010744   .0007707    -1.39   0.163    -.0025849    .0004361
         age |   .0194763     .01788     1.09   0.276    -.0155678    .0545205
  yearsofedu |  -.0967144    .034396    -2.81   0.005    -.1641293   -.0292995
       _cons |   1.827903   .8049947     2.27   0.023     .2501421    3.405663
-------------+----------------------------------------------------------------
4            |
      income |  -.0003516   .0005207    -0.68   0.499    -.0013722     .000669
         age |   .0568159   .0284399     2.00   0.046     .0010747    .1125572
  yearsofedu |    .146708   .0704001     2.08   0.037     .0087263    .2846897
       _cons |  -6.942976   1.485985    -4.67   0.000    -9.855453   -4.030499
-------------+----------------------------------------------------------------
5            |
      income |  -.0005556   .0004731    -1.17   0.240    -.0014828    .0003715
         age |  -.0026325   .0235372    -0.11   0.911    -.0487645    .0434995
  yearsofedu |  -.0146094    .045428    -0.32   0.748    -.1036467    .0744279
       _cons |   1.949125   1.045932     1.86   0.062    -.1008649    3.999114
-------------+----------------------------------------------------------------
6            |
      income |  -.0004237    .000277    -1.53   0.126    -.0009666    .0001193
         age |    .027311   .0135744     2.01   0.044     .0007058    .0539163
  yearsofedu |  -.0176924   .0271537    -0.65   0.515    -.0709126    .0355278
       _cons |    1.90578   .6293261     3.03   0.002     .6723233    3.139236
------------------------------------------------------------------------------

est store all

asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 3, case(id)
alternatives(area) casevars(income age yearsofedu)
note: 231 cases (1155 obs) dropped due to no positive outcome per case

Iteration 0:   log likelihood = -512.86786
Iteration 1:   log likelihood = -493.33747
Iteration 2:   log likelihood = -488.28996
Iteration 3:   log likelihood = -488.25396
Iteration 4:   log likelihood = -488.25394

Alternative-specific conditional logit         Number of obs      =      10600
Case variable: id                              Number of cases    =       2120

Alternative variable: area                     Alts per case: min =          5
                                                              avg =        5.0
                                                              max =          5

                                                  Wald chi2(17)   =     670.75
```

```
Log likelihood = -488.25394                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0767465   .0033076   -23.20   0.000    -.0832293   -.0702637
  hourswater |   .6380906    .072055     8.86   0.000     .4968653    .7793158
 hourswater2 |  -.0229298    .003875    -5.92   0.000    -.0305247   -.0153348
hourselect~y |   1.526679   .0930448    16.41   0.000     1.344315    1.709044
hourselect~2 |   -.080463    .005795   -13.88   0.000    -.0918209   -.0691051
-------------+----------------------------------------------------------------
1            |  (base alternative)
-------------+----------------------------------------------------------------
2            |
      income |   .0004304   .0002537     1.70   0.090    -.0000669    .0009277
         age |   .0447098   .0170846     2.62   0.009     .0112245    .0781951
  yearsofedu |   .1456045   .0402645     3.62   0.000     .0666875    .2245215
       _cons |  -2.509355   .8311625    -3.02   0.003    -4.138404   -.8803069
-------------+----------------------------------------------------------------
4            |
      income |  -.0003018   .0005238    -0.58   0.565    -.0013284    .0007249
         age |   .0546422   .0276286     1.98   0.048     .0004912    .1087932
  yearsofedu |   .1369151   .0661089     2.07   0.038      .007344    .2664863
       _cons |  -6.816114   1.450233    -4.70   0.000    -9.658518    -3.97371
-------------+----------------------------------------------------------------
5            |
      income |  -.0004693   .0004431    -1.06   0.290    -.0013378    .0003993
         age |  -.0006206   .0208961    -0.03   0.976    -.0415763    .0403351
  yearsofedu |   -.016433   .0406955    -0.40   0.686    -.0961947    .0633287
       _cons |   1.081596   .9303602     1.16   0.245    -.7418761    2.905069
-------------+----------------------------------------------------------------
6            |
      income |   -.000341    .000261    -1.31   0.191    -.0008525    .0001705
         age |   .0270033   .0127389     2.12   0.034     .0020355    .0519711
  yearsofedu |  -.0215709   .0257422    -0.84   0.402    -.0720247    .0288829
       _cons |   1.317124   .5966041     2.21   0.027     .1478019    2.486447
------------------------------------------------------------------------------

. est store partial

hausman partial all, alleqs constant

Note: the rank of the differenced variance matrix (19) does not equal the number of coefficients being
tested (21); be sure this is what you expect, or there may be problems computing the test.  Examine the
output of your estimators for anything unexpected and possibly consider scaling your variables so that the
coefficients are on a similar scale.

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
area         |
        rent |   -.0767465    -.0884401        .0116936               .
  hourswater |    .6380906      .971236       -.3331455               .
 hourswater2 |   -.0229298    -.0391148         .016185               .
hourselect~y |    1.526679     1.481254        .0454254        .0123048
hourselect~2 |    -.080463    -.0776209       -.0028421               .
-------------+----------------------------------------------------------------
2            |
      income |    .0004304     .0003719        .0000585               .
         age |    .0447098     .0497745       -.0050647               .
  yearsofedu |    .1456045     .1517774       -.0061728               .
       _cons |   -2.509355    -2.242399       -.2669568               .
-------------+----------------------------------------------------------------
4            |
      income |   -.0003018    -.0003516        .0000499        .0000567
         age |    .0546422     .0568159       -.0021737               .
  yearsofedu |    .1369151      .146708       -.0097929               .
       _cons |   -6.816114    -6.942976        .1268619               .
-------------+----------------------------------------------------------------
5            |
      income |   -.0004693    -.0005556        .0000864               .
         age |   -.0006206    -.0026325        .0020119               .
  yearsofedu |    -.016433    -.0146094       -.0018236               .
       _cons |    1.081596     1.949125       -.8675284               .
-------------+----------------------------------------------------------------
6            |
      income |    -.000341    -.0004237        .0000827               .
         age |    .0270033      .027311       -.0003077               .
  yearsofedu |   -.0215709    -.0176924       -.0038785               .
       _cons |    1.317124      1.90578       -.5886553               .
------------------------------------------------------------------------------
                        b = consistent under Ho and Ha; obtained from asclogit
         B = inconsistent under Ha, efficient under Ho; obtained from asclogit

    Test:  Ho:  difference in coefficients not systematic
```

```
                chi2(19) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                         =  -120.33    chi2<0 ==> model fitted on these
                                       data fails to meet the asymptotic
                                       assumptions of the Hausman test;
                                       see suest for a generalized test

asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, case(id)
alternatives(area) casevars(income age yearsofedu)

Iteration 0:   log likelihood = -588.85292
Iteration 1:   log likelihood =  -567.6415
Iteration 2:   log likelihood = -560.27514
Iteration 3:   log likelihood = -560.16378
Iteration 4:   log likelihood = -560.16368
Iteration 5:   log likelihood = -560.16368

Alternative-specific conditional logit         Number of obs      =      14100
Case variable: id                              Number of cases    =       2350

Alternative variable: area                     Alts per case: min =          6
                                                              avg =        6.0
                                                              max =          6

                                                  Wald chi2(20)   =     799.13
Log likelihood = -560.16368                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0884401    .003451   -25.63   0.000    -.0952039   -.0816763
  hourswater |    .971236   .0755669    12.85   0.000     .8231276    1.119345
 hourswater2 |  -.0391148     .00402    -9.73   0.000    -.0469939   -.0312357
hourselect~y |   1.481254   .0922275    16.06   0.000     1.300491    1.662016
hourselect~2 |  -.0776209   .0059152   -13.12   0.000    -.0892145   -.0660273
-------------+----------------------------------------------------------------
1            |  (base alternative)
-------------+----------------------------------------------------------------
2            |
      income |   .0003719   .0002573     1.45   0.148    -.0001323    .0008762
         age |   .0497745   .0177274     2.81   0.005     .0150294    .0845195
  yearsofedu |   .1517774   .0422894     3.59   0.000     .0688916    .2346631
       _cons |  -2.242399   .8603667    -2.61   0.009    -3.928686   -.5561109
-------------+----------------------------------------------------------------
3            |
      income |  -.0010744   .0007707    -1.39   0.163    -.0025849    .0004361
         age |   .0194763     .01788     1.09   0.276    -.0155678    .0545205
  yearsofedu |  -.0967144    .034396    -2.81   0.005    -.1641293   -.0292995
       _cons |   1.827903   .8049947     2.27   0.023     .2501421    3.405663
-------------+----------------------------------------------------------------
4            |
      income |  -.0003516   .0005207    -0.68   0.499    -.0013722     .000669
         age |   .0568159   .0284399     2.00   0.046     .0010747    .1125572
  yearsofedu |    .146708   .0704001     2.08   0.037     .0087263    .2846897
       _cons |  -6.942976   1.485985    -4.67   0.000    -9.855453   -4.030499
-------------+----------------------------------------------------------------
5            |
      income |  -.0005556   .0004731    -1.17   0.240    -.0014828    .0003715
         age |  -.0026325   .0235372    -0.11   0.911    -.0487645    .0434995
  yearsofedu |  -.0146094    .045428    -0.32   0.748    -.1036467    .0744279
       _cons |   1.949125   1.045932     1.86   0.062    -.1008649    3.999114
-------------+----------------------------------------------------------------
6            |
      income |  -.0004237    .000277    -1.53   0.126    -.0009666    .0001193
         age |    .027311   .0135744     2.01   0.044     .0007058    .0539163
  yearsofedu |  -.0176924   .0271537    -0.65   0.515    -.0709126    .0355278
       _cons |    1.90578   .6293261     3.03   0.002     .6723233    3.139236
------------------------------------------------------------------------------

. est store all

asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 4, case(id)
alternatives(area) casevars(income age yearsofedu)
note: 70 cases (350 obs) dropped due to no positive outcome per case

Iteration 0:   log likelihood = -526.94919
Iteration 1:   log likelihood = -508.32818
Iteration 2:   log likelihood = -501.60365
Iteration 3:   log likelihood = -501.48259
Iteration 4:   log likelihood =  -501.4825
Iteration 5:   log likelihood =  -501.4825

Alternative-specific conditional logit         Number of obs      =      11405
Case variable: id                              Number of cases    =       2281

Alternative variable: area                     Alts per case: min =          5
                                                              avg =        5.0
                                                              max =          5
```

```
                                                  Wald chi2(17)   =     730.60
Log likelihood =  -501.4825                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0866356   .0035177   -24.63   0.000    -.0935301   -.0797411
  hourswater |   1.062628   .0810684    13.11   0.000     .9037372     1.22152
 hourswater2 |  -.0455656   .0042552   -10.71   0.000    -.0539057   -.0372256
hourselect~y |   1.417168   .0926227    15.30   0.000     1.235631    1.598705
hourselect~2 |  -.0726823   .0058812   -12.36   0.000    -.0842092   -.0611554
-------------+----------------------------------------------------------------
1            |  (base alternative)
-------------+----------------------------------------------------------------
2            |
      income |   .0003628   .0002513     1.44   0.149    -.0001298    .0008553
         age |   .0468642   .0176001     2.66   0.008     .0123686    .0813597
  yearsofedu |   .1576626   .0421594     3.74   0.000     .0750317    .2402935
       _cons |  -2.256859   .8452758    -2.67   0.008     -3.91357   -.6001493
-------------+----------------------------------------------------------------
3            |
      income |  -.0010556    .000775    -1.36   0.173    -.0025746    .0004633
         age |   .0196669   .0181311     1.08   0.278    -.0158695    .0552033
  yearsofedu |   -.094829   .0346866    -2.73   0.006    -.1628134   -.0268445
       _cons |     1.9054   .8158847     2.34   0.020     .3062952    3.504504
-------------+----------------------------------------------------------------
5            |
      income |  -.0005346    .000471    -1.14   0.256    -.0014578    .0003886
         age |  -.0024684   .0235692    -0.10   0.917    -.0486631    .0437264
  yearsofedu |   -.012905   .0455769    -0.28   0.777    -.1022341    .0764242
       _cons |   2.010765   1.046621     1.92   0.055    -.0405744    4.062104
-------------+----------------------------------------------------------------
6            |
      income |  -.0004233   .0002762    -1.53   0.125    -.0009647     .000118
         age |   .0262601   .0134913     1.95   0.052    -.0001824    .0527026
  yearsofedu |  -.0172393    .026946    -0.64   0.522    -.0700525    .0355738
       _cons |    1.96312   .6224674     3.15   0.002     .7431068    3.183134
------------------------------------------------------------------------------

. est store partial

. hausman partial all, alleqs constant

Note: the rank of the differenced variance matrix (20) does not equal the number of coefficients being
tested (21); be sure this is what you expect, or there may be problems computing the test.  Examine the
output of your estimators for anything unexpected and possibly consider scaling your variables so that the
coefficients are on a similar scale.

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
area         |
        rent |   -.0866356    -.0884401        .0018045        .0006817
  hourswater |    1.062628      .971236        .0913923        .0293552
 hourswater2 |   -.0455656    -.0391148       -.0064508         .001395
hourselect~y |    1.417168     1.481254       -.0640859        .0085469
hourselect~2 |   -.0726823    -.0776209        .0049386               .
-------------+----------------------------------------------------------------
2            |
      income |    .0003628     .0003719       -9.16e-06               .
         age |    .0468642     .0497745       -.0029103               .
  yearsofedu |    .1576626     .1517774        .0058853               .
       _cons |   -2.256859    -2.242399       -.0144608               .
-------------+----------------------------------------------------------------
3            |
      income |   -.0010556    -.0010744        .0000187        .0000815
         age |    .0196669     .0194763        .0001905        .0030073
  yearsofedu |    -.094829    -.0967144        .0018854        .0044805
       _cons |      1.9054     1.827903        .0774969        .1328581
-------------+----------------------------------------------------------------
5            |
      income |   -.0005346    -.0005556         .000021               .
         age |   -.0024684    -.0026325        .0001642        .0012276
  yearsofedu |    -.012905    -.0146094        .0017044        .0036809
       _cons |    2.010765     1.949125          .06164        .0379585
-------------+----------------------------------------------------------------
6            |
      income |   -.0004233    -.0004237        3.24e-07               .
         age |    .0262601      .027311       -.0010509               .
  yearsofedu |   -.0172393    -.0176924        .0004531               .
       _cons |     1.96312      1.90578        .0573407               .
------------------------------------------------------------------------------
                        b = consistent under Ho and Ha; obtained from asclogit
         B = inconsistent under Ha, efficient under Ho; obtained from asclogit
```

```
    Test:  Ho:  difference in coefficients not systematic

                 chi2(20) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =   -17.01    chi2<0 ==> model fitted on these
                                        data fails to meet the asymptotic
                                        assumptions of the Hausman test;
                                        see suest for a generalized test

. asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, case(id)
alternatives(area) casevars(income age yearsofedu)

Iteration 0:   log likelihood = -588.85292
Iteration 1:   log likelihood =  -567.6415
Iteration 2:   log likelihood = -560.27514
Iteration 3:   log likelihood = -560.16378
Iteration 4:   log likelihood = -560.16368
Iteration 5:   log likelihood = -560.16368

Alternative-specific conditional logit         Number of obs      =      14100
Case variable: id                              Number of cases    =       2350

Alternative variable: area                     Alts per case: min =          6
                                                              avg =        6.0
                                                              max =          6

                                                  Wald chi2(20)   =     799.13
Log likelihood = -560.16368                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0884401    .003451   -25.63   0.000    -.0952039   -.0816763
  hourswater |    .971236   .0755669    12.85   0.000     .8231276    1.119345
 hourswater2 |  -.0391148     .00402    -9.73   0.000    -.0469939   -.0312357
hourselect~y |   1.481254   .0922275    16.06   0.000     1.300491    1.662016
hourselect~2 |  -.0776209   .0059152   -13.12   0.000    -.0892145   -.0660273
-------------+----------------------------------------------------------------
1            |  (base alternative)
-------------+----------------------------------------------------------------
2            |
      income |   .0003719   .0002573     1.45   0.148    -.0001323    .0008762
         age |   .0497745   .0177274     2.81   0.005     .0150294    .0845195
  yearsofedu |   .1517774   .0422894     3.59   0.000     .0688916    .2346631
       _cons |  -2.242399   .8603667    -2.61   0.009    -3.928686   -.5561109
-------------+----------------------------------------------------------------
3            |
      income |  -.0010744   .0007707    -1.39   0.163    -.0025849    .0004361
         age |   .0194763     .01788     1.09   0.276    -.0155678    .0545205
  yearsofedu |  -.0967144    .034396    -2.81   0.005    -.1641293   -.0292995
       _cons |   1.827903   .8049947     2.27   0.023     .2501421    3.405663
-------------+----------------------------------------------------------------
4            |
      income |  -.0003516   .0005207    -0.68   0.499    -.0013722     .000669
         age |   .0568159   .0284399     2.00   0.046     .0010747    .1125572
  yearsofedu |    .146708   .0704001     2.08   0.037     .0087263    .2846897
       _cons |  -6.942976   1.485985    -4.67   0.000    -9.855453   -4.030499
-------------+----------------------------------------------------------------
5            |
      income |  -.0005556   .0004731    -1.17   0.240    -.0014828    .0003715
         age |  -.0026325   .0235372    -0.11   0.911    -.0487645    .0434995
  yearsofedu |  -.0146094    .045428    -0.32   0.748    -.1036467    .0744279
       _cons |   1.949125   1.045932     1.86   0.062    -.1008649    3.999114
-------------+----------------------------------------------------------------
6            |
      income |  -.0004237    .000277    -1.53   0.126    -.0009666    .0001193
         age |    .027311   .0135744     2.01   0.044     .0007058    .0539163
  yearsofedu |  -.0176924   .0271537    -0.65   0.515    -.0709126    .0355278
       _cons |    1.90578   .6293261     3.03   0.002     .6723233    3.139236
------------------------------------------------------------------------------

. est store all

. asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 5, case(id)
alternatives(area) casevars(income age yearsofedu)
note: 289 cases (1445 obs) dropped due to no positive outcome per case

Iteration 0:   log likelihood = -527.11968
Iteration 1:   log likelihood = -508.09167
Iteration 2:   log likelihood = -503.92136
Iteration 3:   log likelihood = -503.87837
Iteration 4:   log likelihood = -503.87835

Alternative-specific conditional logit         Number of obs      =      10310
Case variable: id                              Number of cases    =       2062

Alternative variable: area                     Alts per case: min =          5
                                                              avg =        5.0
```

```
                                                       max =          5

                                                  Wald chi2(17)   =     678.02
Log likelihood = -503.87835                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0833449   .0034863   -23.91   0.000     -.090178   -.0765118
  hourswater |   .8348735   .0731734    11.41   0.000     .6914563    .9782908
 hourswater2 |  -.0324699   .0039088    -8.31   0.000    -.0401311   -.0248088
hourselect~y |   1.421423   .0942068    15.09   0.000     1.236781    1.606065
hourselect~2 |  -.0742321   .0059796   -12.41   0.000    -.0859518   -.0625123
-------------+----------------------------------------------------------------
1            |  (base alternative)
-------------+----------------------------------------------------------------
2            |
      income |   .0003952   .0002624     1.51   0.132     -.000119    .0009095
         age |   .0469977   .0176176     2.67   0.008      .012468    .0815275
  yearsofedu |   .1460012   .0422599     3.45   0.001     .0631732    .2288291
       _cons |  -2.243646   .8417937    -2.67   0.008    -3.893531   -.5937606
-------------+----------------------------------------------------------------
3            |
      income |  -.0010999   .0007469    -1.47   0.141    -.0025637    .0003639
         age |   .0216349   .0169812     1.27   0.203    -.0116477    .0549176
  yearsofedu |  -.0969138   .0327597    -2.96   0.003    -.1611217   -.0327059
       _cons |   1.466938   .7583958     1.93   0.053    -.0194899    2.953367
-------------+----------------------------------------------------------------
4            |
      income |   -.000343   .0005252    -0.65   0.514    -.0013723    .0006864
         age |   .0552969   .0282004     1.96   0.050     .0000251    .1105687
  yearsofedu |    .137339   .0680399     2.02   0.044     .0039834    .2706947
       _cons |  -6.769175   1.451079    -4.66   0.000    -9.613237   -3.925113
-------------+----------------------------------------------------------------
6            |
      income |  -.0003989   .0002739    -1.46   0.145    -.0009358    .0001379
         age |    .028359   .0132118     2.15   0.032     .0024643    .0542537
  yearsofedu |  -.0195917   .0264982    -0.74   0.460    -.0715272    .0323438
       _cons |   1.613087   .6067979     2.66   0.008      .423785    2.802389
------------------------------------------------------------------------------

. est store partial

. hausman partial all, alleqs constant

Note: the rank of the differenced variance matrix (19) does not equal the number of coefficients being
tested (21); be sure this is what you expect, or there may be problems computing the test.  Examine the
output of your estimators for anything unexpected and possibly consider scaling your variables so that the
coefficients are on a similar scale.

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    partial        all         Difference          S.E.
-------------+----------------------------------------------------------------
area         |
        rent |   -.0833449    -.0884401        .0050952        .0004952
  hourswater |    .8348735      .971236       -.1363625               .
 hourswater2 |   -.0324699    -.0391148        .0066449               .
hourselect~y |    1.421423     1.481254       -.0598311        .0192093
hourselect~2 |   -.0742321    -.0776209        .0033888        .0008752
-------------+----------------------------------------------------------------
2            |
      income |    .0003952     .0003719        .0000233        .0000514
         age |    .0469977     .0497745       -.0027767               .
  yearsofedu |    .1460012     .1517774       -.0057762               .
       _cons |   -2.243646    -2.242399       -.0012472               .
-------------+----------------------------------------------------------------
3            |
      income |   -.0010999    -.0010744       -.0000255               .
         age |    .0216349     .0194763        .0021586               .
  yearsofedu |   -.0969138    -.0967144       -.0001994               .
       _cons |    1.466938     1.827903       -.3609643               .
-------------+----------------------------------------------------------------
4            |
      income |    -.000343    -.0003516        8.68e-06        .0000684
         age |    .0552969     .0568159        -.001519               .
  yearsofedu |     .137339      .146708        -.009369               .
       _cons |   -6.769175    -6.942976         .173801               .
-------------+----------------------------------------------------------------
6            |
      income |   -.0003989    -.0004237        .0000248               .
         age |     .028359      .027311         .001048               .
  yearsofedu |   -.0195917    -.0176924       -.0018993               .
       _cons |    1.613087      1.90578       -.2926929               .
------------------------------------------------------------------------------
                        b = consistent under Ho and Ha; obtained from asclogit
         B = inconsistent under Ha, efficient under Ho; obtained from asclogit
```

```
    Test:  Ho:  difference in coefficients not systematic

                 chi2(19) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      311.33
                Prob>chi2 =      0.0000
                (V_b-V_B is not positive definite)

 asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2, case(id)
alternatives(area) casevars(income age yearsofedu)

Iteration 0:   log likelihood = -588.85292
Iteration 1:   log likelihood =  -567.6415
Iteration 2:   log likelihood = -560.27514
Iteration 3:   log likelihood = -560.16378
Iteration 4:   log likelihood = -560.16368
Iteration 5:   log likelihood = -560.16368

Alternative-specific conditional logit         Number of obs      =      14100
Case variable: id                              Number of cases    =       2350

Alternative variable: area                     Alts per case: min =          6
                                                              avg =        6.0
                                                              max =          6

                                                  Wald chi2(20)   =     799.13
Log likelihood = -560.16368                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0884401    .003451   -25.63   0.000    -.0952039   -.0816763
  hourswater |    .971236   .0755669    12.85   0.000     .8231276    1.119345
 hourswater2 |  -.0391148     .00402    -9.73   0.000    -.0469939   -.0312357
hourselect~y |   1.481254   .0922275    16.06   0.000     1.300491    1.662016
hourselect~2 |  -.0776209   .0059152   -13.12   0.000    -.0892145   -.0660273
-------------+----------------------------------------------------------------
1            |  (base alternative)
-------------+----------------------------------------------------------------
2            |
      income |   .0003719   .0002573     1.45   0.148    -.0001323    .0008762
         age |   .0497745   .0177274     2.81   0.005     .0150294    .0845195
  yearsofedu |   .1517774   .0422894     3.59   0.000     .0688916    .2346631
       _cons |  -2.242399   .8603667    -2.61   0.009    -3.928686   -.5561109
-------------+----------------------------------------------------------------
3            |
      income |  -.0010744   .0007707    -1.39   0.163    -.0025849    .0004361
         age |   .0194763     .01788     1.09   0.276    -.0155678    .0545205
  yearsofedu |  -.0967144    .034396    -2.81   0.005    -.1641293   -.0292995
       _cons |   1.827903   .8049947     2.27   0.023     .2501421    3.405663
-------------+----------------------------------------------------------------
4            |
      income |  -.0003516   .0005207    -0.68   0.499    -.0013722     .000669
         age |   .0568159   .0284399     2.00   0.046     .0010747    .1125572
  yearsofedu |    .146708   .0704001     2.08   0.037     .0087263    .2846897
       _cons |  -6.942976   1.485985    -4.67   0.000    -9.855453   -4.030499
-------------+----------------------------------------------------------------
5            |
      income |  -.0005556   .0004731    -1.17   0.240    -.0014828    .0003715
         age |  -.0026325   .0235372    -0.11   0.911    -.0487645    .0434995
  yearsofedu |  -.0146094    .045428    -0.32   0.748    -.1036467    .0744279
       _cons |   1.949125   1.045932     1.86   0.062    -.1008649    3.999114
-------------+----------------------------------------------------------------
6            |
      income |  -.0004237    .000277    -1.53   0.126    -.0009666    .0001193
         age |    .027311   .0135744     2.01   0.044     .0007058    .0539163
  yearsofedu |  -.0176924   .0271537    -0.65   0.515    -.0709126    .0355278
       _cons |    1.90578   .6293261     3.03   0.002     .6723233    3.139236
------------------------------------------------------------------------------

. est store all

. asclogit chosenarea rent hourswater hourswater2 hourselectricity hourselectricity2 if area != 6, case(id)
alternatives(area) casevars(income age yearsofedu)
note: 814 cases (4070 obs) dropped due to no positive outcome per case

Iteration 0:   log likelihood = -377.76929
Iteration 1:   log likelihood =  -362.1109
Iteration 2:   log likelihood = -356.91977
Iteration 3:   log likelihood = -356.87581

Alternative-specific conditional logit         Number of obs      =       7700
Case variable: id                              Number of cases    =       1540

Alternative variable: area                     Alts per case: min =          5
                                                              avg =        5.0
                                                              max =          5
```

```
                                                  Wald chi2(17)   =     503.62
Log likelihood = -356.87579                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
  chosenarea |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
area         |
        rent |  -.0819382   .0040412   -20.28   0.000    -.0898587   -.0740176
  hourswater |   .6711633   .0799254     8.40   0.000     .5145124    .8278141
 hourswater2 |  -.0242302   .0042596    -5.69   0.000    -.0325789   -.0158815
hourselect~y |   1.485716   .1147545    12.95   0.000     1.260802    1.710631
hourselect~2 |  -.0763111   .0070727   -10.79   0.000    -.0901734   -.0624489
-------------+----------------------------------------------------------------
1            |  (base alternative)
-------------+----------------------------------------------------------------
2            |
      income |   .0004529   .0002886     1.57   0.117    -.0001128    .0010185
         age |   .0572359   .0190029     3.01   0.003     .0199908     .094481
  yearsofedu |   .1569126   .0436334     3.60   0.000     .0713926    .2424325
       _cons |  -2.927603   .9202524    -3.18   0.001    -4.731264   -1.123941
-------------+----------------------------------------------------------------
3            |
      income |  -.0011315   .0007475    -1.51   0.130    -.0025966    .0003335
         age |   .0198018   .0171905     1.15   0.249     -.013891    .0534946
  yearsofedu |  -.0987004   .0331024    -2.98   0.003    -.1635799    -.033821
       _cons |   1.171528   .7807741     1.50   0.133    -.3587607    2.701817
-------------+----------------------------------------------------------------
4            |
      income |  -.0003341   .0005383    -0.62   0.535    -.0013891    .0007209
         age |   .0579198   .0290398     1.99   0.046     .0010028    .1148368
  yearsofedu |    .146097   .0712655     2.05   0.040     .0064191    .2857749
       _cons |  -7.064477   1.540794    -4.58   0.000    -10.08438   -4.044577
-------------+----------------------------------------------------------------
5            |
      income |  -.0005149   .0004669    -1.10   0.270      -.00143    .0004001
         age |   .0002313    .022666     0.01   0.992    -.0441933    .0446559
  yearsofedu |  -.0177053   .0439075    -0.40   0.687    -.1037625    .0683519
       _cons |   1.123501   1.015652     1.11   0.269    -.8671407    3.114143
------------------------------------------------------------------------------

. est store partial

. hausman partial all, alleqs constant

Note: the rank of the differenced variance matrix (20) does not equal the number of coefficients being
tested (21); be sure this is what you expect, or there may be problems computing the test.  Examine the
output of your estimators for anything unexpected and possibly consider scaling your variables so that the
coefficients are on a similar scale.

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     partial       all         Difference          S.E.
-------------+----------------------------------------------------------------
area         |
        rent |   -.0819382    -.0884401        .0065019        .0021028
  hourswater |    .6711633      .971236       -.3000728        .0260327
 hourswater2 |   -.0242302    -.0391148        .0148846        .0014084
hourselect~y |    1.485716     1.481254        .0044624        .0682837
hourselect~2 |   -.0763111    -.0776209        .0013097        .0038773
-------------+----------------------------------------------------------------
2            |
      income |    .0004529     .0003719        .0000809        .0001308
         age |    .0572359     .0497745        .0074615        .0068449
  yearsofedu |    .1569126     .1517774        .0051352        .0107463
       _cons |   -2.927603    -2.242399        -.685204        .3265481
-------------+----------------------------------------------------------------
3            |
      income |   -.0011315    -.0010744       -.0000571               .
         age |    .0198018     .0194763        .0003255               .
  yearsofedu |   -.0987004    -.0967144       -.0019861               .
       _cons |    1.171528     1.827903       -.6563744               .
-------------+----------------------------------------------------------------
4            |
      income |   -.0003341    -.0003516        .0000175        .0001364
         age |    .0579198     .0568159        .0011039        .0058722
  yearsofedu |     .146097      .146708        -.000611        .0110725
       _cons |   -7.064477    -6.942976       -.1215012        .4073011
-------------+----------------------------------------------------------------
5            |
      income |   -.0005149    -.0005556        .0000407               .
         age |    .0002313    -.0026325        .0028638               .
  yearsofedu |   -.0177053    -.0146094       -.0030959               .
       _cons |    1.123501     1.949125       -.8256233               .
------------------------------------------------------------------------------
                        b = consistent under Ho and Ha; obtained from asclogit
         B = inconsistent under Ha, efficient under Ho; obtained from asclogit
```

```
Test:  Ho:  difference in coefficients not systematic

             chi2(20) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =      151.32
            Prob>chi2 =      0.0000
            (V_b-V_B is not positive definite)
```

Appendix III: IIA Metaanalysis Results - Detailed Results: Computer Output

Random-effects probit regression

Probit Results for IIA Accept and Reject

All Variables

```
xtprobit iiaresult published jimfact logn logpercapita numofalternatives model2 model3 model4
model5 model6 smallhsiaotest lowincome lowmiddleincome uppermiddleincome studytransportmode
studyresidentiallocation studyfirmlocation studyenvironvaluation studybrandchoicefinance
studyhealthcareandmedicare studytvtelephoneinternet studyemploymentschoolslabour
studyschoolchoice studyvotingpoliticaldecisions

note: studyschoolchoice omitted because of collinearity

Fitting comparison model:

Iteration 0:   log likelihood = -181.53228
Iteration 1:   log likelihood = -115.19891
Iteration 2:   log likelihood = -113.19771
Iteration 3:   log likelihood =  -113.1121
Iteration 4:   log likelihood = -113.10157
Iteration 5:   log likelihood = -113.09955
Iteration 6:   log likelihood = -113.09929
Iteration 7:   log likelihood = -113.09924
Iteration 8:   log likelihood = -113.09923

Fitting full model:
rho =  0.0     log likelihood = -113.09923
rho =  0.1     log likelihood = -113.23119
Iteration 0:   log likelihood = -113.23119
Iteration 1:   log likelihood = -112.96419
Iteration 2:   log likelihood = -112.95935
Iteration 3:   log likelihood = -112.95499
Iteration 4:   log likelihood = -112.95494
Iteration 5:   log likelihood = -112.95494

Random-effects probit regression                Number of obs      =       293
Group variable: studyno                         Number of groups   =       137

Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       2.1
                                                               max =         8

                                                Wald chi2(23)      =     41.80
Log likelihood  = -112.95494                    Prob > chi2        =    0.0096

------------------------------------------------------------------------------
   iiaresult |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   published |   .6339489   .3166066     2.00   0.045     .0134113    1.254486
     jimfact |  -.0139347   .0794476    -0.18   0.861    -.1696491    .1417797
        logn |  -.0630962   .0694322    -0.91   0.363    -.1991808    .0729885
logpercapita |   1.067255   .5361885     1.99   0.047      .016345    2.118165
numofalter~s |   .1556883   .0904771     1.72   0.085    -.0216435    .3330201
      model2 |   1.544285   .6735509     2.29   0.022     .2241494    2.864421
      model3 |   2.133632   .4924357     4.33   0.000     1.168476    3.098788
      model4 |   1.976827   .5167738     3.83   0.000     .9639686    2.989685
      model5 |   .7204579   .3584233     2.01   0.044     .0179611    1.422955
      model6 |     1.0178   .5545914     1.84   0.066    -.0691792    2.104779
smallhsiao~t |   .9452908   .3737268     2.53   0.011     .2127996    1.677782
   lowincome |   3.325634   1.702814     1.95   0.051    -.0118197    6.663088
lowmiddlei~e |   2.204266   1.160952     1.90   0.058    -.0711579     4.47969
uppermiddl~e |   .8792662   .8613345     1.02   0.307    -.8089183    2.567451
studytrans~e |  -3.955173   2599.015    -0.00   0.999     -5097.93     5090.02
studyresid~n |      1.055   .6777913     1.56   0.120    -.2734465    2.383446
studyfirml~n |   .0620813   .6417067     0.10   0.923    -1.195641    1.319803
studyenvir~n |   .8724738   .4198357     2.08   0.038     .0496109    1.695337
studybrand~e |   .4804021   .4981225     0.96   0.335    -.4959001    1.456704
```

```
studyhealt~e |   1.091312   .4841316     2.25   0.024     .1424313    2.040192
studytvtel~t |   2.616409   .7466851     3.50   0.000     1.152933    4.079885
studyemplo~r |   .7894569   .4674083     1.69   0.091    -.1266466     1.70556
studyschoo~e |  (omitted)
studyvotin~s |   -5.43582   1546.363    -0.00   0.997    -3036.251    3025.379
       _cons |  -14.01356   5.732652    -2.44   0.015    -25.24935   -2.777771
-------------+----------------------------------------------------------------
    /lnsig2u |  -1.894275   2.148991                     -6.106219     2.31767
-------------+----------------------------------------------------------------
     sigma_u |   .3878497   .4167428                      .0472119     3.18622
         rho |   .1307579   .2442548                       .002224      .91033
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =     0.29 Prob >= chibar2 = 0.296
```

Selected Model - Whole Dataset

```
xtprobit iiaresult logn jimfact studytvtelephoneinternet studyresidentiallocation
hmcfaddentest studyenvironvaluation HMcFadden-sample alternatives

Fitting comparison model:

Iteration 0:   log likelihood = -192.05042
Iteration 1:   log likelihood =  -163.2121
Iteration 2:   log likelihood =  -163.0122
Iteration 3:   log likelihood = -163.01152
Iteration 4:   log likelihood = -163.01152

Fitting full model:

rho =  0.0     log likelihood = -163.01152
rho =  0.1     log likelihood = -157.30353
rho =  0.2     log likelihood = -153.91629
rho =  0.3     log likelihood = -151.92472
rho =  0.4     log likelihood = -150.97127
rho =  0.5     log likelihood = -150.93531
rho =  0.6     log likelihood = -151.73805

Iteration 0:   log likelihood =  -150.9157
Iteration 1:   log likelihood = -146.22952
Iteration 2:   log likelihood = -145.29957
Iteration 3:   log likelihood = -145.21637
Iteration 4:   log likelihood = -145.21559
Iteration 5:   log likelihood = -145.21559  (backed up)

Random-effects probit regression                Number of obs      =       323
Group variable: studyno                         Number of groups   =       146

Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       2.2
                                                               max =        13

                                                Wald chi2(8)       =     15.88
Log likelihood  = -145.21559                    Prob > chi2        =    0.0441

------------------------------------------------------------------------------
   iiaresult |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        logn |   .4900608   .2564847     1.91   0.056    -.0126399    .9927616
     jimfact |   .0712214   .0964818     0.74   0.460    -.1178795    .2603223
studytvtel~t |   3.561913    1.24112     2.87   0.004     1.129363    5.994463
studyresid~n |   .5714214   1.007316     0.57   0.571    -1.402882    2.545725
hmcfaddent~t |    4.44683   2.292085     1.94   0.052    -.0455734    8.939233
studyenvir~n |   .9692523   .4583333     2.11   0.034     .0709356    1.867569
   HM sample |   -.675113   .2865076    -2.36   0.018    -1.236658   -.1135684
alternatives |   .2536848   .1204021     2.11   0.035      .017701    .4896687
       _cons |  -5.975715   2.306351    -2.59   0.010    -10.49608   -1.455349
-------------+----------------------------------------------------------------
    /lnsig2u |   .8542841   .5235965                     -.1719462    1.880514
-------------+----------------------------------------------------------------
     sigma_u |    1.53287   .4013028                      .9176189     2.56064
         rho |   .7014651   .1096473                       .457119    .8676702
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =    35.59 Prob >= chibar2 = 0.000
```

Selected Model - Hausman-McFadden test Sub-Sample

```
xtprobit iiaresult logn jimfact studytvtelephoneinternet published studyresidentiallocation
studyenvironvaluation  alternatives if hmcfaddentest==1

Fitting comparison model:

Iteration 0:   log likelihood = -160.49732
Iteration 1:   log likelihood = -133.32965
Iteration 2:   log likelihood = -132.91114
Iteration 3:   log likelihood =  -132.9102
Iteration 4:   log likelihood =  -132.9102

Fitting full model:

rho =  0.0     log likelihood =  -132.9102
rho =  0.1     log likelihood = -128.23631
rho =  0.2     log likelihood =  -125.5279
rho =  0.3     log likelihood = -124.00131
rho =  0.4     log likelihood = -123.34662
rho =  0.5     log likelihood = -123.47362

Iteration 0:   log likelihood = -123.35128
Iteration 1:   log likelihood = -117.79581
Iteration 2:   log likelihood = -117.51036
Iteration 3:   log likelihood = -117.50567
Iteration 4:   log likelihood = -117.50567
Iteration 5:   log likelihood = -117.50564
Iteration 6:   log likelihood = -117.50564

Random-effects probit regression                Number of obs      =       273
Group variable: studyno                         Number of groups   =       135

Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       2.0
                                                               max =        13

                                                Wald chi2(7)       =     12.67
Log likelihood  = -117.50564                    Prob > chi2        =    0.0804

------------------------------------------------------------------------------
   iiaresult |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        logn |  -.1904575   .1250175    -1.52   0.128    -.4354873    .0545723
     jimfact |   .0375416   .1659715     0.23   0.821    -.2877566    .3628398
studytvtel~t |   3.967315   1.424482     2.79   0.005     1.175381    6.759248
   published |   1.006381   .5984758     1.68   0.093      -.16661    2.179372
studyresid~n |    .276599   1.080118     0.26   0.798    -1.840394    2.393592
studyenvir~n |   1.539061   .6290115     2.45   0.014     .3062211    2.771901
alternatives |   .3624097   .1534612     2.36   0.018     .0616313    .6631882
       _cons |  -2.567566   1.270781    -2.02   0.043    -5.058251   -.0768816
-------------+----------------------------------------------------------------
    /lnsig2u |   1.036003   .5778076                     -.0964786    2.168485
-------------+----------------------------------------------------------------
     sigma_u |    1.67867    .484974                      .9529057    2.957199
         rho |   .7380781   .1117011                       .475899    .8973836
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =    30.81 Prob >= chibar2 = 0.000
```

Ordered Probit Model – Ranked P-Values as Dependent variable

```
oprobit opvalue  published jimfact logn logpercapita numofalternatives model2 model3 model4
model5 model6 smallhsiaotest lowincome lowmiddleincome uppermiddleincome studytransportmode
studyresidentiallocation studyfirmlocation studyenvironvaluation studybrandchoicefinance
studyhealthcareandmedicare studytvtelephoneinternet studyemploymentschoolslabour
studyschoolchoice studyvotingpoliticaldecisions

note: studytransportmode omitted because of collinearity
note: studyschoolchoice omitted because of collinearity
Iteration 0:   log likelihood = -373.78823
Iteration 1:   log likelihood = -317.05086
Iteration 2:   log likelihood = -316.77751
Iteration 3:   log likelihood = -316.77718
Iteration 4:   log likelihood = -316.77718

Ordered probit regression                         Number of obs   =        197
                                                  LR chi2(22)     =     114.02
                                                  Prob > chi2     =     0.0000
Log likelihood = -316.77718                       Pseudo R2       =     0.1525

------------------------------------------------------------------------------
     opvalue |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   published |   .0712659   .2568673     0.28   0.781    -.4321847    .5747165
     jimfact |   .1320997   .0770317     1.71   0.086    -.0188796     .283079
        logn |   .1401576    .059514     2.36   0.019     .0235123    .2568029
logpercapita |  -.9851696    .432351    -2.28   0.023    -1.832562   -.1377771
numofalter~s |  -.0337672   .0889831    -0.38   0.704    -.2081707    .1406364
      model2 |  -1.806324   .6358237    -2.84   0.004    -3.052515    -.560132
      model3 |  -1.938821    .293512    -6.61   0.000    -2.514094   -1.363548
      model4 |  -1.146848   .3433191    -3.34   0.001    -1.819741   -.4739549
      model5 |  -.3002014   .2922167    -1.03   0.304    -.8729357    .2725328
      model6 |  -.2628657   .3437522    -0.76   0.444    -.9366077    .4108762
smallhsiao~t |  -1.258439   .2449136    -5.14   0.000     -1.73846   -.7784168
   lowincome |   -2.35013   1.473893    -1.59   0.111    -5.238907    .5386474
lowmiddlei~e |  -1.403323    .941038    -1.49   0.136    -3.247723    .4410779
uppermiddl~e |  -.2793315   .6874288    -0.41   0.684    -1.626667    1.068004
studytrans~e |  (omitted)
studyresid~n |  -1.185522   .6751235    -1.76   0.079     -2.50874    .1376955
studyfirml~n |   .3576188   .5502916     0.65   0.516    -.7209329     1.43617
studyenvir~n |  -.9598377   .2952453    -3.25   0.001    -1.538508   -.3811676
studybrand~e |  -.0735912   .4002108    -0.18   0.854    -.8579901    .7108076
studyhealt~e |  -1.214682   .3983069    -3.05   0.002    -1.995349   -.4340148
studytvtel~t |  -.7265131   .6223089    -1.17   0.243    -1.946216      .49319
studyemplo~r |  -1.338611    .355424    -3.77   0.000     -2.03523   -.6419932
studyschoo~e |  (omitted)
studyvotin~s |  -.7017093   .6128193    -1.15   0.252    -1.902813    .4993945
-------------+----------------------------------------------------------------
       /cut1 |  -11.36274   4.530175                     -20.24172   -2.483762
       /cut2 |  -11.07521   4.525942                     -19.94589   -2.204524
       /cut3 |  -10.87165   4.523371                      -19.7373    -2.00601
       /cut4 |  -10.73663   4.521724                     -19.59905   -1.874217
       /cut5 |  -10.61686   4.520019                     -19.47593   -1.757781
       /cut6 |  -10.53697     4.5191                     -19.39425   -1.679699
       /cut7 |  -10.45527   4.518214                      -19.3108   -1.599731
       /cut8 |    -10.353   4.517516                     -19.20717    -1.49883
       /cut9 |  -10.07957   4.516664                     -18.93207   -1.227074
      /cut10 |  -9.877051   4.516623                     -18.72947   -1.024633
------------------------------------------------------------------------------
```

Probit Results for Publication Published and Unpublished*

```
probit published alt3 logn waitingpub authors modelmprobit modelnlogit modelmixedlogit
modelclogit hmcfaddentest studyresidentiallocation studyfirmlocation studyenvironvaluation
studybrandchoicefinance studytvtelephoneinternet studyvotingpoliticaldecisions
studyemploymentschoolslabour

Iteration 0:   log likelihood = -212.05349
Iteration 3:   log likelihood = -173.52575
Iteration 4:   log likelihood = -173.52574

Probit regression                                 Number of obs   =        327
                                                  LR chi2(16)     =      77.06
                                                  Prob > chi2     =     0.0000
Log likelihood = -173.52574                       Pseudo R2       =     0.1817

------------------------------------------------------------------------------
   published |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        alt3 |   .5964022   .1713777     3.48   0.001      .260508    .9322964
        logn |   .1065257   .0463771     2.30   0.022     .0156281    .1974232
  waitingpub |   .6972836   .1715164     4.07   0.000     .3611177     1.03345
     authors |   .4123271   .1836989     2.24   0.025     .0522838    .7723705
modelmprobit |    1.78358   .4921619     3.62   0.000     .8189604      2.7482
 modelnlogit |    .104703   .2519169     0.42   0.678     -.389045     .598451
modelmixed~t |  -.0559629   .3194874    -0.18   0.861    -.6821467    .5702209
 modelclogit |  -.6620151   .2764401    -2.39   0.017    -1.203828   -.1202025
hmcfaddent~t |    .782713   .2601809     3.01   0.003     .2727678    1.292658
studyresid~n |   .8234347   .5135229     1.60   0.109    -.1830517    1.829921
studyfirml~n |   .9443974    .376804     2.51   0.012      .205875     1.68292
studyenvir~n |   -.520938   .2450668    -2.13   0.034     -1.00126    -.040616
studybrand~e |  -.1521566   .3387606    -0.45   0.653    -.8161151    .5118019
studytvtel~t |   .1108215   .4036285     0.27   0.784    -.6802757    .9019188
studyvotin~s |  -.4235154   .4347313    -0.97   0.330    -1.275573    .4285423
studyemplo~r |  -.2847617    .248039    -1.15   0.251    -.7709093    .2013858
       _cons |  -2.510279   .4881376    -5.14   0.000    -3.467011   -1.553547
------------------------------------------------------------------------------
```

Appendix IV: Complete list of studies included in the IIA Meta-Analysis

S/No.	*Author(s)*	*Year*	*Journal/Unpublished**	*Research Topic/Title*	*Sample*	*Reported IIA Result*	*IIA Test Type*	*Number of Alternatives*
1	Adams and Goldsmith	1999	International Food and Agribusiness Management Review	Strategic Alliances in the Food Industry	49	Accepted	Other/NS	3
2	Akin et al	1995	Social Science and Medicine	Quality of Services and Demand for Health Care in Nigeria: A Multinomial Probit Estimation	1763	Rejected	Other/NS	3
3	Alba-Ramírez et al	2007	Labour Economics	Exits from unemployment - Recall or new job	23035	Accepted	HM	3
4	Alix-Garcia	2008	Journal of Development Economics	Effect of inequality on common property forests	346	Accepted	HM	4
5	Aliyu	2009	Unpublished	Microeconometrics of residential location decisions in Nigeria	2439	Accepted	HM	6
6	Alvarez	2007	World Development	Firm Characteristics and Spillover Effects	2592	Accepted	HM	3
7	Álvarez-Farizo et al	2007	Ecological Economics	Individual versus collective interest in environmental valuation	576	Accepted	HM	3
8	An et al	2008	Journal of Real Estate Finance and Economics	Omitted Mobility Characteristics and Property Market Dynamics: Application to Mortgage Termination	1985	Accepted	HM	3
9	Andersen and Christensen	2005		Short-Term Choice Behaviour of Danish Fishermen	117	Rejected	HM	16
10	Andersson et al	2004	Review of Industrial Organization	Demand for basic broadband and premium broadband	1061	Accepted	HM	3
11	Angulo et al	2000	British Food Journal	Hedonic prices for Spanish red quality wine	222	Accepted	HM	3
12	Anyadike-Danes and McVicar	2005	Labour Economics	Childhood influences and male career path clusters	3367	Accepted	HM	6

S/No.	*Author(s)*	*Year*	*Journal/Unpublished**	*Research Topic/Title*	*Sample*	*Reported IIA Result*	*IIA Test Type*	*Number of Alternatives*
13	Asfaw	2006	World Development	Government Food Price Policies and Prevalence of Obesity	902	Accepted	HM and SH	4
14	Asfaw et al	2008	Unpublished	Intra-household Gender Disparities in Children's Medical Care before Death in India	907	Accepted	HM and SH	3
15	Asif	2007	Unpublished	Factors Affecting Employment Choices in Rural Northwest Pakistan	2825	Accepted	HM	6
16	Aslam and Kingdon	2009	Journal of Asian Economics	Public–private sector segmentation in the Pakistani labour market	10884	Accepted	SH	3
17	Atherly	2002	International Journal of Health Care Finance and Economics	The Effect of Medicare Supplemental Insurance on Medicare Expenditures	10853	Accepted	HM	3
18	Audretsch et al	2007	Unpublished	Entrepreneurs Innovation and Financing Constraints	906	Accepted	HM	4
19	Avalos and Hoyos	2008	Review of Industrial Organisation	An Empirical Analysis of Mexican Merger Policy	239	Ambiguous	HM and SH	3
20	Aw and Lee	2008	Journal of International Economics	Firm heterogeneity and location choice of Taiwanese multinationals	884	Accepted	HM	3
21	Bäck	2001	Unpublished	Coalition Formation	8399	Accepted	HM	3
22	Bäck and Dumont	2008	Public Choice	Making the first move - A two-stage analysis of the role of formateurs in parliamentary government formation	1373	Accepted	HM	9
23	Badgett et al	2008	Review of Economics of the Household	Domestic partnerships among gay men and lesbians	1002	Rejected	Other/NS	3
24	Banfi et al	2009	Annals of Public and Cooperative Economics	Child Care Demand in Switzerland	597	Accepted	HM	3
25	Bargain et al	2008	Journal of Population Economics	Making work pay in a rationed labor market	7159	Accepted	Other/NS	6

S/No.	Author(s)	Year	Journal/Unpublished*	Research Topic/Title	Sample	Reported IIA Result	IIA Test Type	Number of Alternatives
26	Battisti et al	2009	Research Policy	e-Business usage across and within firms in the UK: profitability, externalities and policy	5822	Accepted	HM	3
27	Bauer and Riphahn	2008	Unpublished	Age at school entry and intergenerational educational mobility	62535	Accepted	HM	3
28	Belderbos and Carree	2000	Unpublished	The Location of Japanese Investments in China: Agglomeration Effects, Keiretsu, and Firm Heterogeneity	229	Accepted	HM	29
29	Benjamin and Kimhi	2006	European Review of Agricultural Economics	French farm couples' labour decisions	65593	Accepted	HM	16
30	Birol et al	2006b	Ecological Economics	Using a choice experiment to estimate the non-use values of wetlands	2935	Rejected	HM	3
31	Birol et al	2006	Environmental and Resource Economics	Farmers' Valuation of Agrobiodiversity on Hungarian Small Farms	4440	Accepted	HM	3
32	Bondy et al	2009	Vaccine	Identifying the determinants of childhood immunization in the Philippines	1158	Ambiguous	HM and SH	3
33	Boyle and Özdemir	2008	Environmental and Resource Economics	Convergent Validity of Attribute-Based, Choice Questions in Stated-Preference Studies	830	Accepted	HM	4
34	Burton et al	2007	Agricultural Systems	Community attitudes towards water management	1917	Accepted	HM	6
35	Bussiere and Fratzscher	2006	Journal of International Money and Finance	Towards a new early warning system of financial crises	1549	Accepted	HM	3
36	Butler	1999	Unpublished	Estimating Elasticities of Demand for Private Health Insurance in Australia	9199	Accepted	HM	4
37	Campos and Dabusinskas	2008	European Journal of Political Economy	So many rocket scientists, so few marketing clerks: Estimating the effects of economic reform on occupational mobility in Estonia	5848	Accepted	HM and SH	3

S/No.	*Author(s)*	*Year*	*Journal/Unpublished**	*Research Topic/Title*	*Sample*	*Reported IIA Result*	*IIA Test Type*	*Number of Alternatives*
38	Cardona et al	2008	Journal of Regulatory Economics	Demand estimation and market definition for broadband internet services	2825	Rejected	Other/NS	5
39	Cheng	2008	Annals of Regional Science	How can western China attract FDI? A case of Japanese investment	3893	Accepted	HM	27
40	Cheng and Stough	2006	Annals of Regional Science	Location decisions of Japanese new manufacturing plants in China: a discrete-choice analysis	764	Accepted	HM	24
41	Ching	1995	Social Science and Medicine	User Fees, Demand For Children's Health Care and Access across Income Groups: The Philippines Case	520	Accepted	HM	3
42	Choo and Mokhtarian	2002		What Type of Vehicle Do People Drive	1904	Accepted	Other/NS	9
43	Choo et al	2007	Transportation Research Part D	The development of a prescreening model to identify failed and gross polluting vehicles	365488	Accepted	HM	3
44	Christie et al	2007	Journal of Forest Economics	Valuing enhancements to forest recreation using choice experiment and contingent behaviour methods - Cycling	566	Rejected	SH	4
45	Clough	2007	Electoral Studies	Two political worlds?: The influence of provincial party loyalty federal voting in Canada		Accepted	HM	
46	Cohen-Zada and Sander	2008	Journal of Urban Economics	Religion and school choice	2447	Accepted	HM	4
47	Colak	2007	Unpublished	Diversification, Refocusing, and Firm Value	6233	Rejected	HM	3
48	Colaresi and Thompson		Unpublished	Initiation and Escalation of Spatial and Positional Rivalries	32000	Accepted	HM	5
49	Colman and Christie	2006	Unpublished	An economic assessment of the amenity benefits associated with alternative coastal defence options	360	Rejected	HM	3
50	Colombier and Mascle	2008	Small Business Economics	Intergenerational correlation in self employment	47063	Accepted	HM	4

S/No.	Author(s)	Year	Journal/Unpublished*	Research Topic/Title	Sample	Reported IIA Result	IIA Test Type	Number of Alternatives
51	Colombo et al	2006	Ecological Economics	Analysing the social benefits of soil conservation measures	252	Accepted	HM	3
52	Cooper and O'Keefe	2005	Unpublished	The importance of credit transfer in the decision to undertake postcompulsory education: An exercise in experimental choice analysis	274	Accepted	HM	3
53	Cooper et al	2005	Unpublished	Preferences and Values for Urban Waste Water Services in Small Rural Communities in Northern Victoria.		Accepted	HM	7
54	Craig et al	2008	Journal of Clinical Epidemiology	Keep it simple: Ranking health states yields values similar to cardinal measurement approaches	4025	Accepted	HM	3
55	Cronqvist and Nilsson	2005	Journal of Financial Economics	The choice between rights offerings and private equity placements	296	Rejected	SH	4
56	Dahlberg and Eklof	2003	Unpublished	Relaxing the IIA Assumption in Locational Choice Models	1444	Accepted	Other/NS	26
57	Dancer and Fiebig	2004	Australian Economic Papers	Modelling students at risk	1054	Accepted	SH	4
58	Danis and Pennington-Cross	2008	Journal of Economics and Business	The delinquency of subprime mortgages	97852	Ambiguous	HM and SH	6
59	David and Van	2008	Unpublished	Equity Basis Selection in Allocation Environments	84	Accepted	HM	3
60	Davies et al	2001	Journal of Regional Science	A Conditional Logit Approach to U.S. State-to-State Migration		Accepted	HM	47
61	Di	2007	Environmental and Resource Economics	Pollution abatement cost savings and FDI inflows to polluting sectors in China	3208	Accepted	HM	4
62	Dimova and Gang	2007	Journal of Comparative Economics	Self-selection and wages during volatile transition	3112	Accepted	SH	4

S/No.	*Author(s)*	*Year*	*Journal/Unpublished**	*Research Topic/Title*	*Sample*	*Reported IIA Result*	*IIA Test Type*	*Number of Alternatives*
63	Dinkelman and Pirouz	2001	Unpublished	Unemployment and labour force participation in South Africa: A focus on the supply-side	54557	Accepted	HM	4
64	Dolton et al	2000	Unpublished	Survey Attrition: A taxonomy and the search for valid instruments to correct for biases	8925	Accepted	HM	4
65	Dong et al	2008	European Journal of Health Economics	The differences in characteristics between health-care users and non-users	988	Accepted	HM	4
66	Dostie and Leger	2006	Unpublished	Self-selection in migration and returns to unobservable skills	49046	Rejected	Other/NS	8
67	Duffy et al	2004	Unpublished	Health Services Utilization by Individuals with Substance Abuse and Mental Disorders	27646	Ambiguous	HM and SH	4
68	Eastburn and Morrison	2004	Unpublished	Brand Equity in the Australian Beef Market	2268	Rejected	HM	3
69	Engel and Heger	2005	Unpublished	Return-Orientation of Venture Capital Companies	37634	Accepted	HM	5
70	Fader and Hardie	1996	Journal of Marketing Research	Modeling Consumer Choice Among SKUs	3227	Accepted	SH	5
71	Ferto and Fogarasi	2005	Unpublished	The Choice of Farm Organisation	1394	Accepted	HM and SH	3
72	Foster and Mourato	1997	Unpublished	Behavioural Consistency, Statistical Specification and Validity in the Contingent Ranking Method: Evidence from a Survey on the Impacts of Pesticide Use in the U.K.	1683	Accepted	Other/NS	4
73	Foster and Mourato	2003	Environmental and Resource Economics	Elicitation Format and Sensitivity to Scope	234	Rejected	HM	3
74	Franck and Tavares	2008	Unpublished	Income and vote switching between local and national elections	3335	Accepted	HM	3

S/No.	*Author(s)*	*Year*	*Journal/Unpublished**	*Research Topic/Title*	*Sample*	*Reported IIA Result*	*IIA Test Type*	*Number of Alternatives*
75	Fuwa	2003	Unpublished	Pathways from Poverty toward Middle Class: Determinants of Socio-Economic Class Mobility in the Rural Philippines	1199	Rejected	HM	4
76	Gabriel and Rosenthal	1989	The Review of Economics and Statistics	Household Location and Race: Estimates of a Multinomial Logit Model	2497	Accepted	HM	5
77	Gaiha and Imai	2007	Unpublished	Non agricultural employment and poverty in India	41425	Accepted	HM	3
78	Giuri et al	2008	Information Economics and Policy	Explaining leadership in virtual teams: The case of open source software	77039	Rejected	Other/NS	5
79	Glasgowa and Alvarez	2005	Electoral Studies	Voting behavior and the electoral context of government formation	1063	Accepted	HM	4
80	Glenk	2008	Unpublished	Effects of attribute order in choice experiments	310	Rejected	HM	4
81	Goktepe and Mahagaonkar	2008	Unpublished	What do Scientists Want: Money or Fame	1074	Accepted	HM	3
82	Goldfarb	2001	Unpublished	Analyzing Website Choice	301206	Rejected	HM	8
83	Gooroochurn and Hanley	2007	Research Policy	A tale of two literatures: Transaction costs and property rights in innovation outsourcing	1724	Accepted	HM and SH	3
84	Grazier and Sloane	2008	Labour Economics	Accident risk, gender, family status and occupational choice in the UK	80782	Rejected	HM	25
85	Gresenz	1997	Unpublished	Role of AFDC Benefits in Location Choice	1022	Accepted	HM	48
86	Guris et al	2007	Quality and Quantity	The Brand Choice Model of Wine Consumers: A Multinomial Logit Model	1022	Accepted	HM	4
87	Hale	2002	Unpublished	Bonds, Loans and Country Risk	8682	Rejected	SH	3
88	Halvorsen	2000	Environmental and Resource Economics	Comparing Ranking and Contingent Valuation for Valuing Human Lives, Applying Nested and Non-Nested Logit Models	1002	Rejected	Other/NS	3

S/No.	*Author(s)*	*Year*	*Journal/Unpublished**	*Research Topic/Title*	*Sample*	*Reported IIA Result*	*IIA Test Type*	*Number of Alternatives*
89	Hanley et al 2002	2002	Environmental and Resource Economics	Modelling Recreation Demand Using Choice Experiments: Climbing in Scotland	267	Rejected	HM	8
90	Hanley et al 2006	2006	Journal of Environmental Management	Estimating the economic value of improvements in river ecology using choice experiments: an application to the water framework directive	420	Rejected	HM	3
91	Hensel et al	2008	Journal of Conflict Resolution	Bones of Contention: Comparing Territorial, Maritime, and River Issues	9940	Accepted	SH	4
92	Herrington	2001	Unpublished	Consumer Choice of Secondary Supermarkets	263	Accepted	Other/NS	5
93	Hope	2006	World Development	Evaluating Water Policy Scenarios Against the Priorities of the Rural Poor	320	Accepted	HM	3
94	Horbach	2008	Research Policy	Determinants of environmental innovation - New evidence from German panel data sources	1485	Accepted	HM	3
95	Ida and Kuroda	2004	Unpublished	Mobile Telephone Service Demand	939	Rejected	HM	6
96	Ida and Kuroda	2008	Journal of Regulatory Economics	Discrete Choice Analysis of Demand for Broadband in Japan	534	Rejected	HM	3
97	Ida and Kuroda	2008	Empirical Economics	Discrete choice model analysis of mobile telephone service demand in Japan	939	Rejected	HM	6
98	Ida and Sato	2006	The Kyoto Economic Review	Conjoint Analysis of Consumer Preferences for Broadband Services in Japan	1463	Rejected	HM	5
99	Ilahi and Grimard	2000	Economic Development and Cultural Change	Water Supply and Time Allocation of Women in Rural Pakistan	2400	Rejected	HM	3
100	Jo and Lee	2008	Unpublished	Agglomeration Economies, Technological Capability, and Firm Location Decision	352	Rejected	HM	16
101	Juon et al	2006	Journal of Quantitative Criminology	Childhood Behavior and Adult Criminality: Cluster Analysis in a Prospective Study of African Americans	572	Accepted	HM	3

S/No.	Author(s)	Year	Journal/Unpublished*	Research Topic/Title	Sample	Reported IIA Result	IIA Test Type	Number of Alternatives
102	Kahui and Alexander	2008	Environmental and Resource Economics	A Bioeconomic Analysis of Marine Reserves for Paua Management at Stewart Island, New Zealand	3388	Rejected	HM	4
103	Karp and Banducci	2001	Unpublished	Absentee Voting, Mobilization, and Participation	17437	Accepted	Other/NS	3
104	Kassie et al	2008	Unpublished	Adoption of Organic Farming Technologies: Evidence from Semi-Arid Regions of Ethiopia	348	Accepted	HM	4
105	Kato and Uctum	2004	Unpublished	Vanishing Intermediate Exchange Rate Regime - An Assessment of the Two-Pole Hypothesis	138	Accepted	SH	3
106	Kaya and Ulengin	2007	Unpublished	The Impact of Price Changes and Household Specific Attributes on the Choice Behavior of CSD Consumers	2039	Rejected	HM	3
107	Kim and Ulfarsson	2008	Transportation	Curbing automobile use for sustainable transportation: analysis of mode choice on short home-based trips	2737	Accepted	HM	4
108	Kim W.	2008	The Journal of Socio-Economics	Design of unemployment compensation	584	Accepted	HM	5
109	Kim Y.	2004	The Kyoto Economic Review	What Makes Family Members Live Apart or Together?: An Empirical Study with Japanese Panel Study of Consumers	1063	Accepted	HM	3
110	Kosenius	2008	Unpublished	Heterogeneous preferences for water quality attributes: benefit from the reduced nutrient load to the Gulf of Finland, Baltic Sea	3946	Accepted	HM	3
111	Kragt et al	2007	Unpublished	Comparing choice models of river health improvement	5190	Rejected	Other/NS	3
112	Kubis A	2007	Unpublished	Are there gender-specific preferences for location factors	26506	Accepted	HM	439
113	Ladenburg and Olsen	2006	Unpublished	Starting Point Anchoring Effects in Choice Experiments	1710	Rejected	HM	3
114	Ladenburg et al	2007	Unpublished	Enhancing Cheap Talk Scripts in Choice Experiments	170	Accepted	HM	6

S/No.	*Author(s)*	*Year*	*Journal/Unpublished**	*Research Topic/Title*	*Sample*	*Reported IIA Result*	*IIA Test Type*	*Number of Alternatives*
115	Li	2007	Economics of Education Review	Family background, financial constraints and higher education attendance in China	15536	Accepted	HM	3
116	Lockwood and Carberry	1998	Unpublished	Stated Preference Surveys of Remnant Native Vegetation Conservation	2258	Rejected	HM	3
117	Lusk and Schroeder	2004	American Journal of Agricultural Economics	Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks	592	Rejected	HM	6
118	Magnani	2009	Structural Change and Economic Dynamics	How Does Technological Innovation and Diffusion Affect Inter-Industry Workers Mobility	15004	Accepted	HM and SH	4
119	Mallawaarachchi et al	2006	Land Use Policy	Choice modelling to determine the significance of environmental amenity and production alternatives in the community value of peri-urban land	3116	Accepted	HM	4
120	Mansur et al	2008	Journal of Environmental Economics and Management	Climate change adaptation: A study of fuel choice and consumption in the US energy sector	5605	Accepted	HM	4
121	Martin and Stevenson	2001	American Journal of Political Science	Government Formation in Parliamentary Democracies	12466	Accepted	HM	6
122	Mataloni Jr.	2007	Unpublished	Do U.S. Multinationals Engage In Sequential Choice	641	Rejected	HM	7
123	Mazzanti	2003	Journal of Economic Studies	Valuation Experiments Application to Cultural Heritage	185	Accepted	HM	4
124	McCabe et al	2006	Journal of Health Economics	Using rank data to estimate health state utility models	611	Rejected	HM	8
125	Mirchandani and Bishai	2005	Unpublished	Healthcare Utilization and Choice of Provider	4864	Accepted	HM	3
126	Mishra and El-Osta	2008	Review of Economics of the Household	Effect of agricultural policy on succession decisions of farm households	1447	Accepted	HM	3
127	Mitchell and Fields	1983	Unpublished	Economic Incentives to Retire: A Qualitative Choice Approach	390	Rejected	HM	3

S/No.	*Author(s)*	*Year*	*Journal/Unpublished**	*Research Topic/Title*	*Sample*	*Reported IIA Result*	*IIA Test Type*	*Number of Alternatives*
128	Mogas et al	2006	Journal of Forest Economics	A comparison of contingent valuation and choice modelling with second-order interactions	4476	Rejected	HM	3
129	Mok	2007	Urban Studies	Choice of Residential Location	488	Rejected	HM	8
130	Moore	2004	Monthly Labor Review	Effectiveness of a defined benefit pension plan in meeting the income needs of retirees	4925	Accepted	Other/NS	5
131	Moore H.	2004	Unpublished	Measuring Defined Benefit Plan Replacement Rates Using PenSync	2508	Accepted	Other/NS	9
132	Nauges and Strand	2007	Resource and Energy Economics	Non-tap water demand	1379	Accepted	HM	4
133	Niu et al	2006	Economics of Education Review	College selectivity and the Texas top 10% law	5864040	Accepted	HM	5
134	O'Garra et al	2008	Energy Policy	Attitude to hydogen refuelling facilities	370	Accepted	HM	4
135	Oishi and Oshio	2006	The Japanese Journal of Social Security Policy	Coresidence with Parents and a Wife's Decision to Work in Japan	4981	Accepted	HM	3
136	Pardoe and Simonton	2008	Journal of the Royal Statistical Society	Applying discrete choice models to predict Academy Award winners		Accepted	HM	4
137	Patunru	2002	Unpublished	Econometric Consequences of Combining Hedonic Model and Conjoint Analysis for Environmental Valuation	506	Ambiguous	HM	3
138	Piracha and Vadean	2009	Unpublished	Occupational Choice of Return Migrants in Albania	3011	Accepted	HM	4
139	Pokhrel et al	2005	Health Policy	Gender role and child health care utilization in Nepal	8112	Accepted	HM and SH	4
140	Quesnel-Vallee and Morgan	2003	Population Research and Policy Review	Missing the target? Correspondence of fertility intentions and behavior in the U.S.	3172	Accepted	HM	3
141	Ran	2008	Unpublished	Three Papers on the Behavior Modeling of the Shrimp Fishermen in the Gulf of Mexico	9722	Accepted	HM	6
142	Rao et al	2007	Energy	Variations in energy use by Indian households	118000	Accepted	HM	4

S/No.	Author(s)	Year	Journal/Unpublished*	Research Topic/Title	Sample	Reported IIA Result	IIA Test Type	Number of Alternatives
143	Redlawsk and McCann	2005	Political Behavior	Popular Interpretations of 'Corruption' and Their Partisan Consequences	5281	Accepted	HM	4
144	Rodríguez and León	2004	Environmental and Resource Economics	Altruism and the Economic Values of Environmental and Social Policies	2334	Rejected	HM	3
145	Roessler et al	2008	Ecological Economics	Farmers' preferences for pig breeding traits	2091	Accepted	HM	4
146	Schwabe et al	2001	Environmental and Resource Economics	The Value of Changes in Deer Season Length: An Application of the Nested Multinomial Logit Model	5015	Rejected	Other/NS	4
147	Seko and Sumita	2006	Unpublished	Japanese Housing Tenure Choice after the Revision of the Tenant Protection Law:	279	Rejected	HM	3
148	Shafiq	2007	Journal of Asian Economics	Household schooling and child labor decisions in rural Bangladesh	3739	Accepted	HM	3
149	Shahian et al	2000	Journal of Thoracic and Cardiovascular Surgery	Selection of a cardiac surgery provider	6952	Accepted	HM	8
150	Shankar and Mannering	1996	Journal of Safety Research	An Exploratory Multinomial Logit Analysis of Single-Vehicle Motorcycle Accident Severity	650	Accepted	SH	5
151	Shih et al	2007	Pharmacoeconomics	Cost Effectiveness of Selective Drugs in Elderly Depressed Patients	1901	Accepted	HM	5
152	Shimizutani and Todo	2008	Unpublished	Overseas R&D Activities of Japanese Firms	1651	Accepted	HM	3
153	Shishikura et al	2005	Unpublished	Analysis of Subscription Demand for Pay-TV	513	Ambiguous	HM	3
154	Siegfried et al	2007	Unpublished	Choice of Currency in Bond Issuance and the International Role of Currencies	18280	Accepted	HM	7
155	Silvente and Gimenez	2007	Small Business Economics	Information Spillovers and the Choice of Export Destination: A Multinomial Logit Analysis of Spanish Young SMEs	454	Accepted	Other/NS	11

S/No.	Author(s)	Year	Journal/Unpublished*	Research Topic/Title	Sample	Reported IIA Result	IIA Test Type	Number of Alternatives
156	Sookram and Watson	2008	Journal of Eastern Caribbean Studies	Informal Sector and Gender in the Caribbean	3650	Accepted	HM and SH	8
157	Stratton et al 2008	2008	Economics of Education Review	College stopout and dropout behavior	3461	Accepted	HM	3
158	Strauss-Kahn and Vives	2008	Regional Science and Urban Economics	Why and where do headquarters move?	5341	Rejected	HM	10
159	Su et al	2006	European Journal of Health Economics	Determinants of household health expenditure on western institutional health care	2275	Accepted	HM	4
160	Sungyop et al	2007	Transportation Research Part A	Analysis of light rail rider travel behavior: Impacts of individual, built environment, and crime characteristics on transit access	407	Accepted	HM	4
161	Suzuki	2007	Transportation Research Part E	Modeling and testing the ‘‘two-step’’ decision process of travelers in airport and airline choices	459	Rejected	Other/NS	9
162	Tanaka	2008	Journal of The Japanese and International Economies	The gender-asymmetric effect of working mothers on children’s education: Evidence from Japan	2244	Accepted	HM	4
163	Tanner-Smith	2006	Drug and Alcohol Dependence	Pharmacological content of tablets sold as “ecstasy”: Results from an online testing service	1214	Accepted	HM	3
164	Tekin-Koru	2004	Unpublished	Is FDI Indeed Tariff-Jumping? Firm-Level Evidence	8940	Accepted	HM	3
165	Thind	2004	Journal of Community Health	Home Deliveries in Indonesia: Who Provides Assistance?	10692	Accepted	HM and SH	3
166	Timmermans and Borgers	1985	Sistemi Urbani	Consumer spatial shoping behaviour	86	Rejected	Other/NS	5
167	Tzioumis	2008	Journal of Economic Behavior and Organization	Why do firms adopt CEO stock options? Evidence from the United States	13042	Accepted	HM	4
168	van de Vrande et al	2007	Journal of Business Venturing	External technology sourcing: The effect of uncertainty on governance mode choice	1810	Accepted	HM	5

S/No.	Author(s)	Year	Journal/Unpublished*	Research Topic/Title	Sample	Reported IIA Result	IIA Test Type	Number of Alternatives
169	Vartia	2005	Unpublished	Establishing and Closing Down Plants - Assessing the Effects of Firms' Financial Status	52717	Accepted	HM and SH	3
170	Wang et al	2007	Ecological Economics	Estimating non-market environmental benefits of the Conversion of Cropland to Forest and Grassland Program: A choice modeling approach	2920	Rejected	HM	3
171	Weber and Mahringer	2008	Empirical Economics	Choice and success of job search methods	500	Accepted	Other/NS	6
172	Wennberg	2008	Journal of Evolutionary Economics	Knowledge combinations and the survival of financial services ventures	1077	Accepted	HM	3
173	Wennberg	2006	Unpublished	A Real Options Model of Stepwise Entry into Self-employment	236045	Accepted	HM	3
174	Wennberg and Lindqvist	2008	Small Business Economics	The effect of clusters on the survival and performance of new firms	2124	Accepted	HM	4
175	Wennberg et al	2007	Unpublished	A Real Options Model of Stepwise Entry into Self-Employment	236045	Accepted	HM	3
176	Wilander	2004	Unpublished	Currency Denomination in International Trade	192582	Accepted	HM	4
177	Xu et al	2007	Journal of Arid Environments	Choice modeling and its application to managing the Ejina Region, China	4709	Rejected	Other/NS	3
178	Yanik and Assaad	2004	Unpublished	Women's Participation in Paid Urban Work	16075	Rejected	HM	3
179	Yu	2003	Unpublished	A Nested Logit Approach to Airline Operations Decision Process	85539	Rejected	Other/NS	3
180	Zavodny	2008	Review of Economics of the Household	Is there a 'marriage premium' for gay men	4913	Accepted	HM and SH	4
181	Zhai and Suzuki	2008	China Economic Review	Public willingness to pay for environmental management, risk reduction and economic development: Evidence from Tianjin, China	898	Accepted	Other/NS	4

* Unpublished at the time of this research
HM – Hausman-McFadden Test
SH – Small-Hsiao Test
Other/NS – Other and Not Specific/Not Stated

Appendix V: Our Residential Choice Questionnaire

Note: The questionnaire administered was printed in a slightly different format

Target: Current Renting Residents of Kano Metropolitan area

This is a questionnaire for a research on household's residential location decision in Kano, Nigeria. The research is towards a PhD degree in Economics at the University of East Anglia, Norwich, United Kingdom. The research is part-funded, jointly, by the McArthur Foundation and Bayero University Kano.

To be completed by Research Assistants
Local Government Area:
Date:
Time:

ALL INFORMATION DIVULGED IN COMPLETING THIS QUESTIONNAIRE WILL BE TREATED CONFIDENTIALLY AND USED PURELY FOR ACADEMIC RESEARCH

Instructions:
Please mark ☐ in the box which corresponds to your choice and/or fill in the answer in spaces provided

1. Respondents Initials (Optional) ---

2. Current Place of Residence:

a. Area A – The old city ☐
b. Area B – Low density/Government Reserved Areas ☐
c. Area C – Close to the airport ☐
d. Area D – Close to one of the two industrial estates ☐
e. Area E – Other ☐

3. Street Name: ---

4. How many bedrooms are there in your house?:

a. 1 Bedroom ☐
b. 2 Bedrooms ☐
c. 3 Bedrooms ☐
d. 4 Bedrooms ☐
e. 5 Bedrooms ☐
f. 6 Bedrooms and above ☐

5. How many bathrooms/toilets are there in the house? ☐ ☐

6. **Is your residence:**

a. A Flat ☐
b. A Bungalow ☐
c. A Duplex ☐
d. A Traditional House ☐
e. Other (Please Specify) ☐ ------------------------

7. **What is the size of the floor area in sq ft?**

a. Less than 2500 sq ft ☐
b. 2501 – 3000 sq ft ☐
c. 3001 – 3500 sq ft ☐
d. 3501 – 4000 sq ft ☐
e. More than – 4000 sq ft ☐

8. **How large is your garden/yard if any, in sq ft?**

a. Less than 200 sq ft ☐
b. 201 – 300 sq ft ☐
c. 301 – 500 sq ft ☐
d. 501 – 1000 sq ft ☐
e. 1001 – 1000 sq ft ☐
f. More than – 1000 sq ft ☐
g. Not applicable ☐

9. **Who is the house rented from or provided by?**

a. Local Authority/Council ☐
b. Property company ☐
c. Employer ☐
d. Other organisation (Please Specify) ☐ ------------------------
e. Other individual ☐
f. Not applicable ☐

10. **Please indicate if you live close any the following:**

a. Government Approved Private Nursery/Primary School ☐
b. Public Primary School ☐
c. Major Market ☐
d. Highway/by-pass/express ☐
e. Airport ☐
f. Industrial Estate ☐
g. None of the above ☐

11. If you live close to the Airport, Is your residence in line with runway/under the flight path?

a. Yes ☐

b. No ☐

c. Not Applicable ☐

12. How much is the annual rent?

a. ₦30,000 and Below ☐

b. ₦31,000 – ₦39,000 ☐

c. ₦40,000 – ₦49,000 ☐

d. ₦50,000 – ₦59,000 ☐

e. ₦60,000 – ₦69,000 ☐

f. ₦70,000 – ₦79,000 ☐

g. ₦80,000 – ₦99,000 ☐

h. ₦100,000 – ₦149,000 ☐

i. ₦150,000 – ₦199,000 ☐

j. ₦200,000 – and above ☐

k. Not Applicable ☐

13. On average, how many hours do you receive electricity at home everyday

a. None ☐

b. 1 – 2 hours ☐

c. 3 – 6 hours ☐

d. 7 – 12 hours ☐

e. 13 – 15 hours ☐

f. 16 – 24 hours ☐

14. On average, how many hours do you receive water supply at home everyday

a. None ☐

b. 1 – 2 hours ☐

c. 3 – 6 hours ☐

d. 7 – 12 hours ☐

e. 13 – 15 hours ☐

f. 16 – 24 hours ☐

15. Apart from family living with you, do you live within walking distance of close relatives?

a. Yes ☐

b. No ☐

If Yes, please go to question 18: Otherwise continue to next question.

16. How often do you visit your close relatives in an average month?

Number of visits per week ☐ ☐

Not Applicable ☐ **Please go to question 18**

17. Please provide your best estimate of the following cost of commuting for each visit

a. Fuel - Petrol/Diesel ₦ ☐ ☐ ☐ ☐

b. Bus Fares ₦ ☐ ☐ ☐ ☐

c. Other cost 1 (Please Specify)............... ₦ ☐ ☐ ☐ ☐

d. Other cost 2 (Please Specify)............... ₦ ☐ ☐ ☐ ☐

e. Other cost 3 (Please Specify)............... ₦ ☐ ☐ ☐ ☐

18. How do you travel to work/place of business?

a. Car ☐

b. Bus ☐

c. Cycle ☐

d. Motorcycle ☐

e. Walk ☐

f. Others (Please Specify) ☐

g. Not Applicable ☐ **Please go to question 21**

19. In a typical day, how long does it take you to get to work/place of business from home?

a. Less than 30 minutes ☐

b. 30 minutes - 59 minutes ☐

c. 1 hour - 1 hour 30 minutes ☐

d. 1 hour 30 minutes – 2 hours ☐

e. More than 2 hours ☐

20. Please provide your best estimate of the following cost of commuting to work/place of business, <u>monthly</u>

a. Fuel - Petrol/Diesel ₦ ☐ ☐ ☐ ☐

b. Bus Fares ₦ ☐ ☐ ☐ ☐

c. Other cost 1 (Please Specify)............... ₦ ☐ ☐ ☐ ☐

d. Other cost 2 (Please Specify)............... ₦ ☐ ☐ ☐ ☐

e. Other cost 3 (Please Specify)............... ₦ ☐ ☐ ☐ ☐

21. **Consider the last residence you occupied before moving to the current residence. Was it?**

Reminder:

Area A – The old city

Area B – Low density/Government Reserved Areas

Area C – Close to the airport

Area D – Close to one of the two industrial estates

Area E – Other

a. Area A, B, C, D or E ☐

b. In another city/town ☐

c. No previous residence ☐

22. **Respondent's highest education qualification:**

a. Non-formal education ☐

b. Vocational Education ☐

c. Primary ☐

d. Secondary ☐

e. Diploma ☐

f. University degree/higher diploma ☐

g. Postgraduate degree ☐

h. Other qualifications (Please specify) ☐ --------------------------------

i. None of the above ☐

23. **Gross Annual Income (in local currency):**

a. Less than 100,000 ☐

b. 100,000 – 299,999 ☐

c. 300,000 – 499,999 ☐

d. 500,000 – 699,999 ☐

e. 700,000 – 999,999 ☐

f. 1,000,000 – 1,499,999 ☐

g. 1,500,000 – 1,999,999 ☐

h. 2,000,000 – 2,999,999 ☐

i. 3,000,000 and above ☐

24. **Respondent's other major source of Income:**

a. Support from family members ☐
b. Providence ☐
c. Fixed Assets ☐
d. Financial Investment ☐
e. Secondary Occupation (e.g. Part-time) ☐
f. Private Consultancy ☐
g. Other (please specify) ☐ -------------------------
h. None ☐

25. **Nature of Current Occupation:**

a. Manual/Blue Collar ☐
b. Businessman ☐
c. Civil servant/Administrator ☐
d. Teacher ☐
e. Corporate Sector/White Collar ☐
f. Lecturer/Researcher ☐
g. Farmer ☐
h Security Worker ☐
i. Law Enforcement Agent (Police etc) ☐
j. Medical Doctor ☐
k. Nurse/Midwife/Social Worker ☐
l. Retired from paid work ☐
m. Unemployed ☐
n Full time education ☐
o. Other (please specify) ☐ --------------------------------

26. **Spouse Highest Education Qualification**

a. Non-formal education ☐
b. Vocational Education ☐
c. Primary ☐
d. Secondary ☐
e. Diploma ☐
f. University degree/higher diploma ☐
g. Postgraduate degree ☐
h. Islamiyya ☐
i. Other qualifications (Please specify) ☐ -------------------------------
j. None of the above ☐
k. No Spouse ☐

27. Respondent's spouse current occupation

a. Manual/Blue Collar ☐
b. Businessman ☐
c. Administrator/Civil servant ☐
d. Teacher ☐
e. Corporate Sector/White Collar ☐
f. Lecturer/Researcher ☐
g. Farmer ☐
h. Security Worker ☐
i. Law Enforcement Agent (Police etc) ☐
j. Medical Doctor ☐
k. Nurse/Midwife/Social Worker ☐
l. Retired from paid work ☐
m. Unemployed ☐
n. Full Time Education ☐
o. Other (please specify) ☐ --------------------------------
p. No Spouse ☐

28. Consider each of the following determinants of your current residential location choice. Please indicate the importance of each (5 means very important; 0 means irrelevant).

	Irrelevant					**Very Important**
Noise pollution	0☐	1☐	2☐	3☐	4☐	5☐
Air pollution	0☐	1☐	2☐	3☐	4☐	5☐
Proximity to place of work	0☐	1☐	2☐	3☐	4☐	5☐
Security/Law and order	0☐	1☐	2☐	3☐	4☐	5☐
Easy access to other parts of the city	0☐	1☐	2☐	3☐	4☐	5☐
Schools in the neighbourhood	0☐	1☐	2☐	3☐	4☐	5☐
Social interaction	0☐	1☐	2☐	3☐	4☐	5☐
Rent	0☐	1☐	2☐	3☐	4☐	5☐
Proximity to local shops/market	0☐	1☐	2☐	3☐	4☐	5☐
Local business opportunities	0☐	1☐	2☐	3☐	4☐	5☐
Family and friendship ties	0☐	1☐	2☐	3☐	4☐	5☐
Electricity supply	0☐	1☐	2☐	3☐	4☐	5☐
Water supply	0☐	1☐	2☐	3☐	4☐	5☐
Influence of landlord/rent provider	0☐	1☐	2☐	3☐	4☐	5☐

29. Respondent's Age:

a. 20 and Below ☐
b. 21 – 29 ☐
c. 30 – 39 ☐
d. 40 – 49 ☐
e. 50 – 59 ☐
f. 60 and above ☐

30. Marital Status:

a. Single ☐
b. Married ☐
c. Divorced ☐
d. Widowed ☐
e. Separated ☐

31. Household/Family Size: ☐ ☐

32. Number of children in following category

	Age in Years	**Number of Children**
a.	0 – 4	
b.	5 – 6	
c.	7 – 12	
d.	13 – 18	
e.	18 and above	

End of questionnaire

Thank you for completing the questionnaire. Your answers will be very useful. You are reminded that, all your answers will be treated confidentially and purely for academic purposes.

References

Ackerberg, D. A. and M. Rysman (2005), "Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects" http://people.bu.edu/mrysman/research/proddiff.pdf

Adamowicz, W. et al. (2005), "Decision Strategy and Structure in Households: A "Groups" Perspective", ***Marketing Letters***, Vol. 16(3-4) pp. 387-399

Adda, J. and R. Cooper (2000), "The dynamics of car sales: a discrete choice approach", Cambridge, NBER. Working Paper No. 7785 http://www.nber.org/papers/W7785

Ahmad, N. (1994), "A joint model of tenure choice and demand for housing in the city of Karachi", ***Urban Studies***, 31(10), pp. 1691–1706.

Alonso, W. (1964), ***Location and Land Use***, Cambridge Massachusetts, Harvard University Press

Amemiya, T. (1981), "Qualitative Response Models: A survey", ***Journal of Economic Literature***, Vol. 19(4), pp. 1483-1536

Amemiya, T. (1985), ***Advanced Econometrics***, Basil Blackwell, Oxford

Anas, A. and R. J. Arnott (1991), "Dynamic Housing Market Equilibrium with Taste Heterogeneity, Idiosyncratic Perfect Foresight and Stock Conversions", ***Journal of Housing Economics***, Vol. 1(1), pp. 2-32.

Anderson, S. P. (1989), "Demand for Differentiated Products, Discrete Choice Models and the Characteristics Approach", ***Review of Economic Studies***, Vol. 56(1), pp. 21-35.

Anderson, S. P. (2005), ***Product Differentiation***, New Palgrave Dictionary contribution

Anderson, S. P. et al (1992), ***Discrete Choice Theory of Product Differentiation***, Cambridge MA.: MIT Press.

Antonini, G. (2005), "A discrete choice modeling framework for pedestrian walking behavior with application to human tracking in video sequences" http://biblion.epfl.ch/EPFL/theses/2005/3382/EPFL_TH3382.pdf

Arimah, B. C. (1992), An Empirical Analysis of the Demand for Housing Attributes in a Third World City, *Land Economics*, Vol. 68(4), pp. 366-379

Ariyo, A. and A. Jerome (2004), "Utility privatization and the poor: Nigeria in Focus", Global Issue papers, Heinrich Boll Foundation, Berlin, No. 12, Available at: http://www.boell.de/downloads/internationalepolitik/GIP12.pdf

Arne R, H. and F. R. FitzRoy (2004), "Commuting in small towns in rural areas: The case of St Andrews". St. Andrews, Available at: http://129.3.20.41/eps/urb/papers/0312/0312001.pdf

Arnott, R. (1987), “Economic Theory and Housing”, in E. S. Mills (ed), ***Handbook of Regional and Urban Economics***, Vol. 2 – Urban Economics, Elsevier Science, Amsterdam, North Holland, pp.959-88

Atkinson, R. and G. Bridge (eds) (2005), ***Gentrification in a Global Context: The New Urban Colonialism***. Routledge: London.

Asensio, J. (2002), “Transport Mode choice by commuters to Barcelona’s CBD”, ***Urban Studies***, 39(10), 1881-1895

Aufhauser, E. et al (1986), “A disaggregated probabilistic approach to a regulated housing market with an emphasis on the demand side: The Vienna case”, ***Papers in Regional Science***, 60(1), pp. 133-153.

Bacharach, M. (1999), “Interactive team reasoning: a contribution to the theory of cooperation”, ***Research in Economics***, 53(2), pp. 117-47

Baharad, E. and S. Nitzan (2003), “Essential alternatives and set-dependent preferences-an axiomatic approach”, ***Mathematical Social Sciences***, Vol. 45(2), pp. 121-129

Bailar, J. C. III (1995), “The Practice of Meta-Analysis”, ***Journal of Clinical Epidemiology***, Vol. 48(1), pp. 149-157

Bair, E. and J. M. Fitzgerald (2005), “Hedonic Estimation and Policy Significance of the Impact of HOPE VI on Neighborhood Property Values”, ***Review of Policy Research***, Vol. 22(6), pp. 771-786

Bajari, P. and C. L. Benkard (2001), “Discrete Choice Models as Structural Models of Demand: Some Economic Implications of Common Approaches”, Available at:. http://www.stanford.edu/~lanierb/research/al3-13-03.pdf

Bajari, P. and M. Kahn (2001), “Why Do Blacks Live in The Cities and Whites Live in the Suburbs?” Available at: www-econ.stanford.edu/faculty/workp/swp00007.pdf

Barnett, C. J. 1(985), “An Application of the Hedonic Price Model to the Perth Residential”, ***Land Market, Economic Record***, Vol. 61(1), pp. 476-48 1

Bartik, T. J. (1986), "Neighborhood Revitalization's Effects on Tenants and the Benefit-Cost Analysis of Government Neighborhood Programs", Journal of Urban Economics, Vol. 19(2), pp. 234-48

Bartik, T. J. (1987), “The Estimation of Demand Parameters in Hedonic Price Models”, ***Journal of Political Economy***, Vol. 95(1), pp. 81-88

Bartik, T. J. (1988), “Measuring the Benefits of Amenity Improvements”, ***Land Economics***, Vol. 64(2), pp. 172-183

Bartik, T. J. and V. K. Smith (1987), "Urban Amenities and Public Policy", in E. S. Mills (ed) ***Handbook of Regional and Urban Economics***, Elsevier Science, North Holland pp. 1207-1254

Bartik, T. J. et al (1992), "Maximum score estimates of the determinants of residential mobility: Implications for the value of residential attachment and neighborhood amenities", ***Journal of Urban Economics***, Vol. 32(2), pp. 233-256

Barzyk, F. (1999), "Updating the Hedonic Equations for the Price of Computers." Working Paper of Statistics Canada, Prices Division, November. Available at: http://www.statcan.ca/english/sdds/document/5032_D2_T9_V1_E.pdf.

Bastin, F. et al (2005), "Evaluation of optimisation methods for estimating mixed logit models", Paper prepared for presentation at the 84th Annual Meeting of the Transportation Research Board, Washington, D.C. Available at: http://www.trb-forecasting.org/papers/2005/ADB40/05-2340_Bastin.pdf

Basu, K. (2000), ***Prelude to Political Economy: A Study of the Social and Political. Foundations of Economics***, Oxford University Press, Oxford

Basu, K. (2006), "Gender and Say: A Model of Household Behaviour with Endogenously Determined Balance of Power", ***The Economic Journal***, 116 (511), pp. 558-580

Batalhone, S. A. et al. (2002), "Economics of Air Pollution: Hedonic Price Model and Smell Consequences of Sewage Treatment Plants in Urban Areas", Departamento De Economia Universidade de Brasília Texto Para Discussão No 234 Available at: http://www.unb.br/face/eco/cpe/TD/234August2002JNogueira.pdf

Bateman, I. et al. (2001), The Effect of Road Traffic on Residential Property Values: A Literature Review and Hedonic Pricing Study the Scottish Executive Development Department Available at: http://www.scotland.gov.uk/library3/housing/ertpv.pdf

Bateman, I. J. et al (2005), "Economic valuation of policies for managing acidity in remote mountain lakes: Examining validity through scope sensitivity testing", ***Aquatic Sciences***, Vol. 67, pp. 274–291

Batalhone, S. A. et al. (2002), Economics of Air Pollution: Hedonic Price Model and Smell Consequences of Sewage Treatment Plants in Urban Areas, Departamento De Economia Universidade de Brasília Texto Para Discussão No 234 Available at: http://www.unb.br/face/eco/cpe/TD/234August2002JNogueira.pdf

Bayer, P. F. et al (2005), "An Equilibrium Model of Sorting in an Urban Housing Market" Available at: http://aida.econ.yale.edu/~pjb37/BayerMcMillanRueben_IER.pdf

Bayer, P. F. et al (2007), "A Unified Framework for Measuring Preferences for Schools and Neighborhoods***" Journal of Political Economy***, Vol. 115(4), pp. 588-638

Bechtel, G. G. (1990), “Share-Ratio Estimation of the Nested Multinomial Logit Model”, ***Journal of Marketing Research***, Vol. 27(2), pp. 232-237

Bell, K. P. and M. McMahon (2003), “A Spatial Analysis of Household Location Decisions in Maine”, Submission for the AERE 2003 Workshop, Wisconsin Madison., Available at: http://www.aere.org/meetings/0306workshop_Bell.pdf.

Ben-Akiva, M. et al (1997), “Modelling Methods for Discrete Choice Analysis”, ***Marketing Letter***, Vol. 8(3), pp. 273-86

Ben-Akiva, M. and T. Watanatada (1981), “Application of a Continuous Spatial Choice Logit Model”, in Manski, C. F. and McFadden D. (eds.), ***Structural Analysis of Discrete Data with Econometric Applications***, MIT Press http://elsa.berkeley.edu/discrete/ch8.pdf

Ben-Akiva, M. and S. R. Lerman (1985), ***Discrete Choice Analysis: Theory and Application to Predict Travel Demand***. Cambridge MA: MIT Press.

Ben-Akiva, M. and M. Bierlaire (1999), “Discrete choice methods and their applications to short term travel decisions”, in Hall, R. W. (ed), ***Handbook of Transportation Science***, Kluwer Academic Publishers, Norwell, Massachusetts.

Ben-Akiva, M. et al (2002), “Hybrid choice models: progress and challenges”, ***Marketing Letters***, Vol. 13(3), 163-175.

Ben-Akiva, M. and M. Bierlaire (2003), “Discrete choice models with applications to departure time and route choice”, in Hall, R. (ed.), ***Handbook of Transportation Science***, Kluwer

Bergstrom, T. C. (1997), A Survey of Theories of The Family, in M. R. Rosenzweig, and O. Stark (eds) ***Handbook of Population and Family Economics***, Elsevier Science, North Holland pp. 21-79

Berry, S. T. (1994), “Estimating discrete choice models of product differentiation”, ***The RAND Journal of Economics***, Vol. 25(2), pp. 242-262.

Berry, S. et al (1995), “Automobile Prices in Market Equilibrium,” ***Econometrica***, Vol. 63(4), pp. 841-890

Bhattarai, G. R. et al. (2005), “Demand for School Quality, Neighborhood Safety and Environmental Quality: Reflections from Real Estate Market”, Twelfth Annual International Conference on Statistics, Combinatorics, Mathematics and Applications, Auburn University, Auburn, Alabama, USA. Available at: http://ssrn.com/abstract=800887

Bierens, H. J. (2005), ***Introduction to the Mathematical and Statistical Foundations of Econometrics***, Cambridge, Cambridge University Press

Black, D. A. and T. J. Kniesner (2003), “On the Measurement of Job Risk in Hedonic Wage Models”, ***Journal of Risk and Uncertainty***, 27: pp. 205–220

Black, S. (1999), “Do Better Schools Matter? Parental Valuation of Elementary Education”, ***Quarterly Journal of Economics***, Vol. 114(2), pp. 577-99

Bockstael, N. E. and K. E. McConnell (2007), ***Environmental and Resource Valuation with Revealed Preferences: A Theoretical Guide to Empirical Models***, Springer, Dordrecht, The Netherlands

Borenstein, M. et al (2009), ***Introduction to Meta-Analysis***, Chichester, John Wiley & Sons

Borsch-Supan, A. (1987), ***Econometric Analysis of Discrete Choice: With applications on the Demand for Housing in the US and West Germany***, Lecture Notes in Economics and Mathematical System Series, N0 296, Berlin, Spinger-Verlag

Borsch-Supan, A. and J. Pitkin (1988), “On discrete choice models of housing demand”, ***Journal of Urban Economics***, 24, pp. 153–172.

Borsch-Supan, A. (1990), Panel data of housing choices, ***Regional Science and Urban Economics***, Vol. 20, pp. 65–82.

Bourassa, S. C. and M. Hoesli (1999), “The Structure of Housing Submarkets in a Metropolitan Region”, Available at: http://www.hec.unige.ch/recherches_publications/cahiers/1995-1999/99.15.pdf

Bourassa, S. C. at al. (2003), “Do Housing Submarkets Really Matter?” ***Journal of Housing Economics***, Vol. 12(1), pp. 12-28

Brasington, D. M. and D. Hite (2005), “Demand for Environmental Quality: A Spatial Hedonic Analysis”, ***Regional Science and Urban Economics***, Vol. 35(1), pp. 57-82

Breffle, W. and E. R. Morey (2000), “Investigating Preference Heterogeneity in a Repeated Discrete-Choice Recreation Demand Model of Atlantic Salmon Fishing”, ***Marine Resource Economics***, Vol. 15(1), pp. 1-20.

Brewer, M. B. and W. Gardner, (1996), “Who is this `we'? Levels of collective identity and self representation”, ***Journal of Personality and Social Psychology***, 71(3), pp. 83-93

Brook, W. and S. N. Durlauf (2001), “Discrete choice with social interaction”, ***Review of Economic Studies***, Vol. 68(2), pp. 235-260

Browning, M. et al (1994), “Income and outcomes: A structural model of intrahousehold allocation”, ***Journal of Political Economy***, Vol. 102(6), pp. 1067–1096.

Brownstone, D. and K. Train (1999), “Forecasting New Product Penetration with Flexible Substitution Patterns,” ***Journal of Econometrics***, Vol. 89, pp 109-129

Cai, J. (1989), “Are husbands and wives altruistically linked evidence from micro-data” Boston University Working paper

Camerer, C. (1995), “Individual decision making”, in Kagel, J. and A. E. Roth (eds) ***Handbook of Experimental Economics***, Princeton, NJ

Cameron, A. C. and P. K. Trivedi (2005), ***Microeconometrics: Methods and Applications***, Cambridge University Press, Cambridge

Cameron, A. C. et al (2008), "Bootstrap-Based Improvements for Inference with Clustered Errors", ***Review of Economics and Statistics***, Vol. 90(3), pp. 414-427

Cameron, T. A. (1998), "Evolution of Hedonic Property Value Models for the Valuation of Environmental Goods", Department of Economics University of California, Los Angeles Available at: http://www.sscnet.ucla.edu/ssc/labs/cameron/nrs98/hedoninv.htm

Campbell, D. (2007), "Willingness to pay for rural landscape improvements: combining mixed logit and random effects models", ***Journal of Agricultural Economic***, Vol. 58(3), pp. 467-483

Card D. and A. B. Krueger (1995), "Time-Series Minimum wage studies: A Meta-analysis", ***American Economic Review***, 85, 238-243.

Cao, X. and J. A. Hough (2007), "Hedonic Value of Transit Accessibility: An Empirical Analysis in a Small Urban Area" The Upper Great Plains Transportation Institute (UGPTI), North Dakota State University. Available at: http://www.ugpti.org/pubs/pdf/DP191.pdf

Cassel, E. and R. Mendelsohn (1985), "The choice of functional forms for hedonic price equations: Comment", ***Journal of Urban Economics***, Vol. 18(2), pp. 135-142

Chattopadhyay, S. (1999), "Estimating the demand for air quality: New evidence based on the Chicago housing market", ***Land Economics***, Vol. 75(1), pp. 1- 22

Chau K.W. and T. L. Chin (2002), "A Critical Reveiw of the Literature on the Hedonic Pricing Model and Its application to the Housing Market in Penang", The Seventh Asian Real Estate Society Conference. Seoul, Korea, Available at: http://www.kreaa.org/AsRES/doc/Chin%20Tung%20Leong(D3),.doc.

Chay, K. Y. and M. Greenstone (2005), Does Air Quality Matter? Evidence from the Housing Market, ***Journal of Political Economy***, Vol. 113(2), pages 376–424

Cheng, S. and J. S. Long (2007), "Testing for IIA in the Multinomial Logit Model", ***Sociological Methods & Research***, Vol. 35(4) pp. 583-600

Cheshire, P. and S. Sheppard (2002), Capitalising the Value of Free Schools: The Impact of Land Supply Constraints, ERSA conference papers ersa03p8, European Regional Science Association. Available at: http://www.ersa.org/ersaconfs/ersa03/cdrom/papers/8.pdf

Chesher, A. and J. M. C. Santos Silva (2002), "Taste Variation in Discrete Choice Models", ***Review of Economic Studies***, Vol. 69(1), pp. 147-168

Cheshire, P. and S. Sheppard (2002), "Capitalising the Value of Free Schools: The Impact of Land Supply Constraints", ERSA conference papers ersa03p8, European Regional Science Association. Available at: http://www.ersa.org/ersaconfs/ersa03/cdrom/papers/8.pdf

Chin, T. L. and KW Chau (2002), "A Critical Review of Literature on the Hedonic Price Model and Its Application to the Housing Market in Penang", Seventh Annual Conference of the Asian Real Estate Society (AsRES), Available at: http://www.kreaa.org/AsRES/doc/Chin%20Tung%20Leong(D3),.doc.

Christiadi, C. and B. Cushing (2007), "Conditional Logit, IIA, and Alternatives for Estimating Models of Interstate Migration", Research Paper #2007-4 Regional Research Institute, West Virginia University. Available at: http://www.rri.wvu.edu/pdffiles/wpcushing2007-4.pdf

Chuah, S. H. et al (2004), "Do Cultures Clash? Evidence from Cross-National Ultimatum Game Experiments" Nottingham University Industrial Economics Division, Occasional Papers #9. Available at: http://www.nottingham.ac.uk/%7Elizecon/RePEc/pdf/behaviour.pdf

Cho, C.-J. (1997), "Joint Choice of Tenure and Dwelling Type: A Multinomial Logit Analysis for the City of Chongju", ***Urban Studies***, Vol. 34 pp. 99 1459-73

Clark, W. A. V. and W. F. J. Van Lierop (1986), "Residential Mobility and Household Location Modelling", in P. Nijkamp (ed), ***Handbook of Regional and Urban Economics***, Vol. 1 – Regional Economics, Amsterdam: North Holland, pp. 97-132

Cleveland, W. S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots" ***Journal of the American Statistical Association***, Vol. 74(368), pp. 829-836

Cleveland, W. S. and S. J. Devlin (1988), "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting", ***Journal of the American Statistical Association***, Vol. 83(403), pp. 596-610

Coelli, T., J. et al (1991), "Hedonic Pricing for a Cost Benefit Analysis of a Public Water Supply Scheme," ***The Australian Journal of Agricultural Economics***, 35 (1), pp. 1-20

Cohen, J. P. and C. C. Coughlin (2006), "Spatial Hedonic Models of Airport Noise, Proximity, and Housing Prices", Federal Reserve Bank of St. Louis Working Paper 2006-026a. Available at: http://research.stlouisfed.org/wp/2006/2006-026.pdf

Colom, M. C. and M. C. Molés (2008), "Comparative Analysis of the Social, Demographic and Economic Factors that Influenced Housing Choices in Spain in 1990 and 2000", ***Urban Studies***, Vol. 45 pp. 917

Colwell, P. F. and G. Dilmore (1999), "Who was first? An examination of an early hedonic study", ***Land Economics***, Vol. 75(4), pp. 620-626

Correll, M. R. at al. (1978), "The effects of greenbelts on residential property values: Some findings on the political economy of open space", ***Land Economics***, Vol. 54(2), pp. 206-217

Cramer, J. S. (1991), ***The logit model: An Introduction for economists***. London: Edward Arnold.

Cramer, J. S. (2003), ***Logit Models from Economics and Other Fields***, Cambridge, Cambridge University Press

Cramer, J. S. (2003), The origins and development of the logit model, Available at: http://www.cambridge.org/resources/0521815886/1208_default.pdf.

Cropper, M. L. et al (1988), "On the Choice of Functional Form for Hedonic Price Functions", ***Review of Economics and Statistics***, Vol. 70(4), pp. 668-75

Dalhuisen, J.M., R. Florax, F. De Groot, P. Nijkamp, (2003), "Price and Income Elastcities of Residential Water Demand: A Meta Analaysis", ***Land Economics***, 79, 292-308.

Day, B. (2001), "The Theory of Hedonic Markets: Obtaining welfare measures for changes in environmental quality using hedonic market data", Report for the EU Working Group on Noise. Available at: http://www.cserge.ucl.ac.uk/publications.html

Day, B. (2003), "Submarket Identification in Property Markets: A Hedonic Housing Price Model for Glasgow", CSERGE Working Paper Edm 03-09 Available at: http://www.uea.ac.uk/env/cserge/pub/wp/edm/edm_2003_09.pdf

Day, B. et al (2003), "What Price Peace? A Comprehensive Approach to the Specification and Estimation of Hedonic Housing Price Models", CSERGE Working Paper EDM 03-08 Available at: http://www.uea.ac.uk/env/cserge/pub/wp/edm/edm_2003_08.pdf

Daykin, A. R. and P. G. Moffatt (2002), "Analyzing Ordered Responses: A Review of the Ordered Probit Model", ***Understanding Statistics***, Vol. 1(3), pp. 157-166

Diamond Jr., D. B. (1980), "Income and Residential Location: Muth Revisited", ***Urban Studies***, Vol. 17(1), pp. 1-12

DiPasquale, D. (1999), "Why Don't We Know More about Housing Supply?" ***The Journal of Real Estate Finance and Economics***, Vol. 18(1), pp. 9-23

Deininger, A. (1999), "Perception of Public School Quality and It's Effects on Housing Prices: Evidence from Pitt County, North Carolina", Graduate Paper, Department of Economics, East Carolina University. Available at: http://www.ecu.edu/econ/ecer/deininger.pdf

Des Rosiers, F. at al. (1996), "Shopping centres and house values: An empirical investigation", ***Journal of Property Valuation and Investment***, Vol. 14(4), pp. 41-62

Dewentera, R. et al (2007), "Hedonic prices in the German market for mobile phones", ***Telecommunications Policy***, Vol. 31(1), pp.4 -13

Downes, T. A. and J. E. Zabel (1997), "The impact of school characteristics on house prices: Chicago 1987-1991", Available at: http://ase.tufts.edu/econ/papers/9704.pdf

Eastwood, D. B. et al (1986), "Household Nutrient Demand: Use of Characteristics Theory and A Common Attribute Model", ***Southern Journal of Agricultural Economics***, Vol. 18(2), pp. 235-246

Earnhart, D. (2002), "Combining Revealed and Stated Data to Examine Decisions of Housing Location: Discrete-Choice Hedonic and Conjoint Analysis", ***Journal of Urban Economics***, Vol. 51(1), pp. 143-169

Edmonds, R. (1984), "A theoretical basis for hedonic regression: A research primer", ***American Real Estate and Urban Economics Association Journal***, Vol. 12(1), pp. 72-85

Ekeland, I. et al. (2004), "Identification and Estimation of Hedonic Models," ***Journal of Political Economy***, Vol. 112(S1), pp. S60-S109

Elder, H. and Zumpano, L. (1991), Tenure Choice, Housing Demand and Residential Location, ***Journal of Real Estate Research***, 6(3), pp. 341–356.

Ellicott, K. (ed) (2002), ***Cities of the World***, 6th Edition, Vol. 1: Africa, The Gale Group, Farmington Hills, Detroit, pp. 442.

Espey, M. and H. Lopez (2000), "The impact of airport noise and proximity on residential property values", ***Growth and Change***, Vol. 31(3), pp. 408-419

Epp, D. J. and K. S. Al-Ani (1979), "The Effect of Water Quality on Rural Nonfarm Residential Property Values", ***American Journal of Agricultural Economics***, Vol. 61(3), pp. 529-534

Evan, A. W. (1985), ***Urban Studies***, Oxford: Basil Blackwell. pp. 14-36

Feather, P. M. (1994), "Sampling and Aggregation Issues in Random Utility Model Estimation," ***American Journal of Agricultural Economics***, Vol. 76(4), pp 772-780

Feridhanusetyawan, T. and M. Kilkenny (1996), "Rural/Urban Residence Location Choice", Center For Agricultural and Rural Development Working Paper 96-Wp #157

Florax, R. (2002), "Accounting for dependence among study results in meta-analysis: methodology and applications to the valuation and use of natural resources", Amsterdam: Faculty of Economics and Business Administration, Free University, Research Memorandum 5,

Follain, J. R. and E. Jimenez (1985 a), "Estimating the demand for housing characteristics", ***Regional Science and Urban Economics***, Vol. 15(1), pp. 77-107

Follain, J. R. and E. Jimenez (1985 b), The Demand for Housing Characteristics in Developing Countries, ***Urban Studies***, Vol. 22(1), pp. 421-432

Folmer, H. et al eds, (1995), ***Principles of Environmental and Resource Economics***, Aldershot, Edward Elgar.

Fox, J. T. (2006), Semi and Nonparametric Estimation of Multinomial Discrete Choice Models Using a Subset of Choices, Available at: http://home.uchicago.edu/~fox/Jeremy%20Fox_files/fox_multinomial.pdf

Franses, P. H. and A. L. Montgomery, eds (2002), ***Advances in Econometrics***, Vol. 16, Econometric Models in Marketing, North Holland: Elsevier Science

Freeman, A. M, (1979), "Hedonic prices, property values and measuring environmental benefits: a survey of the issues" ***Scandinavian Journal of Economics***, Vol. 81(1), pp 154-171

Freeman, A. M. (2003), ***The Measurement of Environmental Resource Value: Theory and Methods***, Washington D.C., Resource for the Future, 2nd ed.

Frey, H. (1999), ***Designing the city: Towards a More Sustainable Urban Form***, London, E & FN Spon Press

Fry, J. M. et al (1996), "The stochastic specification of demand share equations: Restricting budget shares to the unit simplex," ***Journal of Econometrics***, Vol. 73(2), pp. 377-385

Fry, T. R. L. and M. N. Harris (1996), "A Monte Carlo study of tests for the independence of irrelevant alternatives property", ***Transportation Research Part B: Methodological***, Vol. 30(1), pp. 19-30,

Fry, T. R. L. and M. N. Harris (1998), "Testing for Independence of Irrelevant Alternatives: Some Empirical Results", ***Sociological Methods Research***, Vol. 26(3), pp. 401-423

FT (2008), Special Report - Nigeria, http://www.ft.com/reports/nigeria2008

Gabriel, S. A. and S. S. Rosenthal (1989), "Household location and race: Estimates of a multinomial logit model", ***Review of Economics and Statistics***, Vol. 71(2), pp. 240-249

Gabriel, S. and G. Painter (2003), "Pathways to Homeownership: An Analysis of the Residential Location and Homeownership Choices of Black Households in Los Angeles", ***Journal of Real Estate Finance and Economics***, Vol. 27(1), pp. 87-109

Garrod, G. and K. Willis (1992), "Valuing the goods characteristics – an application of the hedonic price method to environmental attributes", ***Journal of Environmental Management***, Vol. 34(1), pp. 59-76

Gibson, H. R. (2000), "A 'Characteristics' Approach to Content Regulation and Consumer Choice", The University of Newcastle, Callaghan: Available at: http://www2.hunterlink.net.au/~ddhrg/econ/io2000/gibson_io2000_v5.pdf

Gilbert, A. Ed. (1996), ***The mega-city in Latin America***, United Nations University Press, Tokyo

Goodman, A (1981), "Housing submarkets within urban areas: definitions and evidence", ***Journal of Regional Science***, Vol. 21(2), pp. 175-185

Goodman, A. C. (1998), "Andrew Court and the Invention of Hedonic Price Analysis", ***Journal of Urban Economics***, Vol. 44(2), pp. 291-298

Goodman, A. C. and T. G. Thibodeaub, (1998), "Housing Market Segmentation", ***Journal of Housing Economics***, Vol. 7(2), pp. 121-43

Goodman, A. C. and T. G. Thibodeau (2003), "Housing market segmentation and hedonic prediction accuracy", ***Journal of Housing Economics***, Vol. 12, 3, pp 181-201

Goppinath, D. A. (1995), Modelling heterogeneity in discrete choice processes: Application to travel demand. Unpublished PhD thesis, MIT Cambridge

Grafton, R. Q. et al (2004), ***The Economics of the Environment and Natural Resources***, Oxford: Blackwell.

Gravelle, H. and M. Sutton (2003), "Income related inequalities in self assessed health in Britain: 1979–1995", ***Journal of Epidemiology and Community Health***, Vol. 57(2), pp. 125-129

Greene, W. H. (2003), ***Econometric Analysis***, 5th ed. Prentice Hall, USA.

Grieson, R. E. and R. J. Arnott (1982), "The Supply of Housing," in R. E. Grieson, ed. ***The Urban Economy and Housing***, Lexington, Mass.: D.C. Heath and Co.

Guan, W. (2003), "From the help desk: Bootstrapped standard errors", ***The Stata Journal***, Vol. 3(1), pp. 71–80

Haan, P. (2004), "Discrete choice labor supply: Conditional logit vs. random coefficient models" DIW Berlin, Discussion Paper # 394

Hajivassiliou, V. A., et al (1996), "Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results", ***Journal of Econometrics***, Vol. 72(1–2), pp. 85-134

Hall B. H. and C. Cummins (2005), ***TSP 5.0 User's Guide***, TSP International.

Halvorsen, R. and H. O. Pollakowski (1981), "Choice of functional form for hedonic price equations", ***Journal of Urban Economics***, Vol. 10(1), pp. 37-49

Hamilton, J. M. (2007), "Coastal landscape and the hedonic price of accommodation", ***Ecological Economics***, 62(3-4), pp. 594 – 602

Hanemann, W. M. (1984), "Discrete/Continuous Models of Consumer Demand" ***Econometrica***, Vol. 52(3), pp. 541-561

Hanley, N. et al (2001), "Choice Modelling Approaches: A Superior Alternative for Environmental Valuation", ***Journal of Economic Surveys***, Vol. 15(3), pp. 435-62

Haurin, D. R. and D. Brasington (1996), "School quality and real house prices: Inter- and intra-metropolitan effects", ***Journal of Housing Economics***, Vol. 5, pp. 351-368

Hausman, J. (1978), "Specification Tests in Econometrics", ***Econometrica***, 46, 1251-71.

Hausman, J. and D. McFadden (1984), "Specification Tests for the Multinomial Logit Model", ***Econometrica***, Vol. 52(5), pp. 1219-1240

Hawker, M. J. et al (2007), "Linear Regression Modeling of Rim Area to Discriminate Between Normal and Glaucomatous Optic Nerve Heads: The Bridlington Eye Assessment Project", ***Journal of Glaucoma***, Vol. 16(4), pp. 345-351

Heckman, J. (2008), "Selection bias and Self-selection", In S. N. Durlauf and L. E. Blume (Eds), ***The New Palgrave Dictionary of Economics***, Palgrave Macmillan Online 30 June 2009

<http://www.dictionaryofeconomics.com/article?id=pde2008_S000084> doi:10.1057/9780230226203.1504

Heiss, F. (2002), "Specification(s), of Nested Logit Models," MEA discussion paper series 0216, Mannheim Research Institute for the Economics of Aging, University of Mannheim. Available at: http://www.mea.uni-mannheim.de/mea_neu/pages/files/nopage_pubs/dp16.pdf

Hensher, D. A. and M. Bradley (1993), "Using Stated Response Choice Data to Enrich Revealed Preference Discrete Choice Models", ***Marketing Letters***, Vol. 4(2), pp. 139-151

Hensher, D. A. et al (2005), ***Applied Choice Analysis: A Primer***, Cambridge, Cambridge University Press

Hirt, S. (2007), "The Compact versus the Dispersed City: History of Planning Ideas on Sofia's Urban Form", ***Journal of Planning History***, Vol. 6(2), 138-165

Hoffman, S. D. and G. J. Duncan (1988), "Multinomial and Conditional Logit Discrete-Choice Models in Demography", ***Demography***, Vol. 25(3), pp. 415-427

Hoover, E. M. and F. Giarratani (1999), "An Introduction to Regional Economics", Regional Research Institute, Chapter 7, Available at: http://www.rri.wvu.edu/WebBook/Giarratani/contents.htm

Horowitz, J. L. (1984), "Estimating compensating and equivalent income variations from hedonic price models", ***Economics Letters***, Vol. 14(4), pp 303-308

Houthakker, H. (1951), "Compensated Changes in Qualities and Quantities Consumed", ***Review of Economic Studies***, Vol. 19(3) pp. 155-64

Huh, S. and S. J. Kwak (1997), "The choice of functional form and variables in the hedonic price model in Seoul", ***Urban Studies***, Vol. 34(7), pp. 989-998

Hulten, C. R. (2002), "Price Hedonics: A Critical Review", Paper Prepared for the Federal Reserve Bank of New York conference, Economic Statistics: New Needs for the 21st Century http://www.bsos.umd.edu/econ/hulten/PriceHedonics.pdf

Ichoku, H. E. and Leibbrandt, M. (2003), "Demand for health care services in Nigeria: a multivariate Nested logit model", ***African Development Review***, Vol. 15(2-3), pp. 396-424

Ioannides, Y. and Kan, K. (1996), Structural estimation of residential mobility and housing tenure choice, ***Journal of Regional Science***, Vol. 36(3), pp. 335–363

Janssen, C. et al (2001), "Robust estimation of hedonic models of price and income for investment property", ***Journal of Property Investment & Finance***, Vol. 19(4), pp. 342-360

Johansson, P. (1987), ***The Economic Theory and Environmental Benefits***. Cambridge: Cambridge University Press. pp. 126-146

Jones, C., at al. (2002), "Examining the persistence of housing submarket price differences", Paper presented at the Pacific Rim Real Estate Society Conference, Christchurch, New Zealand

Available at: http://business.unisa.edu.au/prres/Proceedings/Proceedings2002/Jones_Examining_the_persistence_of_housing_submarket_price_differences.pdf

Jones, C., at al. (2003), "Structural Change in a Local Urban Housing Market", ***Environment and Planning A***, Vol. 35, pp. 1315-1326

Jones, C., at al. (2004), "Intra-urban migration and housing submarkets: Theory and evidence", ***Housing Studies***, Vol. 19(2), pp. 269-283

Jud, G. D. and J. M. Watts (1981), "Schools and Housing Values," ***Land Economics***, Vol. 57(4), pp. 459-70

Jud, G. D. and D. T. Winkler (2006), "The Announcement Effect of an Airport Expansion on Housing Prices", ***Journal of Real Estate Finance and Economics***, Vol. 33(2), pp. 91-103

Kan, K. (2000), "Dynamic modelling of housing tenure choice", ***Journal of Urban Economics***, Vol. 48, pp. 46–69

Kanemoto, Y. (1988), "Hedonic Prices and the Benefits of Public Projects", ***Econometrica***, Vol. 56(4), pp. 981-989

Kaoru, Y. et al (1995), "Using Random Utility Models to Estimate the Recreational Value of Estuarine Resources", ***American Journal of Agricultural Economics***, Vol. 77(1), pp 141-151

Kim et al (2008), "Modeling Micro-Spatial Employment Location Patterns: A Comparison of Count and Choice Approaches", ***Geographical Analysis***, Vol. 40, pp. 123-151

King, A. T. (1976), "The Demand for Housing: A Lancastrian Approach" ***Southern Economic Journal***, Vol. 43(2), pp. 1077-1087

King, D. M. and M. Mazzotta (undated), "Essentials of Ecosystem Valuation: **Methods, Section 3:** Hedonic Pricing Method", US Department of Agriculture, Natural Resources Conservation Service and National Oceanographic and Atmospheric Administration Available at: http://www.ecosystemvaluation.org/hedonic_pricing.htm#use

Kornstad, T. and T. O. Thoresen (2007), "A discrete choice model for labor supply and childcare", ***Journal of Population Economics***, Vol. 20(4), pp. 781-803

Kraus, M. (2006), "Monocentric Cities", in R. Arnott, and D. McMillen (eds), ***A Companion to Urban Economics***, Blackwell, Malden MA.

Krout, J. A. and E. Wethington (eds) (2003), ***Residential Choices and Experiences of Older Adults: Pathways to Life Quality***, Springer, New York

Kuminoff, N. V. and J. C. Pope (2009), Capitalization and Welfare Measurement in the Hedonic Model, available at: www.aaec.vt.edu/aaec/working%20papers/2009_01.pdf

Lake, I. R, et al (2000), "Using GIS and large-scale digital data to implement hedonic pricing studies" ***International Journal of Geographical Information Science***, Vol. 14(6), pp. 521 - 541

Lancaster, K. J. (1966), "A New Approach to Consumer Theory" ***Journal of Political Economics***, Vol. 74(2), pp.132-157

Lancaster, K. J. (1991), ***Modern Consumer Theory***, Edward Elgar Publishing, Aldershort

Lancsara, E. and E. Savage (2004), "Deriving welfare measures from discrete choice experiments: Inconsistency between current methods and random utility and welfare theory", ***Health Economics***, Vol. 13(2), pp. 901–907

Labeaga, J. M. et al. (2005), "Discrete choice models of labour supply, behavioural microsimulation and the Spanish tax reforms". Madrid, FEDEA. Available at: http://www.pse.ens.fr/document/wp200513.pdf

Leandro, G. (2005), ***Meta-analysis in Medical Research: The handbook for the understanding and practice of meta-analysis***, Oxford, Blackwell-BMJ Books

Lees, L. et al (2008), ***Gentrification***, Routledge: New York.

Leggett, C. G. and N. E. Bockstael (2000), "Evidence of the effects of water quality on residential land prices", ***Journal of Environmental Economics and Management***, Vol. 39(2), pp.121-144

Leu, R. E. and M. Schellhorn (2006), "The Evolution of Income-related Health Inequalities in Switzerland Over Time", ***CESifo Economic Studies***, Vol. 52(4), pp. 666-690

Li, L. (2003), "Linear Regression Analysis with Observations Subject to Interval Censoring", in H. Zhang and J. Huang (eds) Development of Modern Statistics and Related Topics: Singapore, World Scientific Publishing, pp. 236-245

Lind, H. and A. Hellström (2003), "Gentrification – an overview of the literature", Report No 17, The Division of Building and Real Estate Economics Stockholm, Working Paper TRITA-BFE-WP38, Available at: http://www.infra.kth.se/BYFA/publikationer/engelskaUppsatserOchRapporter/38.pdf.

Linneman, P. (1982), "Hedonic Prices and Residential Location," in Diamond, D. B. Jr. and G. S. Tolley (eds.), ***The Economics of Urban Amenities***, New York: Academic Press.

Lipscomb, C. A. (2003), "The Vickrey Auction Motivation of a Spatial Hedonics Model That Identifies Neighborhood Effects" Available at: http://www.econ.uregina.ca/research/seminars/lipscomb.pdf

Littell, J. H. et al (2008), ***Systematic reviews and meta-analysis***, New York, Oxford University Press

Long, J. S. and J. Freese (2003), ***Regression Models for Categorical Dependent Variables Using STATA***. College Station, STATA Press

Louviere, J. et al. (2005), "Recent Progress on Endogeneity in Choice Modeling", ***Marketing Letters***, Vol. 16(3-4), pp. 255-265

Lutz, M. (2004), "Pricing in Segmented Markets, Arbitrage Barriers, and the Law of One Price: Evidence from the European Car Market", ***Review of International Economics***, Vol. 12(3), pp. 456–475

Maconachie, R. (2007), ***Urban Growth and Land Degradation in Developing Cities: Change and Challenges in Kano, Nigeria***. Ashgate, Aldershot

Maddala, G. S., (1983), ***Limited-dependent and Qualitative Variables in Econometrics***, Cambridge University Press, New York, pp. 22-27

Malpezzi, S. (1999), Economic Analysis of Housing Markets in Developing and Transition Economies, in E. S. Mills, and P. Cheshire (eds), ***Handbook of Regional and Urban Economics***, Elsevier Science, North Holland, pp.1791-1864

Manski, C. and D. McFadden (eds) (1981), ***Structural Analysis of Discrete Data with Econometric Applications***, MIT Press, Cambridge

Manski, C. and D. McFadden (1981), "Alternative estimates and sample designs for discrete choice analysis", in ***Structural Analysis of Discrete Data with Econometric Applications***, C. Manski and D. McFadden (eds). Cambridge: MIT Press.

Manski, C. and S. Lerman (1977), "The Estimation of Choice Probabilities From Choice Based Samples," ***Econometrica***, Vol. 45(8), pp. 1977–1988.

Mazzanti, M. (2003), "Discrete choice models and valuation experiments" ***Journal of Economic Studies,*** Vol. 30(6), pp. 584-604

McConnel, K. E. and T. T. Phipps (1987), "Identification of preference parameters in hedonic models: Consumer demands with non-linear budgets", ***Journal of Urban Economics***, Vol. 22(1), pp. 35-52

McDonald, J. F. and D. P. McMillen (2007), ***Urban Economics and Real Estate: Theory and Policy***, Blackwell, Malden MA, pp. 106-136

McFadden, D. (1973), "Conditional Logit Analysis of Qualitative Choice Behavior", in: Zarembka, P. (ed***.), Frontiers in Econometrics***, Academic Press.

McFadden, D. (1978), Modelling the choice of residential location, in A. Karlquist et al. (ed.), ***Spatial interaction theory and residential location***, North-Holland, Amsterdam, pp. 75–96.

McFadden, D. (1981), "Econometric models of probabilistic choice", in Manski, C. and D. McFadden (Eds.), ***Structural analysis of discrete data with econometric application***, MIT Press, Cambridge, Massachusetts.

McFadden, D. (1984), "Econometric Analysis of Qualitative Response Models," in Griliches, Z. and M. Intrilligator (Eds.), ***Handbook of Econometrics***, Vol. II, Elsevier: Amsterdam

McFadden, D. (1987), "Regression-Based Specification Tests for the Multinomial Logit Model", ***Journal of Econometrics***, Vol. 34(1-2), pp. 63-82.

McFadden, D. (1997), "Modelling the Choice of Residential Location", in J. Quigley (ed), ***The Economics of Housing***, Vol. I, London: Edward Elgar, pp. 531-552

McFadden, D. L. (2000), Nobel Prize Lecture, Available at: http://nobelprize.org/nobel_prizes/economics/laureates/2000/mcfadden-lecture.pdf

McFadden, D. (2003), "Discrete Response Models" Available at, http://www.econ.berkeley.edu/~mcfadden/e240b_sp03/Chap21OH.pdf

McFadden, D. et al. (1981), "An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model", ***Transportation Research Board Record***, Vol. 637, pp. 39-46.

McFadden, D. and Train, K. (1997), "Mixed multinomial logit models for discrete response", Technical report, University of California, Berkeley, Ca.

McMillan, M. L. et al. (1980), "An Extension of the Hedonic Approach for Estimating the Value of Quiet," ***Land Economics***, Vol. 56, pp. 315-328

Megbolugbe, I. F. (1989), "A hedonic index model: The housing market of Jos, Nigeria", ***Urban Studies***, Vol. 26(5), pp. 486-494

Megbolugbe, I. F. et al (1991), "The Economic Theory of Housing Demand: A Critical Review", ***Journal of Real Estate Research***, Vol. 6(3), pp. 381-393

Michel, B. (1997), "Discrete choice models", Available at: http://rosowww.epfl.ch/mbi/papers/discretechoice/paper.html

Michaels, R. G. and V. K. Smith (1990), "Market segmentation and valuing amenities with hedonic models: The case of hazardous waste sites", ***Journal of Urban Economics***, Vol. 28(2), pp. 223-242

Mieszkowski, P. and A. Saper (1978), "An estimate of the effects of airport noise on property values", ***Journal of Urban Economics***, Vol. 5(5), pp. 425-440

Mohammadian, A. and P. S. Kanaroglou, (2003), "Applications of spatial multinomial logit model to transportation planning", The10th International Conference on Travel Behaviour Research, Available at: http://www.ivt.baug.ethz.ch/allgemein/pdf/mohammadian.pdf

Moulton, B. R. (2001), "The Expanding Role of Hedonic Methods in the Official Statistics of The United States", OECD, Available at: http://www.oecd.org/dataoecd/34/55/1907790.doc.

Munizaga, M. A. and R. Alvarez-Daziano (2005), "Testing Mixed Logit and Probit by simulation", ***Transportation Research Record***, Vol. 1921, pp. 53-62

Murray, M. P. (1983), “Mythical demands and supplies for proper estimation of Rosen’s hedonic price model”, ***Journal of Urban Economics***, Vol. 14(3), pp 327-337

Murty, M. N. et al (2003), “Choosing Appropriate Functional Forms of Hedonic Price Equations in Hedonic Property Value Models: Measuring Benefits from Reduced Air Pollution in the Cities of Delhi and Kolkata in India”, Institute of Economic Growth, Delhi University Enclave, Available at: http://eaere2004.bkae.hu/download/paper/murty-gulati-banerjeepaper.doc.

Musgrave, R. A. (1969), ***Fiscal Systems***, New Haven and London: Yale University Press

Muth, R. F. (1960), “The demand for non-farm housing”, in A. C. Haberger (ed.), The ***Demand for Durable Goods***. Chicago: University of Chicago Press. pp. 29-96.

Nechyba, T. J. and R. P. Strauss (1997), “Community choice and local public services: a discrete choice approach”, NBER Working Paper 5966, Available at: http://www.nber.org/papers/w5966.pdf

Nelson, J. P. (2003), “Meta-Analysis of Airport Noise and Hedonic Property Values: Problems and Prospects”, Available at: http://econ.la.psu.edu/papers/nelson_metanoise31.pdf

Nevo, A. (1999), “A research assistant's guide to random coefficients discrete choice models of demand”, NBER Technical Working Paper 221 Available at: http://www.nber.org/papers/t0221.pdf

Nevo, A. (2003), “Using Weights to Adjust for Sample Selection When Auxiliary Information is Available”, ***Journal of Business & Economic Statistics***, Vol. 21(1), pp. 43-52

Newsom, J. T. (2004), “Distinguishing between Random and Fixed: Variables, Effects, and Coefficients” Available at: http://www.ioa.pdx.edu/newsom/mlrclass/ho_randfixd.doc.

Nigeria Vanguard (2009), 60m Nigerians now own power generators – MAN, Nigerian Daily Newspaper, Monday, 26 January 2009. Available at: http://www.vanguardngr.com/content/view/27253/41/

NIST/SEMATECH (no date), e-Handbook of Statistical Methods, Available at: http://www.itl.nist.gov/div898/handbook/pmd/section1/pmd144.htm

North, J. H. and C. C. Griffin (1993), “Water Source as a Housing Characteristic Hedonic Property Valuation and Willingness to Pay for Water”, ***Water Resource Research***, Vol. 29(7), pp. 1923-1929

O'Garra, T. and S. Mourato (2007), “Public Preferences for Hydrogen Buses: Comparing Interval Data, OLS and Quantile Regression Approaches”, ***Environmental and Resource Economics***, Vol. 36(4), pp. 389-411

Oliver, M. L. and T. M. Shapiro (2006) ***Black Wealth/White Wealth: A New Perspective on Racial Inequality***, Tenth Anniversary Edition. New York: Routledge, Olsen, E. O. (1987), "The demand and supply of housing services: a critical review of the empirical literature", in E. S. Mills (Ed), ***The Handbook of Regional and Urban Economics***, Elsevier, New York, NY

Oxley, M. (2004), ***Economics, Planning and Housing***, Palgrave Macmillan, Houndmills, Basingstoke

Pagliara, F., J. Preston, and D. Simmonds (eds.) (2010), ***Residential Location Choice: Models and Applications***, Berlin, Springer

Palmquist, R. B. (1984), "Estimating the demand for characteristics of housing", ***The Review of Economics and Statistics***, Vol. 66(3), pp. 394-404

Palmquist, R. B. (1991), "Hedonic Methods," in Braden, J. B. and C. D. Kolstad, (eds.), ***Measuring the Demand for Environmental Quality***, North Holland

Palmquist, R. B. (1999), "Hedonic Models", in van den Bergh, J. C.J.M. (ed), ***Handbook of Environmental and Resource Economics***, Aldershot: Edward Elgar.

Pandit, R. and D. N. Laband (2007), "Spatial autocorrelation in country-level models of species imperilment", ***Ecological Economics***, Vol. 60(3), pp. 526-532

Pearce D. et al (2002), "Economic Valuation with Stated Preference Techniques Summary Guide", London, Department for Transport, Local Government and the Regions.

Pearson, K (1904), "Report on certain enteric fever inoculation statistics", ***British Medical Journal***, Vol. 3, pp. 1243-1246

Pennington, G. et al. (1990), "Aircraft Noise and Residential Property Values Adjacent to Manchester International Airport", ***Journal of Transport Economics and Policy***, Vol. 25(1), pp. 49-59

Piekkola, H. and K. Snellman (Eds.) (2004), ***Collective Bargaining and Wage Formation Performance and Challenges***, Physica-Verlag Gmbh & Co pp. 108

Porojan, A. (2001), "Trade Flows and Spatial Effects - The Gravity Model Revisited", ***Open Economies Review***, 12(3), pp. 265–280

Quigley, J. M, (1979), "What have we learned about urban housing markets?", in Mieszkowski, P. and M. Straszheim Eds. ***Current Issues in Urban Economics***, Johns Hopkins University Press, Baltimore, Maryland. pp 391 – 429

Quigley, J. M. (1982), "Non-Linear budget constraints and consumer demand: An application to public programs for residential housing", ***Journal of Urban Economics***, Vol. 12(2), pp. 177-201

Ramaswamy, V. (1997), "Evolutionary preference segmentation with panel survey data: An application to new products", ***International Journal of Research in Marketing***, Vol. 14(1), pp. 57-80

Ramos, F. R. (1999), "Estimating the benefit and the environmental quality preference in multinomial probit models using a simulation approach", ***Envirometrics***, Vol. 10(3), pp.309-319

Rapaport, C. (1997), "Housing demand and community choice: an empirical analysis", ***Journal of Urban Economics***, Vol. 42, pp. 243–260

Rasmussen, D. W. and T. W. Zuehlke (1990), "On the choice of functional form for hedonic price functions", ***Applied Economics***, Vol. 22(4), pp. 431-438

Ratchford, B. T. (1975), "The New Economic Theory of Consumer Behavior: An Interpretive Essay", ***Journal of Consumer Research***, Vol. 2(2), pp. 65-75

Ridker, R. G. and J. A. Henning (1967), "The determinants of residential property values with special reference to air pollution", ***The Review of Economics and Statistics***, Vol. 49(2), pp. 246-257

Riera, P. and A. Mhawej (2005), A Fixed-Effects Hedonic Price Model for Statistical Value of Live Estimations, Universidad Autonoma De Barcelona, Available at: http://volcano.uab.es/priera/Gines/Hedonic.doc.

Roback, J. (1988), "Wages, Rents, and Amenities: Differences among Workers and Regions" ***Economic Inquiry***, 26: pp 23-42.

Robinson, R. (1979) ***Housing Economics and Public Policy: Studies in Planning, Macmillan***, London

Rosen, S. (1974), "Hedonic prices and implicit markets: Product differentiation in pure competition", ***Journal of Political Economy***, 82(1), pp 34-55.

Salensminde, K. (2002), "The Impact of Choice Inconsistencies in Stated Choice Studies", ***Environmental and Resource Economics***, Vol. 23, pp. 403-20

Salomon, I. and M. Ben-Akiva (1983), "The use of the life-style concept in travel demand models", ***Environment and Planning A***, Vol. 15(5), pp. 623-638

Scotchmer, S. (1986), "The Short Run and Long Run Benefits of Environmental Improvement", ***Journal of Public Economics***, Vol. 30(1), pp. 61-81

Schulze, R. (2004), ***Meta-Analysis: A Comparison of Approaches***, Cambridge MA, Hogrefe and Huber

Sen, A. K. (1988), "Freedom of choice: concept and content", ***European Economic Review***, Vol. 32(2-3), pp. 269-294

Shen, J. (2008), "Understanding the determinants of consumers' willingness to pay for eco-labeled products: An empirical analysis of the China Environmental Label", OSIPP Discussion Paper, Available at: http://www.osipp.osaka-u.ac.jp/archives/DP/2008/DP2008E001.pdf

Sharpe, D. (1997), “Of Apples and Oranges, File drawers and garbage: Why validity issues in Meta-Analysis will not go away”, ***Clinical Psychology Review***, Vol. 17(8), pp. 881-901

Sheppard, S. (1999), “Hedonic Analysis of Housing Markets”, in E. S. Mills and P. Cheshire Eds. ***Handbook of Regional and Urban Economics***, North Holland, Elsevier Science, pp. 1595-1635

Skaburskis, A. (1997), “Gender Differences in Housing Demand”, ***Urban Studies***, Vol. 34(2), pp. 275-320

Skaburskis, A. (1999), “Modelling the choice of tenure and building type”, ***Urban Studies***, Vol. 36(13), pp. 2199–2215

Slater, T. (2002), “What is Gentrification?”, Available at: http://members.lycos.co.uk/gentrification/whatisgent.html

Small K. A. and C. Hsiao (1985), “Multinomial Logit Specification Tests”, ***International Economic Review***, Vol. 26(3), pp. 619-627

Smith, B. (1976), “The Supply of Urban Housing”, ***Quarterly Journal of Economics***, Vol. 90(3), pp. 389-405

Sonnier, G. et al (2007), “Heterogeneity distributions of willingness-to-pay in choice models”, ***Quantitative Marketing and Economics***, Vol. 5(3), pp. 313–331

Stanley, T. D. (2001), “Wheat From Chaff: Meta-Analysis As Quantitative Literature Review”, ***Journal of Economic Perspectives***, Vol. 15(3), pp. 131-150

STATA (2007), “STATA Data Analysis Examples: Interval Regression”, UCLA Academic Technology Services, Available at: http://www.ats.ucla.edu/stat/stata/dae/intreg.htm

Stewart, M. B. (1983), “On Least Squares Estimation when the Dependent Variable is Grouped”, ***The Review of Economic Studies***, Vol. 50(4), pp. 737-753

Straszheim M. R., (1975), ***An Econometric Analysis of the Urban Housing Market***, Urban and Regional Studies No. 2, Columbia University Press and National Bureau of Economic Research, Cambridge, MA. Available at http://www.nber.org/books/stra75-1

Sugden, R. (2000), “Team preferences”, ***Economics and Philosophy***, Vol. 16(2), pp. 175-204

Sutton, A. J. et al (2000), ***Methods for Meta-analysis in Medical Research***, Chichester, Wiley-Blackwell

Swait, et al (2002), “Context dependence and aggregation in disaggregated analysis”, ***Marketing Letters***, Vol. 13(3), pp. 195-205

Taylor, L. O. (2003), “Hedonic Price Method” in P. A. Champ, T. Brown, and K Boyle, eds, ***A Primer on the Economic Valuation of the Environment***, Kluwer: Dordrecht.

Thaler, D. (1978), “A note on the value of crime control: Evidence from the property market”, ***Journal of Urban Economics***, Vol. 5(1), pp. 137-145

The Economist (2002), Cities - “The doughnut effect”: The suburbs are paying for Britain's success in reviving its inner cities, Jan 19th, Available at: http://economist.com/displaystory.cfm?story_id=940671

Tipplea, G. et al (1999), “Housing supply in Ghana: a study of Accra, Kumasi and Berekum”, ***Progress in Planning***, Vol. 51(4), pp. 255-324

Tomkins, J., et al (1998), “Noise versus access: The impact of an airport in an urban property market”, ***Urban Studies***, Vol. 35(2), pp. 243-258

Train, K. E. (1999), “Halton Sequences for Mixed Logit” ELSA Berkeley Working Paper, Available at: http://elsa.berkeley.edu/wp/train0899.pdf

Train, K. E. (2003), ***Discrete Choice Methods with Simulation***. Cambridge: Cambridge University Press.

Travisi, C. M., and P. Nijkamp (2004), “Willingness to Pay for Agricultural Environmental Safety: Evidence from a Survey of Milan, Italy, Residents”, Fondazione Eni Enrico Mattei SIEV Working Paper Series, 100

Trijonis, J. et al. (1985), Air Quality Benefit Analysis for Los Angeles and San Francisco Based on Housing Values and Visibility, Available at: ftp://ftp.arb.ca.gov/carbis/research/apr/past/a2-088-32_exsum.pdf

Triplett, J. E. (2004), Handbook on hedonic indexes and quality adjustments in price indexes: Special Application to Information Technology Products, Directorate for Science, Technology and Industry. OECD, STI Working Paper Number 9, OECD Available at: http://www.oecd.org/dataoecd/37/31/33789552.pdf

Tu, Y. and J. Goldfinch (1996) A Two-stage Housing Choice Forecasting Model, ***Urban Studies***, Vol. 33(3), pp. 517-37

Turnbull, G. K. et al (1991), “Uncertain Income and Housing Price and Location Choice”, ***Journal of Regional science***, Vol. 31(4), pp. 417-33

UN-HABITAT (2010), ***The State of African Cities 2010, Governance, Inequality and Urban Land Markets***: United Nations, Nairobi

van Doorslaer, E. and A. M. Jones (2003), “Inequalities in self-reported health: validation of a new approach to measurement”, ***Journal of Health Economics***, Vol. 22(1), pp. 61-87

van Praag, B. M. S. and B. E. Baarsma (2005), “Using Happiness Surveys to Value Intangibles: The Case of Airport Noise”, ***The Economic Journal***, 115(500), pp. 224–246

van Soest, A. (1995), “Structural Models of Family Labor Supply: A Discrete Choice Approach”, ***Journal of Human Resources***, Vol. 30(1), 63-88.

Vandell, K. D. (1995), “Market Factors Affecting Spatial Heterogeneity among Urban Neighbourhoods”, ***Housing Policy Debate***, Vol. 6(1), pp. 103-139

Vartia, L. (2005), “Establishing and Closing Down Plants -Assessing the Effects of Firms’ Financial Status”, 32nd Conference of the European Association for Research in Industrial Economics

Veall, M. R. and K. F. Zimmermann (1996), “Pseudo-R2 Measures for Some Common Limited Dependent Variable Models,” ***Journal of Economic Surveys***, Vol. 10(3), pp. 241-59

Venti, S. F. and D. A. Wise (1984), “Moving and housing expenditure: transactions cost and disequilibrium,” ***Journal of Public Economics***, Vol. 23(1-2), pp. 207–243. Also available http://papers.nber.org/papers/w1012.v5.pdf

Walden, M. L. (1990), “Magnet schools and the differential impact of quality on residential property values”, ***Journal of Real Estate Research***, Vol. 5(2), pp. 221-230

Walford, M. (2001), “Dangers of the ‘doughnut effect’”, The Baltic Times, Available at: http://www.baltictimes.com/news/articles/4065/

Walker, J. L. and J. Li (2007), “Latent lifestyle preferences and household location decisions”, ***Journal of Geographical Systems***, Vol. 9(1), pp. 77–101

Wallace, A. (2004), Understanding Local Housing Markets? The need for a complementary institutional approach, Centre for Housing Policy University of York, Available at: http://www.york.ac.uk/inst/chp/hsa/papers/spring%2004/Allison%20Wallace.pdf

Washington, S. et al **(2001), *Scientific Approaches to Transportation Research: Volumes 1 and 2***, National Cooperative Highway Research Program Transportation Research Board of the National Academies, Washington, DC

Watkins, C. (2001), “The definition and identification of housing submarkets”, ***Environment and Planning A***, Vol. 33, pp. 2235-2253

Waugh, D. (2003), ***The New Wider World***, Nelson-Thornes, Cheltenham

Weinberg, D. et al (1981), “Intraurban residential mobility: The role of transaction costs, market imperfections and household disequilibrium”, ***Journal of Urban Economics***, Vol. 9(3), pp. 332-348

Wen, H. et al. (2005), “Hedonic price analysis of urban housing: An empirical research on Hangzhou, China”, ***Science*** (Journal of Zhejiang University), Vol. 6(8), pp. 907-914

Whitehead, C. M. E. (1999), Urban Housing Markets: Theory and Policy in E. S. Mills and P. Cheshire (Eds) ***Handbook of Regional and Urban Economics***, Elsevier Science, North Holland, pp. 1559-94

WHO/UNICEF (2005), *Water for Life*, Available at: http://www.who.int/water_sanitation_health/waterforlife.pdf

Williamson, J. M. Hale, W. Thurston, and M. T. Heberling (2008), "Valuing acid mine drainage remediation in West Virginia: a hedonic modelling approach" ***The Annals of Regional Science,*** Vol. 42(4), pp. 987-999

Wong, G. K. M. (2002), "A Conceptual Model of Household's Decision Making Process: The Economic Perspective", ***Review of Urban & Regional Development Studies***, Vol. 14(3), pp. 217-34

Wu, J. (2006), "Environmental amenities, urban sprawl, and community characteristics", ***Journal of Environmental Economics and Management***, Vol. 52(2), pp. 527-547

Wyly, E. K. and D. J. Hammel (2004), "Gentrification, segregation, and discrimination in the American urban system", ***Environment and Planning A***, Vol. 36(7), pp. 1215 - 1241

Yates, J. and Mackay, D. F. (2006), "Discrete choice modelling of urban housing markets: A critical review and an application", ***Urban Studies***, Vol. 43(3), pp. 559-581

Yusuf, A. A. and P. Koundouri (2005), "Willingness to pay for water and location bias in hedonic price analysis: evidence from the Indonesian housing market", ***Environment and Development Economics***, Vol. 10(06), pp. 821–836

Yoo, S. H. (2001), "A robust estimation of hedonic price models: least absolute deviations estimation", ***Applied Economics Letters***, Vol. 8(1), pp. 55-58

Yu, D. (2004), "Modeling Housing Market Dynamics in the City of Milwaukee: A Geographically Weighted Regression Approach", Department of Geography, University of Wisconsin, Milwaukee Available at: http://www.ucgis.org/UCGISFall2004/studentpapers/files/danlinyu.pdf