Research article

Volatility in the Housing Market: Evidence on Risk and Return in the London Sub-market

Steve Cook[1,*] and Duncan Watson[2]

[1] Department of Finance, Swansea University, Bay Campus, Swansea SA1 8EN, UK
[2] School of Economics, University of East Anglia, Norwich NR4 7TJ, UK

* **Correspondence:** Email: s.cook@swan.ac.uk.

Abstract: The impact of volatility in housing market analysis is reconsidered via examinaton of the risk-return relationship in the London housing market is examined. In addition to providing the first empirical results for the relationship between risk (as measured by volatility) and returns for this submarket, the analysis offers a more general message to empiricists via a detailed and explicit evaluation of the impact of empirical design decisions upon inferences. In particular, the negative risk-return relationship discussed frequently in the housing market literature is examined and shown to depend upon typically overlooked decisions concerning components of the empirical framework from which statistical inferences are drawn.

Keywords: regional housing markets; risk analysis; volatility; rolling samples

1. Introduction

The importance of the housing market to the wider economy has been well documented in empirical research with a number of studies noting, *inter alia*, its substantive contribution to private sector wealth, dominance over the stock market in determining household consumption decisions, central role within the macroeconomy and close relationship with economic fundamentals (Brueckner, 1997; Holly and Jones, 1997; Gallin, 2006; Goetzmann, 1993; Goodhart and Hoffman, 2007; Bayer et al., 2010; Costello et al., 2011; Case et al., 2013; Han, 2013). As a consequence, the behaviour of housing markets and the properties of house prices have received much attention. In recent years, one element of this research has considered whether housing displays the risk-return characteristics predicted for other financial assets by standard finance theory. A feature of this literature is the repeated discussion of the existence of a counterintuitive negative risk-return relationship within housing markets (Dolde and Tirtiroglu, 1997; Morley and Thomas, 2011; Han, 2013; Lin and Fuerst, 2014). Clearly this runs contrary to the positive risk-return relationship depicted

by theoretical finance where higher risk is compensated by higher returns.

The contribution of the present paper to the existing literature is twofold. First, previous research is extended by producing the first findings on this issue for the highly topical London housing submarket. Second, an explicit investigation is provided of the extent to which decisions regarding empirical design impact upon inferences concerning risk and return in housing markets. More precisely, the present research considers the influence of the components of empirical design upon the subsequent inferences drawn by investigators when examining the relationship between risk, as measured by volatility, and returns. Consequently, it is examined how the significant negative relationship which has featured so prominently in the literature is dependent upon stances taken with regard to decisions on variable definition, sample selection, optimisation methods, dynamic specification, regional disaggregation and modelling techniques. Interestingly the results of the current analysis show that while negative risk-return relationships are observed for an empirical design with very specific options selected for the sample, modelling technique and approach to dynamic specification employed, a more thorough analysis produces mixed findings.

To achieve its objectives, this paper will proceed as follows. In Section 2 a selected review of the literature on the analysis of volatility and risk in housing markets is presented. The various components of the empirical design employed to examine risk and return in the present analysis are provided in Section 3, with the empirical results from this analysis presented in Section 4. Section 5 provides some concluding remarks.

2. Literature Review

In this section a selected review of the literature in relation to the analysis of volatility and risk in housing markets is provided. Considering these two issues in turn, house price volatility has received much attention in the empirical literature for a number of years, as illustrated by, Foster and Van Order (1984), Crawford and Rosenblatt (1995), Dolde and Tirtiroglu (1997), Crawford and Fratantoni (2003), Miller and Peng (2006), Miles (2008, 2011), Miller and Pandher (2008), Morley and Thomas (2011) and Barros et al. (2015). However, while a wealth of empirical studies have emerged examining volatility, the development of theoretical explanations for its presence have received less attention with the inertia-based explanation of Case and Shiller (1988, 1989, 1990) and Wheaton's (2015) decomposition of volatility into demand- and supply-side factors being notable exceptions to this. Inspection of the empirical literature examining volatility shows the use of the autoregressive conditional heteroskedasticity (ARCH) model and its various extensions to feature prominently. Interestingly, it can be seen that their use has produced conflicting results. As an illustration of this, while Dolde and Tirtiroglu (1997), Crawford and Fratantoni (2003) and Miles (2008) present evidence of volatility in house prices for the USA, the results of Miller and Peng (2006) are less supportive. Similarly, the results of Miles (2011) suggest an element of volatility in UK house prices with just over half of the regions considered providing evidence of significant volatility, the evidence provided by Lin and Fuerst (2014) is more compelling for Canada.

With regard to 'risk', a number of differing perspectives have been adopted in the literature to consider its both is presence and its impact. Portfolio risk management provides the motivation for both Huang et al. (2016) and Zhou and Gao (2012). While Huang et al. (2016) consider the management of risk via diversification in real estate investment trust (REIT)-housing and

stocks-housing portfolios, Zhou and Gao (2012) employ copula-based methods to explore risk management in real estate securities. A further alternative consideration of the role of risk in housing markets is provided by Tsang et al. (2016) where stochastic dominance analysis is employed to examine the impact of risk on housing purchase decisions in the Hong Kong market. Additional research more closely related to the present analysis is provided by Domian et al. (2015) where two specifications of CAPM are employed to consider the risk-return relationship across Metropolitan Statistical Areas (MSAs) in the USA using the Case-Shiller house price index. Extending the analysis to include leverage and liquidity risks, Domian et al. (2015) provide evidence of geographical variation across MSAs with counter-cyclical behaviour noted in some areas such as cities in Californian contasting with the higher levels of risk noted in, for example, New York. However, the literature on the risk-return relationship in housing is dominant by the use of ARCH-based models which are employed in the current analysis. Specific examples of this include Dolde and Tirtiroglu (1997), Morley and Thomas (2011, 2016), Lin and Fuerst (2014) and Lee (2017). As noted by Han (2013) and in the studies above, a negative relationship between risk and returns has been noted in this literature. This variation in the sign of the risk-return relationship and its possible dependence on the nature of the approach adopted towards modelling provides the motivation for the current study. To explore this issue and provide an extension to the existing literature, the current analysis explicitly considers the impact of variations in the empirical design of econometric framework upon the results obtained with regard to the risk-return relationship.

3. Risk-return and the Housing Market

In this section the structure, or components, of the empirical analysis are outlined. The material is structured via consideration of the decisions made concerning variable definition, sample selection, modelling techniques and dynamic specification when undertaking an analysis of risk and return. Typically these decisions and assumptions are implicit (or unrecognised) when performing empirical analysis. In contrast, the present research makes explicit the alternative options available to investigators in the process of moving from an initial hypothesis of interest motivating empirical analysis to the testable framework from which inferences are drawn.

3.1. Data

To consider the nature of risk and return within housing markets, an obvious issue to consider concerns the actual series to be examined. In the present analysis, the London housing market is considered. The London housing market is a highly topical example receiving frequent attention within both academia and the media, and hence an attractive choice. The specific series considered herein are those for the 32 boroughs of London plus the aggregate London series. The house price data considered are monthly, seasonally adjusted observations on average house prices in the over the period January 1995 to December 2015.* While the London submarket has been examined recently by Abbott and De Vita (2012), the data set employed in the present analysis differs from that employed in their

---

*The data set is available from http://landregistry.data.gov.uk/app/hpi. The 32 boroughs are: Barking, Barnet, Bexley, Brent, Bromley, Camden, Croydon, Ealing, Enfield, Greenwich, Hackney, Hammersmith & Fulham, Haringey, Harrow, Havering, Hillingdon, Hounslow, Islington, Kensington and Chelsea, Kingston upon Thames, Lambeth, Lewisham, Merton, Newham, Redbridge, Richmond upon Thames, Southwark, Sutton, Tower Hamlets, Waltham Forest, Wandsworth, City of Westminster. The aggregate series is denoted simply as London.

examination of housing market convergence. First, the aggregate London price series considered here was not considered by Abbott and De Vita (2012). Second, the 32 boroughs of London are considered herein without the addition of the City of London local authority district.[†] Third, in a further departure from the previous study, the house price series are considered in nominal terms as well as in real (inflation adjusted) terms following deflation using the consumer price index. Consideration of both nominal and real data recognises the common use of real data in analysis of risk and return, along with the argument within the economics literature that decision making is often undertaken by consideration of nominal values (Shafir et al., 1997). [‡] Fourth, the sample considered differs from that employed by Abbott and De Vita (2012) as it is observed at a monthly, rather than quarterly frequency, with this higher frequency allowing an improved analysis of volatility. Finally, seasonally adjusted, rather than unadjusted, observations are considered for a sample which starts a year earlier and finishes over 6 years later than that of Abbott and De Vita (2012). As a result, the current analysis considers a sample of 252 observations compared to 54 in Abbott and De Vita (2012). Therefore, despite initially appearing similar, the data in the current analysis differ markedly from those in Abbott and De Vita (2012) due to sample span, number of observations, frequency, seasonal adjustment, aggregation and the nominal/real measurement dichotomy. Importantly, the presence of these options illustrates the relevance of the issue of empirical design discussed later in this paper.

Denoting the natural logarithms of the house price series as p_t, house price returns are calculated as their difference $r_t = \Delta p_t$. [§] The use of changes in house prices to measure returns to housing is an issue that warrants some discussion.[¶] This issue can be illustrated via consideration of the work of Bayer et al. (2010) where the return to housing is given as a combination of house price changes plus rental income. However, while this has a clear justification, the literature is dominated by studies measuring returns as house price changes without the inclusion of rental income. Examples of studies adopting this approach include the works of Dolde and Tirtiroglu (2002), Miller and Peng (2006), Morley and Thomas (2011, 2016). In light of the prevalence of the use of house price changes as a measure of returns, this approach will be adopted in the present study to allow consideration of the returns to purely holding housing as an asset or, alternatively expressed, the returns to owner occupation.[‖]

The use of house price data for London in both disaggregated and aggregate form and in both nominal and real terms reflects clear decisions in the design of the empirical analysis undertaken in the current analysis. These variable-related assumptions or decisions are denoted here as V_i. Similarly, options available concerning decisions on the sample employed can be denoted as S_i. Within the present analysis, the full or maximum sample available is considered as a starting point for the empirical analysis. However, to explore the sample dependence of inferences drawn, rolling samples are considered also. To include a relatively large number of observations in each of the rolling samples examined, the 251 observations of the full effective sample are employed to create an

[†]Abbott and DeVita (2012) include the City of London in their analysis. While this is a local authority district within the Greater London, it is a very different in nature to the boroughs of London. In addition, data for the City of London are not available for the preferred frequency considered herein.

[‡]The CPI series was obtained from the Office of National Statistics (https://www.ons.gov.uk/)

[§]Examination of the order of integration of the returns series (both nominal and real) using the Im *et al*. (2001) panel unit root test resulted in rejection of the null. Hence the returns series are all treated as stationary processes. Further details are available from the authors upon request.

[¶]We are grateful to the editor and an anonymous referee for raising the issue of measuring returns.

[‖]As noted, the measurement of returns has been extended by Bayer et al. (2010) to include rental income. Interestingly while this has a positive impact upon the level of returns, factors such as renovation costs and mortgage payments which have a negative impact are not considered.

additional 72 samples containing 180 observations.** Via the use of these rolling samples, robustness of results across alternative sampling periods is explored.

3.2. Alternative modelling techniques

The prediction of a positive relationship between the returns on an asset and its associated risk is a standard feature of financial theory. However, as noted by Han (2013), a negative relationship between risk and returns has been noted repeatedly in the housing market literature (Dolde and Tirtiroglu, 1997; Morley and Thomas, 2011; Lin and Fuerst, 2014). When considering the examination of risk-return relationships, an obvious and typically employed model to utilise is the GARCH-M specification (Engle et al., 1987). With returns on house prices denoted as r_t, a standard GARCH(1,1)-M model can be expressed as follows:

$$r_t = \mu + \delta\sigma_t + u_t \qquad u_t \sim \left(0, \sigma_t^2\right) \tag{1}$$

$$\sigma_t^2 = \phi_0 + \phi_1 u_{t-1}^2 + \phi_2 \sigma_{t-1}^2 \tag{2}$$

Application of (1)-(2) allows examination of the risk-return relationship via the coefficient δ attached to the conditional standard deviation term (σ_t) in the mean equation. Although this specification is commonly applied in the literature, as has been noted by Scruggs (1998), variations on this model can be considered in terms of the specification of both the mean and variance equations. With regard to the variance equation, an obvious alternative to consider is the exponential GARCH (EGARCH) model of Nelson (1991). In addition to allowing a broader coverage of behaviour to capture potential asymmetric responses, the EGARCH model is attractive to practitioners as it does not require consideration of non-negativity constraints associated with the $\{\phi_1, \phi_2\}$ parameters in the GARCH model of (2) above. The benefits of the EGARCH specification are apparent in its widespread adoption in the literature with Miller and Peng (2006), Lee (2009), Miles (2011), Morley and Thomas (2011, 2016) and Lin and Fuerst (2014) all employing this model. Extension of the GARCH-M model of (1)-(2) to consider an analogous EGARCH(1,1)-M specification results in the model below:

$$r_t = \mu + \delta\sigma_t + u_t \qquad u_t \sim \left(0, \sigma_t^2\right) \tag{3}$$

$$log\left(\sigma_t^2\right) = \gamma_0 + \gamma_1 log\left(\sigma_{t-1}^2\right) + \gamma_2 \left|\frac{u_{t-1}}{\sigma_{t-1}}\right| + \gamma_3 \left(\frac{u_{t-1}}{\sigma_{t-1}}\right) \tag{4}$$

To allow for the heavy, or thick, tails observed in financial series, the above models are estimated using the generalised error distribution (GED) for the error u_t. In addition to allowing for heavy-tailed errors, use of the GED affords flexibility in the analysis conducted via unrestricted estimation of its underlying shape parameter (v) for every model considered. Hence the degree of thickness of the tails of the error distribution can be tailored to the series, with movement from the Normal distribution $(v = 2)$ to heavier tailed processes $(v < 2)$ arising due to the estimation of smaller values for the shape parameter. Following the above discussion, the options available for these model-related components of the empirical design are denoted as M_i.

**The full sample from January 1995 to December 2015 contains 252 observations. Following creation of a first difference to generate returns, this is reduced to 251. The 72 rolling samples of 180 observations are then given as Feb 1995 to Jan 2010, Mar 1995 to Feb 2010 through to Jan 2001 to Dec 2015.

3.3. Dynamic specification

With regard to the specification of the mean equations in (1) and (3), these static expressions can be expanded to include a dynamic structure as noted by Scruggs (1998). Inclusion of an autoregressive component results in the GARCH(1,1)-M specification of (5)-(6):

$$r_t = \mu + \delta\sigma_t + \sum_{i=1}^{p} \lambda_i r_{t-i} + u_t \qquad u_t \sim \left(0, \sigma_t^2\right) \tag{5}$$

$$\sigma_t^2 = \phi_0 + \phi_1 u_{t-1}^2 + \phi_2 \sigma_{t-1}^2 \tag{6}$$

and the EGARCH(1,1)-M model of (7)-(8):

$$r_t = \mu + \delta\sigma_t + \sum_{i=1}^{p} \lambda_i r_{t-i} + u_t \qquad u_t \sim \left(0, \sigma_t^2\right) \tag{7}$$

$$log\left(\sigma_t^2\right) = \gamma_0 + \gamma_1 log\left(\sigma_{t-1}^2\right) + \gamma_2 \left|\frac{u_{t-1}}{\sigma_{t-1}}\right| + \gamma_3 \left(\frac{u_{t-1}}{\sigma_{t-1}}\right) \tag{8}$$

Estimated risk-return coefficients $\left(\widehat{\delta}\right)$ can then be drawn from models employing alternative lag structures, or values of p. Investigators could consider the risk-return coefficients obtained from the static models discussed above $\left(\widehat{\delta_0}\right)$ or the maximum and minimum values $\left(\widehat{\delta_{max}}, \widehat{\delta_{min}}\right)$ observed to be statistically significant across a range of values of p. Alternatively, a more typical approach is to consider coefficients obtained from models employing an optimised value of p, with the Akaike Information Criterion (AIC) frequently utilised as the means of selecting the optimum lag length. These coefficients are denoted herein as $\widehat{\delta}_{AIC}$ with the optimum lag length defined as that generating the minimum value of the AIC, with the AIC specified as:

$$AIC = -\frac{2}{\ell} + \frac{2p}{T} \tag{9}$$

where ℓ and T denote the log-likelihood and sample size respectively. In light of the monthly frequency of the data considered in the present analysis, lags from a maximum value of $p = 12$ down to a minimum of no lags ($p = 0$) are considered herein. Continuing the above notation, these options or alternative possibilities concerning dynamic specification can be denoted as D_i.

3.4. Empirical design components

When considering the risk-return relationship via the use of (E)GARCH-M models, the parameter of interest is the risk coefficient in the relevant mean equation. The two issues of importance concerning this coefficient are its significance so as to determine whether a significant relationship exists and, if so, whether it is positive, as predicted by theory, or whether it is negative. However, as the above discussion in this section has made explicit, a series of decisions are required to structure the subsequent empirical analysis to permit testing of this hypothesis and inferences to be drawn. This movement from the initial focus upon the risk-return coefficient through to inference upon its nature is summarised in Figure One below. As this illustration depicts, the hypothesis of interest (H) providing the motivation for empirical analysis is surrounded by assumptions and decisions relating to variable selection and definition (V_i), sample selection (S_i), the models employed (M_i) and dynamic specification (D_i). It is argued that

as a result of being embedded within these surrounding assumptions/options, any resulting inferences will be dependent upon not just the truth or falsity of the underlying hypothesis of interest, but also the impact of the decisions relating to $\{V_i, S_i, M_i, D_i\}$. Recognition of this 'jointness of testing' as a result of moving from an initial hypothesis to a composite testable form is present in the philosophy of science, particularly in relation to the Duhem-Quine thesis and the (im)plausibility of 'crucial experiments' allowing the evaluation of hypothesis of interest in isolation from surrounding factors.[††] This issue has been considered in the economics literature also where Cross (1982) provides a theoretical analysis and discussion of the auxiliary assumptions associated with the empirical examination of monetarism. In the present study this form of analysis is extended to provide an empirical evaluation of the impact of variation in empirical design in relation to a very topical issue. As a result of this detailed consideration, the robustness of risk-return inferences is explored rather than consider a specific, single analysis.
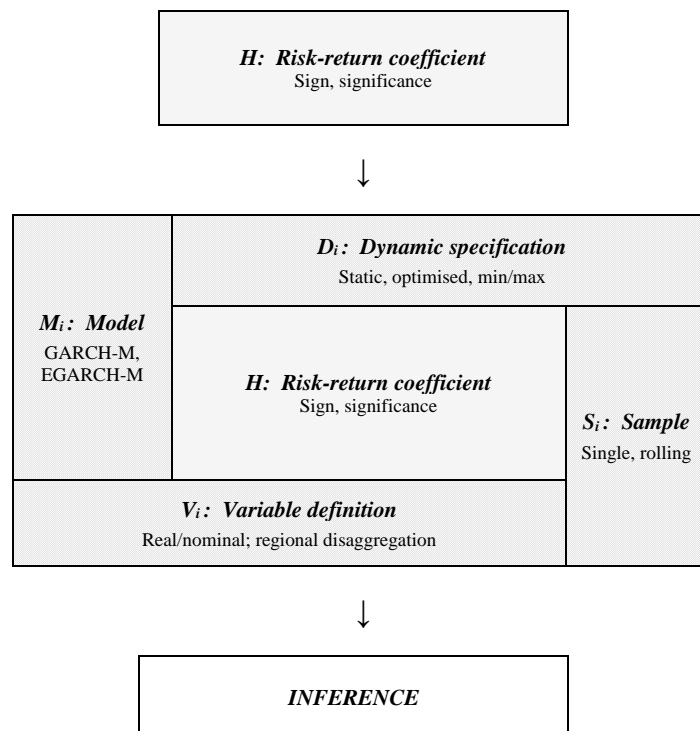
Figure 1. From hypothesis of interest to inference.

4. Results

The results obtained from estimation of the GARCH(1,1)-M and EGARCH(1,1)-M models over the full sample are presented in Tables One and Two. As a result of the use of two models, alternative options for M_i are considered. Similarly, the use of both nominal and real returns for 33 regions (32 boroughs plus the aggregate London series) provides results for alternative V_i. Further to this, the

---

[††]See Harding (1976) for a very readable collection of essays concerning the Duhem-Quine thesis.

tabulated results present a variety of risk-return coefficients obtained from alternative approaches to dynamic specification (D_i). In particular, statistically significant (at the 5% level) estimated risk-return coefficients from static models $\left(\widehat{\delta_0}\right)$, obtained from AIC optimisation of the lag length $\left(\widehat{\delta_{AIC}}\right)$ and the maximum and minimum values obtained across alternative lag specifications $\left(\widehat{\delta_{max}}, \widehat{\delta_{min}}\right)$ are provided to evaluate the impact of variations in D_i. While Tables One and Two allow examination of the impact of variation in $\{V_i, M_i, D_i\}$, the results are generated using the full sample available and hence reflect use of a single value for S_i. In recognition of this, further results are presented in Tables Three and Four where rolling samples are employed to consider the impact of variations in S_i. These full sample and rolling sample results are considered in turn below.

4.1. Full sample results

Turning to the results for the GARCH(1,1)-M model in Tables One and Two, there is evidence of a significant, negative risk-return relationship for a number of series examined. More precisely, there are 9 (13) series for which the static model produces a significant coefficient for nominal (real) returns respectively. In contrast, results obtained from use of AIC optimisation depict a single significant risk-return coefficient for both the nominal and real series. Hence, alternative decisions on both V_i and D_i influence the extent of the derivation of significant results. In terms of V_i, real returns produce a greater number of significant results than nominal returns, and (in)significance varies widely across regions. Similarly, the influence of decisions concerning D_i is apparent via the difference in findings for coefficients from static and AIC optimised models. However, the extent of the detection of significant findings is relatively low given 33 series are examined, as is reflected in the small number of instances in which significant (at the 5% level) $\widehat{\delta_{max}}$ and $\widehat{\delta_{min}}$ coefficients are observed across the full set of lag lengths considered.

Considering the results for the EGARCH(1,1)-M model, a vast increase in significant results for the static model is apparent with nominal and real returns producing 29 and 27 significant negative coefficients respectively. In addition, nominal and real returns generate 2 and 4 significant positive coefficients respectively for the static model. Consequently, variation in M_i via the movement from GARCH-M to EGARCH-M has a substantial impact upon the detection of significant risk-return relationships and leads also to the introduction of positive coefficient estimates to the analysis. Turning to the results for the AIC optimised coefficients, the nominal (real) series produce 4 (6) positive coefficients and 5 (9) negative coefficients. This represents a far more balanced outcome than that observed for the static model. The issue of the sign of the risk-return coefficient is more apparent when considering $\left\{\widehat{\delta_{max}}, \widehat{\delta_{min}}\right\}$ for the EGARCH(1,1)-M model where only one series fails to produce significant coefficients when considering nominal and real returns. Interestingly, these latter results show with regard to the sign of the risk-return coefficient, 31 (33) series produce negative coefficients for the nominal (real) series, while 14 (17) series produce positive coefficients. This range of outcomes for the EGARCH model compared to the GARCH specification, along with the variation in results observed under use of AIC optimisation and static models illustrates clearly the impact of M_i and D_i upon inferences. While the use of nominal or real has an influence upon results derived, with a greater number of significant findings observed for the latter, the impact of V_i upon inferences is most apparent in terms of the effects of disaggregation, with widespread variation observed for the alternative series examined. In summary, these findings display a variation in sign and significance which indicates the counterintuitive negative risk-return relationship depicted by static EGARCH-M

models is far less prevalent when alternative decisions are made concerning dynamic specification and modelling technique within the empirical design.

4.2. Rolling sample results

To explore the impact of the selected sample period upon inferences and whether the conclusions drawn from Tables One and Two are dependent upon the use of the specific full sample available, the EGARCH(1,1)-M is estimated over 72 rolling samples of 180 observations. The results obtained from this analysis are presented in Tables Three and Four below. For each of the series considered, the maximum and minimum statistically significant values for the $\widehat{\delta}_{AIC}$ across these 72 samples are reported. This provides information on the variation in, or range of, results arising from consideration of different samples. Further information on this variation is provided by C_{AIC} which presents the percentage of samples for which statistically significant values of the $\widehat{\delta}_{AIC}$ are observed (under the heading 'total') and their sign (under the headings 'positive' and 'negative'). To illustrate these results, consider the findings in Table Three for the first series (Barking). It can be seen that significant values of $\widehat{\delta}_{AIC}$ from 0.65 down to −6.13 are observed across the 72 samples, and that while 19% of samples return significant positive $\widehat{\delta}_{AIC}$ values and 3% return significant negative values, the $\widehat{\delta}_{AIC}$ is insignificant for the remaining 78% of samples. As the $\widehat{\delta}_{AIC}$ denotes the optimised value of $\widehat{\delta}$ according to the AIC across 13 alternative lag lengths ($p = 0, 1, ..., 12$), further significant values of $\widehat{\delta}$ are potentially available within each sample across the 12 other (non-optimal) lag lengths considered. In recognition of this, C_{ALL} is reported to provide information on the percentage of samples producing significant positive and negative values of $\widehat{\delta}$.

Considering the results in Tables Three and Four, 28 (31) regions produce negative values of $\widehat{\delta}_{AIC}$ for nominal (real) returns, while 25 (30) regions produce positive values. These figures are clearly very similar in terms of the extent of negative/positive values. With regard to the percentages of samples for each series where the $\widehat{\delta}_{AIC}$ was negative/positive, this is again balanced with an average of 14% (23%) of samples returning negative values and 18% (23%) producing positive values for nominal (real) returns. Therefore, the extension of the analysis to consider alternative sample periods has resulted in both increased detection of significant risk-return relationships using the AIC and evidence of its variability with a positive relationship being more frequent than a negative relationship. The results for C_{ALL}, where the percentage of significant values of δ are reported across all lag lengths considered over the 72 samples, show that nearly all samples return significant negative coefficients, with the percentages being 95% and 96% respectively for nominal and real returns. In contrast, on average less than half the samples return significant positive coefficients, with the average number of samples being 42% for both nominal and real returns. Therefore, while consideration of all available significant risk-return coefficient results in a prevalence of counterintuitive negative values, inspection of coefficients obtained from optimisation of the lag length leads to a relatively balanced finding in terms of negative and positive values.

Table 1. Risk-return coefficients: Nominal returns.

Borough	GARCH(1,1)-M				EGARCH(1,1)-M			
	$\widehat{\delta}_0$	$\widehat{\delta}_{AIC}$	$\widehat{\delta}_{max}$	$\widehat{\delta}_{min}$	$\widehat{\delta}_0$	$\widehat{\delta}_{AIC}$	$\widehat{\delta}_{max}$	$\widehat{\delta}_{min}$
Barking	−0.33	0.15	−	−	−3.03[a]	0.18	−1.24	−3.03
Barnet	−0.56[b]	0.67	−0.56	−0.56	−3.88[a]	−21.21	−3.88	−11.89
Bexley	−0.23	−0.77	−	−	−3.45[a]	−4.46[b]	−1.41	−4.46
Brent	−0.30	0.11	−	−	4.62[a]	0.10	7.59	−9.43
Bromley	−0.08	0.52	−	−	−4.08[a]	0.22	49.62	−31.17
Camden	−0.11	−0.49	−	−	−3.58[a]	−0.36	−3.58	−5.57
Croydon	0.03	0.22	−	−	−4.83[a]	0.62	−2.06	−5.22
Ealing	−0.77[b]	0.20	−0.77	−0.77	−3.46[a]	−0.10	3.87	−3.46
Enfield	−0.02	0.40	−	−	−3.05[a]	25.40	−0.96	−3.05
Greenwich	−0.57	−0.93	−	−	−3.19[a]	−0.10	−3.19	−3.19
Hackney	−0.49	−0.16	−	−	−3.30[a]	−0.98	−3.30	−44.60
Hm & Ful	−0.23	8.58	−	−	−3.51[a]	1.74[a]	24.50	−3.51
Haringey	−0.72[b]	−0.39	−0.72	−0.72	−2.74[a]	4.88[a]	4.88	−2.81
Harrow	−0.10	−0.22	−	−	−2.99[a]	−0.20	0.64	−4.23
Havering	−0.14	8.29	−	−	−3.07[a]	1.10[b]	11.40	−6.05
Hillingdon	−0.62[b]	−0.60	−0.62	−0.62	−2.62[a]	−1.00	0.50	−10.05
Hounslow	−0.15	−0.54	−	−	−2.53[a]	−17.83[b]	3.08	−24.57
Islington	−0.06	0.35	−	−	3.76[a]	−8.76[a]	3.76	−12.92
Kensington	−0.78[b]	−0.53	−0.78	−0.78	−2.72[a]	−0.53	−0.78	−2.72
Kingston	−0.16	0.02	−	−	−0.06	0.23	−	−
Lambeth	−0.18	0.51	3.76	3.76	4.59[a]	0.45	4.59	0.96
Lewisham	0.19	−0.37	−1.67	−1.67	−3.21[a]	−4.31	−1.86	−22.43
Merton	−1.10[a]	−1.12	−1.10	−1.10	−3.25[a]	−1.54	−0.76	−11.28
Newham	0.00	−0.04	−	−	−0.03	−11.98[a]	−5.05	−27.98
Redbridge	−0.20	−0.35	−	−	−2.81[a]	−11.43	−0.63	−2.81
Richmond	0.12	−0.21	−	−	−4.33[a]	−0.09	−4.11	−4.33
Southwark	−0.45	0.19	3.21	3.21	−3.36[a]	0.86[a]	0.90	−3.36
Sutton	0.20	−0.11	−	−	−4.82[a]	−0.23[a]	41.25	−4.82
Tower Ham	−0.81[a]	−1.14	−0.81	−0.81	−2.14[a]	−1.46	−2.14	−9.73
Walt For	−0.08	−0.52	−	−	−4.37[a]	−0.28	−4.37	−35.67
Wandsworth	−0.30	0.00	−	−	−6.02[a]	−0.10	−1.11	−6.02
Westminster	−0.89[a]	−0.89[b]	−0.89	−0.95	−3.11[a]	3.13	325.21	−17.90
London	−0.81[a]	−0.36	−0.64	−0.81	−4.57[a]	−0.26	−4.57	−9.02

Notes: The above figures are the static mean equation, AIC optimised, maximum and minimum risk-return coefficients obtained from the GARCH-M and EGARCH-M specifications for nominal house returns. Significance at the 5% and 1% levels are denoted by [a] and [b] respectively.

Table 2. Risk-return coefficients: Real returns.

| Borough	GARCH(1,1)-M				EGARCH(1,1)-M			
	$\widehat{\delta}_0$	$\widehat{\delta}_{AIC}$	$\widehat{\delta}_{max}$	$\widehat{\delta}_{min}$	$\widehat{\delta}_0$	$\widehat{\delta}_{AIC}$	$\widehat{\delta}_{max}$	$\widehat{\delta}_{min}$
Barking	−0.48	0.21	–	–	−3.02^a	0.05	−1.07	−7.55
Barnet	−1.04^a	−166.15	−1.04	−1.04	−3.41^a	24.60^a	24.60	−4.39
Bexley	−3.30^a	0.49	−1.30	−3.30	−3.57^a	−6.92	−3.57	−4.46
Brent	−0.54	−0.09	–	–	−2.53^a	−22.20^a	−2.53	−22.20
Bromley	−0.39	−1.15	–	–	−4.08^a	58.69^a	58.69	−4.08
Camden	−0.16	0.28	–	–	−3.32^a	1.28	−3.32	−44.45
Croydon	−0.15	−0.97	–	–	−4.05^a	−22.60	13.78	−4.05
Ealing	−1.36^a	15.39	−0.99	−1.36	−3.32^a	6.86^a	6.86	−14.57
Enfield	−0.89^a	42.03	−0.89	−0.89	−2.83^a	−2.23^a	4.70	−2.83
Greenwich	−0.69	−1.37	−1.10	−1.10	−3.13^a	13.67^a	37.93	−32.41
Hackney	−0.42	−41.64	–	–	−3.83^a	−3.16^a	−3.16	−5.49
Ham & Ful	−0.35	−0.35	–	–	−3.22^a	4.01	−3.16	−25.01
Haringey	−0.73^b	−0.82	−0.73	−0.73	−2.84^a	16.89^a	16.89	−14.52
Harrow	−0.13	−0.02	–	–	−3.24^a	−0.94	−2.47	−4.60
Havering	−1.75^a	−1.04^b	−1.04	−1.75	−3.16^a	−6.34^a	2.35	−11.82
Hillingdon	−1.03^a	−0.09	−0.75	−1.03	−2.46^a	−18.81^a	66.56	−18.81
Hounslow	−0.11	0.03	−2.08	−2.08	3.27^a	−43.10^a	3.27	−43.10
Islington	0.00	−0.12	–	–	7.12^a	−23.43^a	7.12	−23.43
Kensington	−0.84	−0.22	–	–	−2.66^a	−0.42	−2.51	−7.19
Kingston	−0.30	−0.23	–	–	−3.59^a	0.54	2.99	−3.59
Lambeth	−0.36	0.85	–	–	−8.85^a	25.88	−8.85	−8.85
Lewisham	0.07	−0.57	−1.73	−1.73	0.13	39.25^a	58.49	−9.22
Merton	−1.18^a	−1.35	−1.18	−1.18	−3.28^a	−31.42^a	−3.28	−31.42
Newham	0.00	−0.01	–	–	−0.08	−0.02	−52.65	−52.65
Redbridge	−1.58^a	−0.29	−1.58	−1.58	−3.04^a	−0.32	−1.54	−9.27
Richmond	−0.43	−0.63	–	–	−3.92^a	−0.13	−2.82	−12.98
Southwark	−0.46	0.24	–	–	6.78^a	71.93^a	71.93	0.84
Sutton	−0.06	226.50	–	–	4.97^a	−16.62^a	18.34	−16.62
Tower Ham	−1.12^a	0.02	−1.12	−1.12	−2.38^a	−0.23	−2.38	−22.89
Walt For	−0.12	−0.06	–	–	−4.23^a	−0.01	−4.23	−6.71
Wandsworth	−0.46^b	0.25	−0.46	−0.46	−4.98^a	0.59^a	0.59	−4.98
Westminster	−0.85^a	−0.71	−0.59	−0.85	−2.37^a	−0.90	8.51	−3.81
London	−1.14^a	−0.61	−0.79	−1.14	−6.76^a	−0.15	−6.76	−6.76

Notes: The above figures are the static mean equation, AIC optimised, maximum and minimum risk-return coefficients obtained from GARCH-M and EGARCH-M specifications for real house returns. Significance at the 5% and 1% levels are denoted by ^{a} and ^{b} respectively.

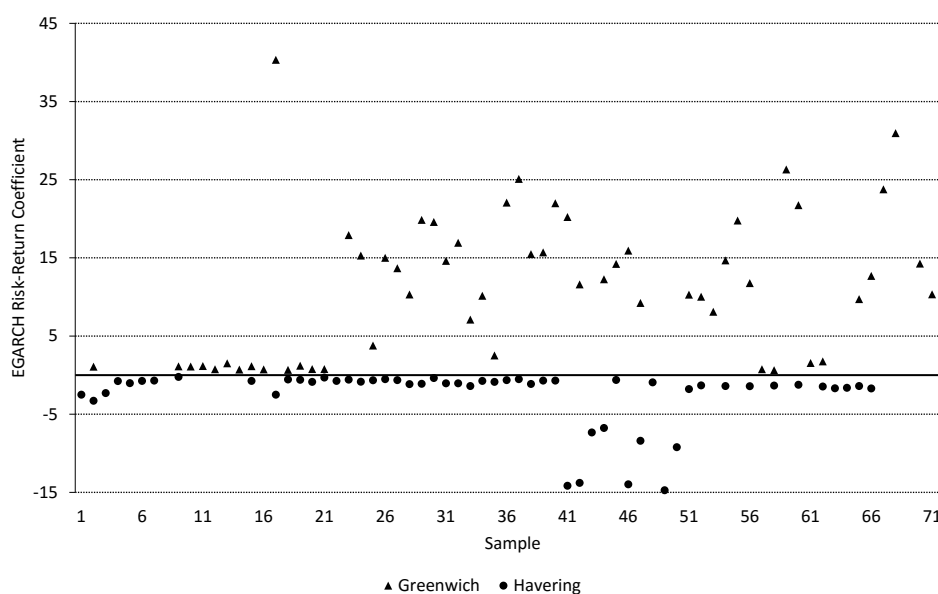**Table 3. Rolling sample risk-return analysis: Nominal returns.**

| Borough | $\widehat{\delta}_{AIC}$ | | C_{AIC} | | | C_{ALL} | |
	max	min	+ve	−ve	total	+ve	−ve
Barking	0.65	−6.13	19	3	22	56	100
Barnet	10.14	−19.61	25	17	42	40	96
Bexley	0.56	−9.57	3	19	22	6	100
Brent	0.72	−1.12	14	1	15	44	99
Bromley	1.71	−32.01	58	11	69	90	79
Camden	−0.71	−1.33	0	22	22	1	100
Croydon	9.83	−6.16	13	14	26	42	82
Ealing	17.03	−6.81	40	3	43	79	97
Enfield	13.64	1.90	21	0	21	74	88
Greenwich	4.61	−40.68	14	14	28	43	100
Hackney	1.96	−6.05	47	4	51	60	100
Ham & Ful	26.10	−0.97	14	4	18	75	100
Haringey	53.86	0.49	26	0	26	78	99
Harrow	0.88	−7.70	31	6	36	83	100
Havering	15.57	−14.33	7	4	11	44	90
Hillingdon	0.98	−6.20	10	24	33	43	100
Hounslow	0.99	−8.71	3	18	21	21	100
Islington	6.54	−20.06	67	8	75	88	97
Kensington	–	–	0	0	0	3	96
Kingston	37.36	2.54	36	0	36	58	99
Lambeth	2.26	0.53	82	0	82	97	63
Lewisham	21.83	−62.64	6	92	97	39	100
Merton	−1.63	−18.57	0	51	51	28	100
Newham	0.79	−0.89	4	3	7	29	89
Redbridge	−4.51	−24.16	0	31	31	0	100
Richmond	−1.72	−6.20	0	17	17	0	100
Southwark	1.09	−11.18	13	1	14	25	92
Sutton	57.17	−84.19	19	17	36	54	100
Tower Ham	−1.72	−12.63	0	26	26	1	100
Walt For	−0.69	−41.02	0	33	33	38	82
Wandsworth	6.16	−0.40	6	1	7	17	99
Westminster	−1.44	−4.51	0	26	26	40	93
London	0.76	−7.49	1	4	6	1	99

Notes: The first two columns of figures provide maximum and minimum values of the $\widehat{\delta}_{AIC}$ significant at the 5% level over 72 rolling samples. C_{AIC} is the percentage of samples generating significant $\widehat{\delta}_{AIC}$ values with the percentage of negative and positive values along with the total percentage reported. C_{ALL} provides the percentage of samples generating significant $\widehat{\delta}$ values with the percentages for both positive and negative coefficients provided.

The above findings illustrate the variation in results concerning risk and return in the housing market due to alternative decisions regarding the use of models, variables, dynamic specification and samples. These findings have illustrated the empirical relevance of the jointness of hypothesis testing as depicted in Figure One. More specifically, the analysis has made explicit the dependence of the often cited counterintuitive finding of a negative risk-return relationship upon the approach taken to empirical design. While the full sample results for EGARCH-M model show the prevalence of negative coefficients, results observed for AIC optimisation and rolling samples do not support this finding. To illustrate further the variability of results across alternative samples and series (that is, alternative S_i and V_i), Figures Two and Three present the significant $\widehat{\delta}_{AIC}$ values for three London boroughs using the EGARCH(1,1)-M model. In Figure Two, the values of the $\widehat{\delta}_{AIC}$ for real house prices for Waltham Forest are plotted. To ease consideration of the crucial issue of the sign of the coefficient, a line at zero is included for the horizontal axis. From inspection of this graph, it can be seen that the first half of the 72 rolling samples are dominated by positive estimated coefficients while the second half of the samples does not return anything but negative values. As a consequence, variability in results is apparent when just the decision regarding samples is allowed to vary and decisions on the model, dynamic specification and variable are held constant. Such variation when just 1 factor is variable (and 3 are not), is compelling evidence against a certainty in the nature of the risk-return relationship. Similarly, to illustrate variability across alternative variables, Figure Three depicts analogous results for real house prices in Greenwich and Havering. In this instance, variation across series is illustrated as the former region generates positive $\widehat{\delta}_{AIC}$ values only, while the latter has negative values only. In summary, Figures Two and Three present variability within series and across series respectively.
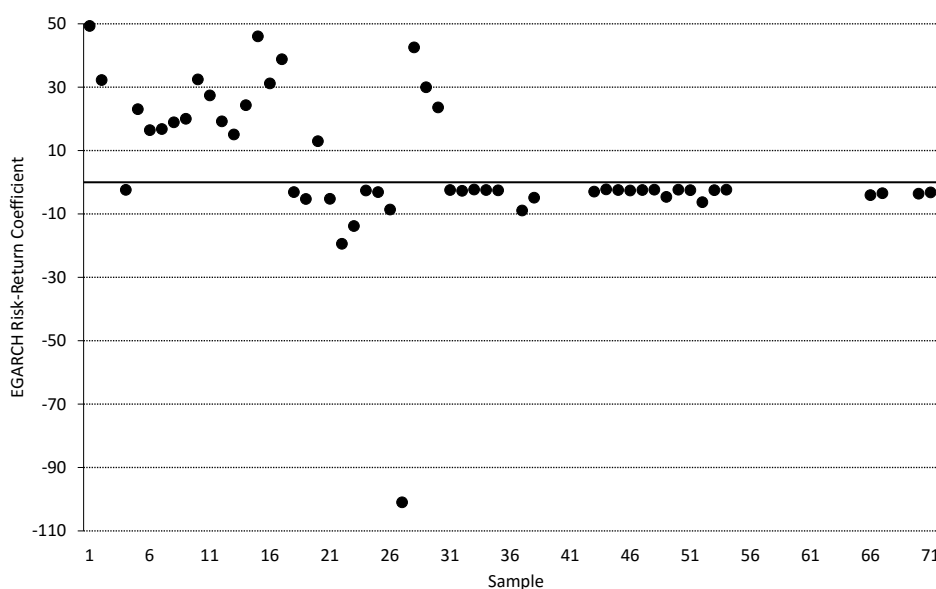
Figure 2. Rolling sample risk-return for Waltham Forest real house prices.

Figure 3. Rolling sample risk-return for Greenwich and Haveing real house prices.

5. Concluding Remarks

The above analysis has provided a detailed examination of the risk-return relationship in the London housing market. In addition to providing the first findings within the literature for this issue, the analysis contains a more general message relating to the impact of decisions concerning empirical design upon risk-return inferences. The results provided indicate that while the counterintuitive negative risk-return relationship discussed frequently in the literature does arise, this is most prevalent when very specific decisions are taken with regard to models, dynamic specification and sample periods. In contrast to this, a more flexible and more preferred approach involving optimisation results in the generation of mixed results where a relatively balanced number of positive and negative findings are apparent. The analysis therefore provides a clear message concerning the impact of typically implicit and overlooked decisions on empirical design upon inferences. An obvious future line of research would involve a meta-analysis to consider the linkages between empirical design and inferences in previous research and hence a broader evaluation of the impact of design upon the robustness of conclusions relating to the risk-return relationship.

Conflict of Interest

All authors declare no conflicts of interest in this paper.

References

Abbott A, De-Vita G (2012) Pairwise convergence of district-level house prices in London. *Urban Stud* 49: 719-738.

Barros C, Gil-Alana L, Payne J (2015) Modeling the long memory behavior in U.S. housing price volatility. *J Hous Res* 24: 87-106.

Bayer P, Ellickson B, Ellickson P (2010) Dynamic asset pricing in a system of local housing markets. *Am Econ Rev* 100: 368-372.

Brueckner J (1997) Consumption and investment motives and the portfolio choices of homeowners. *J Real Estate Financ Econ* 15: 159-180.

Case K, Cotter J, Gabriel S (2011) Housing risk and return: Evidence of a housing asset-pricing model. *J Portf Manag* 35: 89-109.

Case K, Quigley J, Shiller R (2013) Wealth effects revisited 1975-2012. *Crit Financ Rev* 2: 101-128.

Case K, Shiller R (1988) The behavior of home buyers in boom and post-boom markets. *New Engl Econ Rev* 6: 29-46.

Case K, Shiller R (1989) The efficiency of the market for single-family homes. *Am Econ Rev* 79: 125-37.

Case K, Shiller R (1990) Forecasting prices and excess returns in the housing market. *Am Real Estate Urban Econ Associ J* 18: 253-73.

Costello G, Fraser P, Groenewold N (2011) House prices, non-fundamental components and interstate spillovers: The Australian experience. *J Bank Financ* 35: 653-669.

Crawford G, Fratantoni M (2003) Assessing the forecasting performance of regime-switching, ARIMA and GARCH models of house prices. *Real Estate Econ* 31: 223-243.

Crawford G, Rosenblatt E (1995) Efficient mortgage default option exercise: Evidence from loan loss severity. *J Real Estate Res* 10: 543-555.

Cross R (1982) The Duhem-Quine thesis, Lakatos and the appraisal of theories in macroeconomics. *Econ J* 92: 320-340.

Dale D, Wolf R, Yang HF (2015) An assessment of the risk and return of residential real estate. *Manag Financ* 41: 591-599.

Dolde W, Tirtiroglu D (1997) Temporal and spatial information diffusion in real estate price changes and variances. *Real Estate Econ* 25: 539-565.

Dolde W, Tirtiroglu D (2002) House price volatility changes and their effects. *Real Estate Econ* 30: 41-66.

Domian D, Wolf R, Yang HF (2015) An assessment of the risk and return of residential real estate. *Manag Financ* 41: 591-599.

Engle R, Lilien D, Robbins R (1987) Estimating time varying risk premia in the term structure: The ARCH-M model. *Econom* 55: 391-407.

Foster C, Order RV (1984) FHA terminations: A prelude to rational mortgage pricing. *J American Real Estate Urban Econ Assoc* 13: 273-291.

Gallin J (2006) The long-run relationship between house prices and income: Evidence from local housing markets. *Real Estate Econ* 34: 417-438.

Goetzmann W (1993) The single family home in the investment portfolio. *J Real Estate Financ Econ* 6: 201-222.

Goodhart C and Hoffman B (2007) House Prices and the Macroeconomy: Implications for Banking and Price Stability Oxford: Oxford University Press.

Han L (2013) Understanding the puzzling risk-return relationship for housing. *Rev Financ Stud* 26: 877-928.

Harding S (1976) Can theories be refuted? Dordrecht: Springer.

Holly S, Jones N (1997) House prices since the 1940s: cointegration, demography and asymmetries. *Econ Model* 14: 549-565.

Huang M, Wang TC (2015) Housing-bubble vulnerability and diversification opportunities during housing boom–bust cycles: evidence from decomposition of asset price returns. *Ann Regional Sci* 54: 605-637.

Huang M, Wang TC (2017) An early alarm system for housing bubbles. *Q Rev Econ Financ* 63: 34-49.

Im K, Pesaran M, Shin Y (2003) Testing for unit roots in heterogeneous panels. *J Econ* 115: 53-74.

Lee C (2009) Housing price volatility and its determinants. *Int J Hous Mark Anal* 2: 293-308.

Lee C (2017) An examination of the risk-return relation in the Australian housing market. *Int J Hous Mark and Anal* 10: 431-449.

Lin P, Fuerst F (2014) Volatility clustering, risk-return relationship and asymmetric adjustment in Canadian housing markets. *J Real Estate Portf Manag* 20: 37-46.

Nelson D (1991) Conditional heteroskedasticity in asset returns: A new approach. *Econom* 59: 347-370.

Miles W (2008) Volatility clustering in U.S. home prices. *J Real Estate Res* 30: 73-90.

Miles W (2011) Clustering in U.K. home price volatility. *J Hous Res* 20: 87-100.

Miller N, Pandher G (2008) Idiosyncractic volatility and the housing market. *J Hous Res* 17: 13-32.

Miller N, Peng L (2006) Exploring metropolitan price volatility. *J Real Estate Financ Econ* 33: 5-18.

Morley B, Thomas D (2011) Risk-return relationships and asymmetric adjustment in the UK housing market. *Appl Financ Econ* 21: 735-742.

Morley B, Thomas D (2016) An empirical analysis of UK house price risk variation by property type. *Rev Econ Financ* 6: 45-56.

Scruggs J (1998) Resolving the puzzling intertemporal relation between the market risk premium and conditional market variance: A two-factor approach. *J Financ* 26: 877-928.

Shafir E, Diamond P, Tversky A (1997) Money illusion *Q J Econ* 112: 341-374.

Tsang CK, Wong WK, Horowitz I (2016) Arbitrage opportunities, efficiency, and the role of risk preferences in the Hong Kong property market. *Stud Econ Financ* 33: 735-754.

Wheaton W (2015) The volatility of real estate markets: A decomposition. *J Portf Manag* 41: 140-150.

Zhou J, Gao Y (2012) Tail dependence in international real estate securities markets. *J Real Estate Financ Econ* 45: 128-151.