



Assessing Competition Policy: Methodologies, Gaps and Agenda for Future Research

Stephen Davies*

and

Peter Ormosi**

***ESRC Centre for Competition Policy and School of
Economics, University of East Anglia**

****ESRC Centre for Competition Policy and Norwich Business
School**

CCP Working Paper 10-19

Abstract: Research by academics and competition agencies on evaluating competition policy has grown rapidly during the last two decades. This paper surveys the literature in order to (i) assess the fitness for purpose of the main quantitative methodologies employed, and (ii) identify the main undeveloped areas and unanswered questions for future research. It suggests that policy evaluation is necessarily an imprecise science and that all existing methodologies have strengths and limitations. The areas where the need is most pressing for further work include: understanding why Article 102 cases are only infrequently evaluated; the need to bring conscious discussion of the counterfactual firmly into the foreground; a wider definition of policy to include success in deterrence and detection. At the heart of the discussion is the impact of selection bias on most aspects of evaluation. These topics are the focus of ongoing work in the CCP.

November 2010

JEL Classification: L40, K21

Keywords: Competition enforcement, evaluation methods, simulation, event studies, difference-in-differences, detection, selection bias, deterrence

Acknowledgements: This draft is circulated with the anticipation that any readers might point us in the direction of omitted references and potential inaccuracies and misunderstandings. Comments and corrections will be very gratefully received and acknowledged. The support of the Economic and Social Research Council is gratefully acknowledged.

Contact Details:

Stephen Davies, ESRC Centre for Competition Policy, and School of Economics, University of East Anglia, Norwich NR4 7TJ, UK; s.w.davies@uea.ac.uk

Peter Ormosi, Norwich Business School and ESRC Centre for Competition Policy, University of East Anglia, Norwich, NR4 7TJ; p.ormosi@uea.ac.uk

Table of Contents

TABLE OF CONTENTS.....	1
1. INTRODUCTION	4
1.1 THE PURPOSES OF EVALUATION	4
1.2 EVALUATION BY POLICY AREAS	6
1.3 STRUCTURE OF THE PAPER.....	11
2. LITERATURE SURVEY OF METHODOLOGIES.....	11
2.1 A TAXONOMY	11
2.2 SIMULATION.....	12
2.3 EVENT STUDIES	15
2.4 DIFFERENCE IN DIFFERENCES	20
2.5 OTHERS	24
2.6 PRELIMINARY CONCLUSIONS	25
3. COMPARING COUNTERFACTUALS: CARTELS AS A CASE STUDY	30
4. SELECTION BIAS AND DETERRENCE	34
4.1 SOME POTENTIAL SOURCES OF SELECTION BIAS	34
4.2 HOW MUCH OF THE ICEBERG LIES BELOW THE WATERLINE	35
5. CONCLUSIONS AND FUTURE RESEARCH PRIORITIES.....	47
6. REFERENCES.....	50

2. Introduction

Research on evaluating competition policy has grown rapidly in the last 10-20 years, and this is a good time to ask: “what have we learned so far?” and “what are the pressing questions for future research”? This survey¹ is not so much about the effectiveness of competition policy *per se*, but more the methodologies used to evaluate it. It points to the gaps or unanswered questions in the evaluation area, and sets an agenda of issues where we believe the need for future research is most pressing.

It complements a number of other recent surveys. Some have been produced or commissioned by the CAs themselves, such as Davies (2010), and Buccirosi et al. (2006); other works include a critical overview of impact evaluation by Bergman (2008), and surveys of methodologies in the US by Werden (2008), and by Kemp and Sinderen (2008) for the Netherlands.

2.1 *The purposes of evaluation*

Evaluation is undertaken both by the Competition Authorities (CAs) and independent academics. It can take many forms, not least because it is required for different purposes. At the risk of over-simplification, these fall into four broad categories:

Accountability

Increasingly, there is an obligation on CAs around the world, to their Governments² to quantify the aggregate benefits of competition policy

¹ Earlier versions of some of the ideas in this paper can be found in Davies’s keynote lecture, “Ex-post evaluation methodologies of competition enforcement “, presented to the Conference on Ex-post evaluation of competition policy, Mannheim June 4, 2009, at: ftp://ftp.zew.de/pub/zew-docs/veranstaltungen/evalcomp2009/Programme_Evaluation2009.pdf. Davies’s subsequent evaluation of the evaluation methods used by OFT is reported in “A Review of OFT’s Impact Estimation Methods” (January 2010), OFT1164, at: http://www.of.gov.uk/shared_of/reports/Evaluating-OFTs-work/oft1164.pdf This includes a survey of the evaluation methodologies used by CAs.

² For example, in the UK, HM Treasury has set a performance target for OFT that it should deliver direct financial benefits to consumers of at least five times its cost to the taxpayer.

(measured perhaps by increased consumer surplus). The aggregate estimate is often then assessed against some pre-specified target to judge whether the CA has met its required objectives. These impact evaluations typically involve using some of the methodologies described below.

Assessment of specific policies

Sometimes, the purpose is to evaluate a specific area of policy (e.g. merger control), and/or of the success of a particular intervention (e.g. a prosecuted cartel, or prohibited merger.) Again, this is often undertaken (or at least overseen) internally by the CA itself, perhaps to check on the quality of its own decision-making – the rigour of its analysis, data collection etc – and/or to help set internal priorities

Assessment in the broader academic literature

Policy evaluation in the academic literature appears in a multitude of guises. Sometimes it might be designed as an independent check on the performance of the CAs. More often however, the immediate objectives of the research may be more academic, focusing primarily on the development/testing of theory and/or empirical techniques. Nevertheless, these studies often employ important anti-trust cases, and have frequently led to developments in evaluation methodologies.

Estimating damages and fines

In any private damages case (or indeed when a CA sets a fine, where applicable), this entails an evaluation of harm, and, by implication, the gains resulting from the CA intervention to remove that harm.

Depending on the purpose, and who undertakes the evaluation, the sophistication of the evaluation can vary considerably. At one extreme, simple rules of thumb are often used in accountability exercises for obvious reasons of practicability; at the other extreme are highly sophisticated extensions/applications of theory using frontier econometric methodologies.

The question of who undertakes the evaluation, and how deeply, raises important issues of the potential distortion of incentives and priorities within the CA, and we return to this in section 4.2.3 below.

2.2 *Evaluation by policy areas*

We define the constituent areas of competition policy (Table 1) initially as all activities undertaken by enforcement agencies. While these are not all strictly law enforcement, the non-enforcement activities are often extremely resource-demanding and it would seem appropriate, at least for the CAs themselves, to evaluate their impact too.

For each area of activity, the Table provides our assessment³ of the extent of the academic literature and CAs' evaluations, classified by three key methodologies. These methodologies are discussed at greater length in the following section, but briefly they can be defined as follows. *Simulation* typically entails formally or informally modelling the nature of competition in a market, calibrating the parameters using real world information (sometimes estimated econometrically) and then assessing how the intervention will change the equilibrium relative to what would have happened without the intervention. In the academic literature, this may be 'full-fledged' (i.e. theoretically and econometrically often highly sophisticated); when used for evaluation by CAs, it is often more of a 'back-of-envelope' calculation, based on simplified models and calibrated with 'typical' values for parameters such as the demand elasticity. *Event Studies* use the financial markets' assessment of the impact of an event. In the context of say merger enforcement, the events in question are the initial announcement of the merger, the CA's announcement of investigation, and then its subsequent decision. The effects are quantified by comparing movements in stock prices – both of the parties and their immediate rivals – with movements in more general stock price indices. The *Difference-in-differences (DiD)* methodology involves a comparison of, say, prices before and after an event (e.g. merger

³ Our assessments are probably uncontroversial and are consistent with the emphases in previous literature surveys, for example, Bergman (2008) and Werden (2008)

or dawn raid) relative to some other real world control, i.e. a similar market without the event, or within the same market for firms not involved in the event. Again, there is a variety of related less time- and data-intensive more ad-hoc methods, for example, estimating the extent of cartel overcharge by comparing prices within the cartel period with pre- or post-cartel prices.

Table 1: Extent of evaluation literature by broad area of policy

	Methodology	Academic Literature	CAs
Enforcement			
Merger control	Simulation	Extensive	Extensive
	Event Studies	Extensive	Some
	DiD etc	Extensive	Some
Cartels/Article 101	Simulation	Some	Few
	Event Studies	Some	Few
	DiD etc	Some	Some
Abuse/Article 102	Simulation	Few	None
	Event Studies	None	None
	DiD etc	Few	None
Non-enforcement:			
Advocacy	None	None	None
Compliance	None	None	None
Consumer Education	None	None	None

Most of the previous literature has concentrated heavily on mergers and cartels⁴. The academic literature is most developed and extensive for mergers, and this includes highly influential contributions on simulation, but also event studies and DiD. In turn, these have had a strong impact on the evaluations undertaken by the CAs themselves, and it is commonplace for the major CAs to apply simulation models in their own evaluations. There is also an extensive literature on cartel overcharge, with ad-hoc versions of DiD being more frequent (see section 3 below). CAs sometimes employ the latter in their evaluations of the consumer benefits from cartel-busting.

⁴ The following sections reference many of the key academic contributions, but see also the surveys mentioned in the introduction for additional references.

Since mergers and cartels feature extensively in the remainder of this paper, the remaining preliminary comments here are confined to the other areas where evaluation has been much less common.

2.2.1 *Article 102*

Although our assessment that cases involving abuse of power (Art 102) have featured far less heavily in the existing evaluation literatures is impressionistic and not yet documented by reference to hard statistics, the following fragments of evidence strongly suggest that it is valid. Davies's review (2010, Table 4.2), of the OFT's Impact Estimates, reveals that for 2006-9, while the OFT were able to evaluate 20 mergers, only 1 case of Article 102 enforcement was assessed. OXERA, in its report for the EC (2010, p.16) suggests that "There have been relatively few cases of exploitative abuse of dominance found by competition authorities or courts, either at the EU level, or in the Member States". In the US, Werden (2008, p.446) suggests that "Non-merger civil enforcement accounts for relatively few cases and for far less consumer savings than either criminal or merger enforcement"⁵. Similarly, within the academic literature, relevant studies are few. This is probably symptomatic of a more general scarcity of empirical IO work in the broad area of Article 102. For example, Slade (2008) concludes her survey of the empirical literature on the effects of vertical restraints (p.28) by suggesting that "Perhaps the most important lesson that can be learned ...is how scant that evidence is, especially when compared to the amount of theoretical research."

There are two obvious explanations for this. The first is, simply, that competition authorities bring relatively few cases of Article 102 abuse to fruition, and the second is that evaluation is unusually difficult to conduct in these cases. Werden (2008, pp. 442-3) suggests that, certainly, the latter is true: "In assessing the effects of antitrust enforcement, cases involving

⁵ He explains that "Non-merger civil enforcement relates to single-competitor exclusionary conduct, vertical restraints and agreements among competitors other than hard-core cartels or mergers".

exclusionary conduct present the greatest challenge. The effects of potentially exclusionary conduct are apt to be subtle and can be experienced long after the conduct itself". He points to the difficulties in establishing the extent to which rivals are harmed and the impact on consumers, and explains that delicate trade-offs are involved where the practice may entail an element of efficiency enhancement. He also identifies predatory behaviour cases as particularly problematic, given the need to quantify short-run and long-run impacts of opposite directions.

2.2.2 *Advocacy*

It is sometimes argued that the benefits of advocacy activities may exceed those of a enforcement actions. Typically, competition authorities allocate a significant proportion of their budget to advocacy activities.⁶ Despite this, there has been little research measuring the impact of competition advocacy, and the work that has been done is predominantly qualitative. The OFT (2010b) conducted a survey, based on three case studies, of UK government officials to assess whether its competition related advice was taken into consideration and influenced policymakers. The FTC and the Italian Competition Authority have also both carried out detailed surveys to evaluate the impact of their advocacy activities⁷.

Looking to the future, an obvious question is whether the methods used for assessing merger and cartel enforcement could be applied also to measuring the impact of advocacy. An obvious problem would be to identify the right counterfactual. For example to measure the impact of advocacy that resulted in the dropping of a piece of draft legislation which would have had anticompetitive consequences, we would need to know what would have happened had the anticompetitive proposal been enacted. For enforcement activities there is always the possibility to use a similar market as

⁶ The International Competition Network (2002) study on competition advocacy reported that amongst those countries that were able to quantify the resources they devoted to advocacy almost one third reported between 20 and 30%; the rest below 20%.

⁷ The FTC study is referred to in Majoras (2005) For the Italian study see: Arisi and Esposito (2007)

counterfactual, but in this case it would have to be a similar jurisdiction, something that would be even harder to find.⁸ It would also be hard to identify what version of the proposed legislation would have been accepted in absence of successful advocacy efforts. These concerns have also been identified by Evenett (2006).⁹

2.2.3 *Education: compliance by firms*

Encouraging business compliance with competition law is a potentially important dimension of CA activity. However, attempts to measure the efficacy of compliance programmes *per se* are rare¹⁰. In one example, OFT (2010a) presents some survey findings designed to understand what motivates business compliance and includes examples of their compliance activities. Responding firms mentioned reputational damage, financial penalties, individual sanctions (e.g. risk of criminal proceedings) as key motivations for compliance. Compliance is obviously linked to deterrence, which we discuss more generally in section 4.

2.2.4 *Education: consumers*

Much consumer education relates to consumer protection law and falls outside our remit. However, insofar as better educated consumers should be less frequently exposed to asymmetric information, reduced market imperfections might enhance competition between firms. Again, this is an under-researched area, but some CAs are beginning to work on evaluation in this area.¹¹

⁸ It may be equally hard to show that a proposed piece of legislation was prevented as a result of advocacy rather than other political considerations.

⁹ As Evenett points out, the problems are even greater in assessing competition advocacy in the broader sense, such as increasing public awareness.

¹⁰ The scarcity of evidence also reflects the fact that in many countries the CA is not allowed to carry out compliance activities.

¹¹ For the OFT, see: OFT (2010c) p.30 fn.45. See also OECD's report on best practice of the Communication of competition authorities at: <http://www.oecd.org/dataoecd/20/40/2492536.pdf>

2.3 *Structure of the paper*

Section 2 describes and assesses the three main evaluation methodologies. Sections 3 and 4 address some of the questions raised in Section 2, but which are not peculiar to any particular methodology. Firstly, Section 3 discusses the fundamental issue in any evaluation exercise: the choice of counterfactual. It illustrates some of the issues by reference to a comparative assessment of how cartel overcharge has been evaluated in the past. Section 4 turns to arguably the biggest challenge faced by the evaluation project – the possibility that results may be seriously undermined by pervasive biases in the nature of sample cases typically used in evaluation. Section 5 concludes by drawing together our ‘future research priorities’.

3. Literature Survey of methodologies

3.1 *A Taxonomy*

This section summarises the previous literatures, focusing mainly on the three methodologies, with a brief discussion of all other methodologies at the section end. We refer to this as our ‘taxonomy’ of methodologies: there are at least three other taxonomies in the recent literature: Bergman (2008), Buccirosi et al. (2006), OXERA (2009)¹². The first two are sufficiently similar to our own for us to merely refer the interested reader to the original papers, but OXERA (2009) is sufficiently different to merit a short summary here. The authors group the existing methods and models into three broad types:

- **Comparator** based, which includes (i) DiD (ii) cross-section (across firms, markets or countries) econometrics or averages, and (iii) time series econometrics or interpolation for during, before and/or after.
- **Market structure** based using IO models especially structural models.

¹² This is a commissioned study report for the European Commission on Quantifying antitrust damages. There are obvious parallels (although by no means always exact) between the savings that CAs achieve for consumers from removing anti-competitive practices and claims for damages against defendants responsible for inflicting losses on their customers and rivals and/or suppliers.

- **Financial analysis** based, including event studies but also other more descriptive techniques well established in accountancy/finance.

This taxonomy is more exhaustive than our own, but their three categories correspond fairly closely to our own categories DiD (loosely defined), simulation and event studies respectively.

3.2 *Simulation*

3.2.1 Definition

By simulation, we refer to evaluation based on (i) explicit formal modelling of the nature of competition in the market, then (ii) calibrating the model with real world information (sometimes estimated econometrically), before (iii) using it to assess how the equilibrium will change with and without an event/intervention.

Thus, in the first stage, a model is selected based on a reading of the nature of oligopoly in the market concerned: Are products homogeneous or differentiated? Are there capacity constraints? Do firms act unilaterally or in a coordinated way? If products are differentiated, are they symmetrically so, or is competition more localised? This stage often involves a structural model derived from a game theoretic perspective, coupled with a particular model of the demand system, e.g. logit, nested logit or random utility. In the second stage, the model's parameters are calibrated. This will sometimes be derived from direct observation (e.g. existing market shares, prices and extraneous estimates of demand elasticities), but in other contexts, it may be necessary to undertake full-blown econometric estimation of the demand system in order to derive estimates of those parameters. By substitution of these 'known' parameters into the equilibrium conditions derived in stage 1, one can then recover (i.e. solve for) other, unknown, parameters. In some cases these unknowns might be the firms' marginal costs, in others they might be parameters which summarise the extent of 'market power'.

In the third stage, the post-event equilibrium is simulated using the calibrated parameters, to derive comparative statics. For example when evaluating the

effects of prohibiting a merger, the pre-merger equilibrium is compared with the hypothetical non-intervened equilibrium, allowing for coordination in the price-setting of the brands of the newly merged firm, any claimed synergy effects and reactions by outsiders (i.e. competitors). In principle, one might also simulate a switch in the prevailing behaviour of firms as a consequence of the merger; for example, if it is suspected that the merger would result in a coordinated effect between merging and non-merging firms, one could compute alternative collusive post-merger equilibria. Rather more difficult are the possibilities that firms might want to reposition their products, or where the potential for new entry becomes important.

Simulation may be either *ex-ante* or *ex-post* counterfactual analysis. The latter is backward-looking - what outcome would have happened, had, say, a cartel not actually existed; while *ex-ante* looks forward – anticipating whether or not, say, a merger would have had coordinated effects, if cleared. When the perspective is *ex-ante*, but conducted after the event, a decision must be made as to whether the counterfactual estimates should draw on all information available, or merely the information that would have been available at the time of the policy decision. So, for example, when retrospectively evaluating the impact of a particular merger intervention, should actual post-merger prices be compared with those projected from the simulation model, or should these be adjusted for any post-intervention exogenous shocks, (say a demand shock)? Of course, the latter requires a reworking of the original model, now simulating the impact of that shock, as well as the original intervention. This is discussed further in the context of merger remedies by Davies and Lyons (2007 pp.106-7).

3.2.2 *Previous literature*

Seminal and/or illustrative contributions in the academic literature include Hausman and Leonard (1997), Werden (2000), Nevo (2000), Pinske and Slade (2004), Ivaldi and Verboven (2005), and Peters (2006). Simulation is less common in evaluating the effects of cartel interventions, but see Verboven and Van Dijk (2009).

3.2.3 *Pros and cons*

The major strength of the simulation approach is the explicit use of theory to identify the counterfactual. This facilitates the 'joining-up' of the analysis undertaken at the time of the intervention with any subsequent evaluation of the effects of the policy, and, in turn, provides a clear opportunity for evaluating the assumptions made at the time of the intervention.

However, as is well documented, simulation is very sensitive to modelling assumptions. Sometimes this is a strength, e.g. in revealing how sensitive predictions are to the precise nature of the counterfactual. But sometimes this sensitivity is unhelpful, deriving from alternative specifications between which there are no strong theoretical reasons to choose, e.g. the functional form of the demand curve. These qualifications are well documented elsewhere (see for example Buccirossi et al. (2006), Appendix II.)

Equally important, simulation is better suited for some types of oligopoly models (and therefore markets) than others. The trusted and well-trying workhorses are the Cournot homogeneous product model, and logit type models of product differentiation. Invariably, the emphasis is on price and quantity to the exclusion of innovation, repositioning etc, and possible changes in conduct (relevant to Coordinated Effects.) Buyer power has also proved difficult to incorporate satisfactorily, and simulation of bidding markets is still in its relative infancy. This raises the strong likelihood that evaluation based on simulation is heavily skewed towards certain types of markets, i.e. sample selection bias.

Another source of potential selection bias derives from the heavy demands on data. Many of the seminal studies are based on high quality disaggregated datasets constructed from scanner sources; but, of course, these are typically drawn from a relatively small set of consumer good products (often sold through supermarkets.)

Finally, there is mixed evidence on how well simulation predicts actual outcomes. Ashenfelter and Hosken (2008, p.36) summarise on this count by suggesting that "careful evaluation of their effectiveness seems long overdue".

3.2.4 *Back of the envelope (simple) simulation*

As mentioned, full-fledged simulation is extremely demanding of data and research time. Faced with this, CAs in their own evaluation exercises often employ simplified versions of simulation models (the only viable option if the CA is to evaluate a range of mergers investigated in a given year.) Typically, most of their effort is directed to those mergers which are acceptably described by the Cournot model or simple models of differentiation¹³ In order to calibrate these, extraneous estimates of demand elasticities are required, but there is often a scarcity of good estimates. In these circumstances, Werden (2008) reports that the US typically employs a range of 1 to 1.5 for the aggregate industry demand elasticity, but is not obvious whether this is a reasonable ball-park range. Such inelastic demand will inevitably lead to high predicted price increases. As noted by Bergman (2008, p.394) “There exist amazingly few econometric studies of the price effects of mergers, considering the economic importance of mergers and given that merger effects is a topic that is well suited for this type of quantitative analysis”. Below, we discuss Connor’s painstaking meta-analyses of cartels, but there appears to be no equivalent for mergers. At the very least, there is an obvious gap to be filled by a similar sort of meta-analysis of the industry demand price elasticity – given its key role in most simulation.

3.3 ***Event studies***

3.3.1 *Definition*

An event study draws on financial market data to measure the effect of an economic event on the market valuation of a firm. If financial markets are efficient, then the effect of any event on a firm’s discounted profits will be instantaneously observable through the changes in the prices of its shares. The methodology entails measuring any abnormal returns associated with an event (e.g. the announcement of a merger. Abnormal returns are identified as

¹³ For those markets which do not fit the bill, simple ‘rules of thumb’ tend to be used, e.g. the OFT assumes that an intervened merger (which it cannot simulate) would have raised price by 1%.

the difference between the observed movement in stock valuation and those that would have occurred absent the event.

3.3.2 *Previous literature*

Event studies of mergers typically examine the effect of the announcements of the merger and the CA's decision (i.e. the type of intervention) on the valuation of the merging firms *and their rivals*.¹⁴ For Europe, Duso et al. (2006a) examine the effectiveness of merger control in general, and Duso et al. (2005) focuses more specifically on merger remedies. For Australia, Diepold et al. (2006) apply the methodology to 50 mergers handled by the Australian Competition and Consumer Commission between 1996-2003. Examples of industry specific studies include Simpson and Hosken (1998), for FTC investigations on four retail mergers between 1984 and 1993, and Warren-Boulton and Dalkir (2001), for the Staples – Office Depot merger.

Event studies on cartels typically examine the impact of dawn raids and the subsequent CA decisions, for example Langus and Motta (2007) for EC cartel enforcement, and Bosch and Eckard (1991) for US DOJ decisions. Lübbers (2009) studied the effect of cartelisation (a coal syndicate) in Germany, 1893-1913, where the events analysed were the foundation of the syndicate and two major modifications to the original contract.

More generally, Carletti et al. (2009) analyse the impact of the introduction, or modification in competition laws in 18 jurisdictions. They find that pro-competitive regulatory changes lead to negative abnormal returns on firms' shares, with the exception of the banking sector, where the opposite effect is observed - suggesting that merger control is anticipated to act as a check and balance on prudential control.¹⁵ Other related subjects where event

¹⁴ Eckbo (1985) is an early discussion of why the effects on rivals' valuation provides important additional information.

¹⁵ Other examples of measuring the effect of regulatory change, although not for competition laws, include: Becher (2009), who looks at the effect of US interstate deregulation in the banking sector; and Prager (1989), who examines the effect of the Interstate Commerce Act in the railroad industry. A major problem with using event studies to measure the impact of regulatory changes is the definition of the event window. Regulatory changes are lengthy procedures, typically starting with drafting at administrative level, before the

studies have been used include private antitrust litigation (Bizjak and Coles (1995) and the effects of entry (Whinston and Collins (1992)). However, to the best of our knowledge, the effectiveness of Article 102 enforcement has not been assessed using the event study methodology.

3.3.3 *Pros and cons*

Central to the event study is the assumed rationality of markets, the efficient market hypothesis (EMH), according to which share prices instantly reflect the value to investors of all the relevant information available to them. It builds upon information that is generated by the interaction between a large number of self-interested, independent, rational market agents. This information then can be thought of as the best estimate, given the set of all available information.

If the EMH holds, then the change in the market's valuation of a company will always reflect an unbiased estimate which is both "objective" and quick. This therefore enables a quicker assessment than from using more direct measures such as product prices. The methodology may be particularly attractive to CAs as it tackles the issue of information asymmetry between them and the firms involved in the event. This makes event studies more appealing than analysis of accounting data, which typically suffers from the potential bias that such information is produced by the interested parties. It is also argued that event studies are undemanding of data - the necessary data are easily accessible for listed firms (but see below).

However, the *plausibility* of the EMH assumption is open to question, and many commentators are sceptical. Werden (2008) suggests that the presumption that 'the instant analysis of uninformed investors is more

proposal is carried on to parliamentary committees, brought to the chamber floor, and finally approved by the parliament. During this process, information about the proposal is often made public. As the content of the legislative draft may continuously evolve, it is hard to define an appropriate event window.

accurate than the painstaking work of enforcement agencies with access to confidential documents and data' is not supported by evidence.¹⁶

Turning more specifically to mergers, the main area of competition policy where event studies have been used, Table 2 illustrates, by summarising how expectations about how each of the three constituent events should change the merging parties' market value: at each stage, there is the potential for ambiguity. An increase in valuation of the merging parties may reflect either pro-competitive effects (efficiency gains), or anticompetitive effects (exclusion, dominance, or collusion). Duso et al. (2006b) argue that this ambiguity can be resolved by observing the change in the valuation of rival firms. According to most static oligopoly models, horizontal mergers will result in a higher product price unless there are offsetting efficiency gains; while the former will also benefit rivals, the latter will not. However, as the final row shows, the ambiguity does not completely disappear. For example although an observed negative change in the rivals' valuation may reveal pro-competitive (efficiency) expectations, it may also indicate anticompetitive (exclusion) effects.

Table 2: Expected effects of a (horizontal) merger on merging firms' and rivals' asset values

Source of post-merger gains	Merger proposals		Announcement of investigation		Merger clearance	
	Merging firms	Rivals	Merging firms	Rivals	Merging firms	Rivals
Dominance or collusion	+	+	-	-	+	+ ¹⁷
Efficiency	+	-	-	+	+	- ¹⁸
Exclusion	+	-	-	+	+	-

There is also the possibility that the merger may be interpreted as a signal that other firms in the same market will be subject to merger activities in the

¹⁶ See also Malkiel (2003) for a more general criticism of EMH.

¹⁷ The situation would be different for vertical mergers, which can potentially have foreclosure effects, and would therefore result in negative abnormal returns.

¹⁸ Assuming that competitors cannot free-ride on merger generated efficiencies.

near future,¹⁹ which would increase the market valuation of rivals, resulting in the same sign of change as with collusive and/or dominant outcomes. Although Da Graca (2006) proposes a method for eliminating simultaneity biases, this still remains an under-explored issue. A further criticism is that the methodology does not separate out the market's anticipation of the CA's eventual decision, therefore its reaction to an 'antitrust event' may equally be explained by the market updating its beliefs about a particular antitrust decision, once the uncertainty about the merger investigation is resolved. Duso et al. (2006a) address this problem by using observable merger characteristics to estimate the probability of a particular decision and correct the average abnormal returns accordingly.

The pros and cons are similar for assessing cartel enforcement. However, in this case (characterised by a higher level of CA secrecy), in order to ensure the success of the investigation), there is more chance that the first event (typically in the form of a dawn-raid) will actually be 'unanticipated news' to the market. Also, the theoretical expectations are less ambiguous. The effect of news of the investigation and the ultimate decision nearly unambiguously reduce the valuation of the parties. The one exception is when insufficient sanctions are announced. In this case the negative effect stemming from the elimination of cartel profit may be mitigated by a smaller than expected fine.

An obvious question in general is how well do event studies predict actual outcomes? For mergers, Duso et al. (2006c) compare the results of an event study with an ex post analysis of balance sheet profits. They find that in some cases the abnormal returns measured in event studies are positively and significantly correlated with the ex post measured profitability of the same mergers. Although this provides some affirmative evidence, it is not totally persuasive and further and possibly deeper statistical analysis is required.²⁰

¹⁹ Cox and Portes (1998) claim that this happened in the merger of SBC Communications and Pacific Telesis Group.

²⁰ For example when calculating the ex post profit effect of the merger, the counterfactual used is the median firm in the same market minus the effective rivals. The authors assume that these firms are not strongly affected by the merger, but this assumption needs more evidence especially for large mergers, which may affect the profit figures of other – not

Finally, in spite of the general presumption that event studies are easy to use and data are easily accessed from financial databases, the practical reality is that there are many circumstances when appropriate data are not available. It is, of course, a necessary condition that the parties and their rivals should all be quoted on the stock market, but this is often not the case in small numbers markets especially where firms are small and rivals are scarce. Moreover, very often the parties are large conglomerates and/or multinational, and the market concerned may constitute only a small part of its aggregate activities; where this is the case, it can often prove difficult to identify any effect on the firm's valuation resulting from an event in a small market in a particular country. These practical difficulties are cited by *inter alia* Buccirosi et al. (2006, p.187-8), but perhaps the most telling indictment is provided by Beverley (2008) who attempts to apply the methodology to a sample of 9 UK Competition Commission merger inquiries but is ultimately frustrated by an inability to locate sufficient competitors with traded equities and merging parties for whom the market concerned accounts for a sufficiently large proportion of their activities. It seems very likely that event studies of samples of mergers will suffer from an inherent selection bias for these reasons.

3.4 ***Difference in Differences***

3.4.1 *Definition*

Difference-in-differences (DiD) methods belong to a broad category of methodologies sometimes known (rather unhelpfully) as evaluation methods (see Buccirosi et al. (2006), Appendix I), which also includes natural experiments and matching methods. The basic idea is to evaluate 'performance' before and after an event (or sometimes before, during and after) in the market concerned relative to performance in another similar (control) market, unaffected by the event. The standard DiD application is

directly competing – firms in the market. The main worry however is that their results in general show that the stock market reaction is positively correlated with the ex post development of profit typically in cases where a longer event-window is used. This raises doubts whether the stock market reaction is actually a reaction to the merger or it picks up other confounding effects as well.

typically econometric, in which the performance measure (usually product price) is tracked over time to include the pre- and post-periods in the treatment (e.g. merger) and compared against the same in the control market, with the use of dummy variables. For the purposes of this paper we shall also loosely include in this category the much simpler methods (sometimes used by the CAs) described at the end of this sub-section.

DiD analysis is typically conducted ex-post. The time lag before it is undertaken reflects a trade-off between choosing a sufficiently long post-event period to gain a better grasp of long-term effects, and avoiding a time period which is so long as to compromise the chances of finding a practicable control to emulate the counterfactual. There are also some instances of ex ante DiD analyses, for example for predicting the effect of cartel intervention (which presumes overcharge is removed).

3.4.2 *Previous literature*

Examples of DiD for cartels include Symeonidis (2002), who assesses the impact of the introduction of the 1956 Restrictive Trade Practices Act. In the UK. Levenstein and Suslow (2006) examined the impact of the 1982 United States' Export Trading Company Act by comparing manufacturing industries with and without export cartel exemption.

However, DiD is most typically applied as an ex post tool for evaluating the effect of mergers. Although very popular, there is an inherent problem with using DiD for mergers: if the chosen control group is other non-merging rivals, then how does one allow for the possible externalities of the merger on those rivals? Tenn and Yun (2010) for example looked at the effect of US divestitures in the Johnson & Johnson/Pfizer acquisition. Although the authors concede that their choice of control group (similar brands in the same category) may mean that the treatment (divestitures) had an effect on the control group, they claim that this does not raise serious concerns when simply looking at whether divestitures had an effect at all on prices. We are not convinced that this is indeed the case. Ashenfelter and Hosken (2008) looked at the price impact of 5 US mergers using a set of different control

groups: in the event, their preferred choice was private label products sold in the same industry. Their rationale for this preferred control group was that private labels are only weak substitutes, from the consumer's viewpoint, for the higher quality branded products affected by the merger. At least to the current authors, familiar with the UK supermarket industry, this would seem to be highly contestable. There are similar doubts with Dobson and Piga (2009) use as controls airline routes from different but close airport-pairs.

R.A. Connor et al. (1998) attempted to measure the impact of 112 American hospital mergers, using a database of around 3500 hospitals. Their method essentially follows a difference-in-differences approach, where the treatment markets were those with mergers, and the control was taken from the remaining set of markets.

There are few examples where DiD has been applied to Article 102 cases, deterrence, or other effects of public intervention. An early equivalent is Shaw and Simpson (1986) analysis of the erosion of market dominance in markets where the Monopolies and Mergers Commission was active (markets not investigated by the MMC were the control group). There are also some examples where DiD has been used to assess the impact of other types of regulatory intervention. For instance, Cooper et al. (2010) used a DiD estimator to test whether the introduction of hospital competition in the English NHS in January 2006 has led to increased efficiency.

3.4.3 *Pros and cons*

The difference-in-differences approach enjoys an (at least superficial) appeal that it uses observed data from the relevant product market (i.e. what *actually* happens) in comparison with a control (which if carefully selected), where the event does not occur. Thus the counterfactual is not dependent on untestable, maybe restrictive, theoretical assumptions.

However, the other side to the coin is that the methodology, is inevitably atheoretical. Since much depends in evaluation on the nature of the counterfactual this means that a key part of the methodology – identifying an appropriate control group – is also atheoretical. As such, there is a danger

that the choice of counterfactual (control group) is constrained by ‘what is out there’, i.e. the best of a set of alternatives, none of which is entirely appropriate. When choosing the control market the ideal is that it is characterised by the same supply and demand shocks as the treatment market, which makes it possible to filter out the net effect of the analysed event by controlling for these supply and demand shocks. But Simpson (2008) warns about the danger of this assumption, claiming that even the same supply and demand shocks may influence prices differently in the two markets.

On a practical level, the control should have a sufficient number of members and time observations to emulate the random variation which would occur in the treatment group post intervention that is unrelated to the intervention itself. Most research however just assumes that the only difference between the treatment and the control groups is the treatment²¹. Meyer (1994) identifies some of the limitations of this approach, such as omitted variables, trends in outcomes, measurement error, simultaneity, selection bias, omitted interactions, etc.

Overall, difficulties in identifying a satisfactory control must raise the worry that DiD may only be useable for a fairly small sample of markets, and a sample which may not be representative of the population as a whole..

3.4.4 *Simple/ Back of envelope*

Simple variations on DiD are sometimes used by CAs in ex-ante decision-making and/or evaluation. This might involve identifying similar geographic or product markets where similar events occurred in the past.²² The literature on estimating cartel overcharge includes many instances (academic papers and CA evaluations) where the overcharge is simply equated to the difference

²¹ See Buccirosi et al. (2006) for a discussion of the selection bias which will occur if the set of unobservable characteristics which affect the decision to merge also affect the performance of the parties post-merger.

²² Informal examples of this approach can sometimes be found in CC’s evaluations of merger interventions and market investigations.

between price during cartel with pre- and post-cartel prices in the same market, or in similar non-cartelised markets.

3.5 **Others**

There is also a variety of other, mainly qualitative, approaches which we discuss more briefly since they lie largely outside the subject area of the present paper.

3.5.1 *Aggregate Economy Studies*

Some more aggregate studies employ cross-country or panel data, attempting to identify the effects of competition policy on macro or sector aggregates, e.g. price-cost margins, GDP, productivity²³. These are discussed and well summarised by Bergman (2008, section 3.4 'Macro-level studies'), who provides some of the key references. Necessarily they confront daunting measurement issues, often relying on the construction of subjective indexes of the 'severity of competition policy' which are allowed to differ across countries. Other econometric problems are familiar in any international comparisons based on production functions or related concepts, e.g. identification, simultaneity and the requirement that the underlying functional forms are stable across countries.

3.5.2 *Follow up surveys of the parties/rivals/customers/suppliers*

It is quite common for CAs to undertake or commission reviews of their previous cases (especially mergers) based on follow up questionnaires and/or interviews with the interested parties and related firms²⁴. These often provide invaluable insights but are inevitably prone to a number of potential limitations: low response rates, respondent bias, the parties often have short corporate memories, and in their view, interventions can often be overtaken by other subsequent and more important events. Clarke et al. (1998) includes a number of examples of the latter.

²³ See Bergman (2008) for a discussion and selected references.

²⁴ A recent example for UK merger policy is Deloitte (2009)

3.5.3 *Expert commentaries on specific cases*

Book collections of expert (economic) commentaries on specific, often well known anti-trust case studies, such as Kwoka and White (2004) and Lyons (2009) for the US and EU respectively, can also be viewed as contributing to the evaluation literature, but for obvious reasons, it is difficult to generalise from a set of heterogeneous non-randomly selected small samples, especially with respect to the evaluation methodologies used (if any).

3.5.4 *Court decisions*

Some studies have assessed the quality of CA decision-making by the frequency of court appeals and/or the success rates in those appeals (see Bergman (2008, p.389-91) for a brief survey.) For certain purposes this may be a valuable extra source of (presumably well-informed and objective) evaluation, but obvious limitations include the likelihood of selection bias, and the fact that court decisions will sometimes involve judgement on the correctness of legal process rather than economic substance.

3.5.5 *Surveys of peer/practitioner opinions*

Certain high-profile annual reviews provide an alternative approach, based on peer review evaluation of the performance of different CAs around the world, e.g. the Global Competition Review and OECD country reviews.²⁵ These enable international comparisons over time at the aggregate level, but these lie outside our current remit, being based on subjective opinion rather than quantitative methodologies.

3.6 ***Preliminary conclusions***

Rather than argue for the primacy of any one methodology over the others, we suggest, uncontroversially, that different methodologies are better suited for some purposes than others. When evaluating a specific intervention (e.g. a particular cartel or merger), the event study and DiD may be appropriate,

²⁵ http://www.oecd.org/document/43/0,3343,en_2649_33759_2489707_1_1_1_1,00.html

assuming availability of the necessary data; expert retrospective case study commentaries can also be attractive. On the other hand, when evaluating the quality of the CA's decision-making - analysis, data collection etc, the event study and DiD are less attractive, since neither is based on explicit theory. Here, simulation scores more strongly, since it forces us to articulate the model, make specific assumptions (about entry, capacity etc.) and to specify the data to be collected – precisely the purpose of the evaluation, i.e. the accuracy of assumptions, rather than the magnitudes of price increases. When aggregating the overall impact assessment of policy over full range of cases handled in a given year it is noticeable how often the CAs turn to simple simulation methods, supported where necessary by rules of thumb. Indeed, it is not obvious how event studies and DiD could be applied over a large number of different cases due to their much greater demands on time and data.

Having said this, we believe that the most important messages to emerge from reviewing this literature concern not so much the relative merits of specific methodologies, but more with the problems which seem to afflict all of them. Here we highlight four general issues, each meriting future research.

3.6.1 *Paucity of evaluation of Article 102 cases*

In both the academic literature and the CAs' own evaluations, there has been far less work on Article 102 than on mergers and cartels, and this raises a number of questions. Is it that existing methodologies are not suited to Article 102 cases, or is it simply that CAs prosecute far fewer Article 102 cases? If the latter, is this because competition policy is more effective in deterring the violation of Article 102, or are CAs less adept at uncovering them, or less inclined to intervene? If it is true that CAs are significantly less active and/or successful in this area, is this because cases are more difficult to prove, and that CAs are tempted towards the softer areas of merger enforcement? Or is it that net harm in most 102 cases is considered to be too small to merit much priority? Of course, these questions are of wider importance than just for the evaluation agenda, and improving our understanding of the reasons should be high on the research agenda.

3.6.2 *Selection bias*

Each of the broad methodologies discussed here tends to be better suited for some types of market than other – but for different reasons. Simulation is most easily and most often applied to unilateral effects horizontal mergers, where the product is homogenous or can be described with a standard model of product differentiation; but coordinated effects, innovation markets and vertical mergers pose greater problems and have attracted far fewer studies. Event studies are impracticable in markets of non-quoted or highly conglomerate firms, and where rivals are scarce. DiD requires the existence of genuinely unaffected firms in the same market or the coincidence of a similar control market. In each case therefore, we cannot rule out the possibility of pervasive selection bias in the make-up of samples on which many evaluation studies have been based.

3.6.3 *The need for longer-term studies*

Invariably, the above methodologies model impact only over a relatively short-term, while a longer term perspective might often be desirable. Consider merger simulation for example - here, the essence of the evaluation is a comparison of two equilibria (with and without) which are both conceptually timeless. Typically, it is assumed that an evaluation period of one or two years after the event will be sufficient. Despite an awareness that there may be longer-term consequences (e.g. efficiency savings, new entry, brand repositioning), these are rarely incorporated into the modelling. For instance, very often, but not always, simulation abstracts from the possibility of efficiency savings and in the short-run, this may be appropriate. However, where efficiency gains may only emerge say 2-3 years after merger,²⁶ as the benefits of restructuring start to appear; if so, what is essentially a static counterfactual becomes less less appropriate. The same is often true for DiD, in which the post event period is typically short because the interventions examined are usually fairly recent. It might be that event studies avoid this

²⁶ Which seems to be the average pre-merger expectation by merging parties, see Ormosi (2010b)

problem insofar as financial markets are supposed to anticipate all future events, but this does strain one's faith in the EMH.

There is also another, more general, concern about the time dimension which applies to any one-off evaluation (no matter what the methodology) – that it runs the risk of closing the story prematurely. The wider IO literature (both theoretical and empirical) suggests various possibilities for how a specific event might trigger a sequence or chain of subsequent events - each of which might be evaluated independently, but which are in reality clearly path-dependent. This means that when evaluating an isolated intervention, we might ignore the longer-term consequences. For example, the literature on endogenous mergers alerts us to the possibility that, if merger A is cleared, this makes a subsequent merger B more or less likely. Similarly, in failing firm merger cases, the consequences of intervention may include subsequent alternative merger proposals by other parties. There is also case study evidence which suggests that sometimes when some practice is prohibited, it is replaced by others. Clarke et al. (1998) cite examples from the UK where firms responded to the prohibition of one form of vertical restraint by introducing an alternative form (Ice Cream), or where a prohibited restraint was replaced by full-fledged integration (Travel Agents). More generally, it has long been recognised that horizontal mergers may sometimes be an alternative to cartelisation: Symeonidis (2002) shows that cartel legislation in the UK in the 1950s provoked a subsequent merger wave through the 1960s. In other words, it matters what happens next, and next should sometimes be interpreted as long-term and not too narrowly. Some evaluation studies have acknowledged this, albeit indirectly, by considering sequences of cases, for example Sabbatini (2008) for Italian baby-milk; Pinske and Slade (2004) for a sequence of mergers in UK beer and Nevo (2000) who includes various different mergers in his simulations, but these are the exception rather than the rule.

3.6.4 *The counterfactual*

A central issue running throughout most antitrust analysis is the choice of the counterfactual – what *would* have happened, had some event, policy or

practice not occurred? This is not just a concern for academic economists, it also occupies the attention of the courts and lawyers, for example when attempting to quantify damages²⁷. It has both conceptual and empirical dimensions – which counterfactuals are theoretically tenable, and how do we calibrate them with plausible estimates of key parameters?

Clearly, the counterfactual has a crucial role in the present context - any evaluation methodology must necessarily entail a counterfactual, even if it is sometimes only implicit. In effect, much of our discussion so far can be interpreted as questioning the nature of the counterfactual assumptions employed by different methodologies - both in specific cases, but also more generally – are some methodologies intrinsically more counterfactual-aware than others?

Simulation, by its nature, places the choice of counterfactual conspicuously at centre stage: a specific oligopoly model is selected, and this immediately reveals the nature of the assumed counterfactual equilibrium. Similarly, it must be calibrated very transparently with key parameter estimates.

In DiD, the control plays the role of the counterfactual – for example, what would have happened in the UK, had RPM for books not been repealed, might be captured by what actually happened over the same period in Germany, where it was not repealed (Davies and Olczak (2008)). Here, the choice of counterfactual is less theoretically driven, and the strength of the methodology rests on whether there really is a control (with adequate data) which is sufficiently similar. In practice, data expediency may sometimes distract attention from how closely this condition is met.

While it is less common to think of the counterfactual in the typical event study, implicitly it is still there - captured by whatever comparator share price index the practitioner uses to compute abnormal returns. Here there is a trade-off between using a general index, which is less likely to be sensitive to

²⁷ For a useful cross-discipline of the use of counterfactuals in antitrust and mergers, see Colley and Marsden (2010) at: http://www.bicl.org/files/5106_the_use_of_the_counterfactual_in_antitrustv2.pdf.

market-specific exogenous events, and more customised sector-specific indexes, which may not be truly independent of the event at issue.

While there are various other gaps in the literature deserving of further research²⁸, we confine the remainder of this paper to further discussion of the counterfactual and selection bias.

4. Comparing counterfactuals: cartels as a case study

An ideal approach for comparing the alternative evaluation methodologies would be to assemble an (ideally very large) random sample of cases, and attempt to apply all the methodologies to all cases in the sample. This would help both in assessing their relative practicabilities and in identifying any systematic differentials in their estimates. While it is likely that some CAs and advising consultancies do sometimes conduct parallel assessments during their conduct of a particular case (experimenting simultaneously with, say, an event study and a simulation), these do not appear in the published literature of course. More generally, attempting such a task across a large sample of cases would be difficult, and, in the event, has not occurred to date.

In this section therefore, we explore the feasibility of an alternative, second-best, approach. We examine an existing database already in the public domain, on cartel overcharge, constructed by Connor and various co-authors. It is a meta-analysis, drawing together the results from a very large number of primary studies by previous authors. It includes 800 observations, taken from nearly 400 cartel episodes across the world. This is a only second best to our ideal, because the overcharge estimate for each cartel has been computed typically using only a single methodology - for very few of the cartels do we have alternative estimates using different methodologies. This means that relatively higher power matched sample tests are impossible, but the sheer size of the sample should nevertheless permit reasonably meaningful tests between sample means for different methodologies.

²⁸ Notable amongst the other gaps are the absence of any large scale meta analysis of the price raising effects of mergers, and of a practitioners' encyclopaedia of price elasticities.

The version of the version of the database we use here is summarised in Bolotova and Connor (2005) (C&B). The main purpose of their paper was to examine the determinants of overcharge, and the methodologies used to derive the estimates was merely a minor side issue – to be controlled for. But for us here, this is the main focus of attention.

C&B classify the estimation methods used in the primary studies into eight broad groups. Of these, we discard about 30% of cases which belong to three of the groups since they do not correspond to any of our methodologies above: sundry/unknown ‘historical records’, and ‘cost calculations’. Three of the retained categories correspond roughly to component parts of DiD: PBEFORE (PAFTER) are comparisons between the within-cartel-period price and the price immediately before (after) the cartel period, and YARDST are relative to prices in yardstick ‘analogous markets that were believed to be free from cartelization’. The PWAR, ‘price-war’ method compares prices within the lifetime of a cartel, but distinguishes sub-periods with and without price wars. Finally, the ECON, ‘Econometric’, category is very broad based, but appears to include estimates based on simulation-type methods. C&B (2006, p.1120)

Table 3 shows the proportions of cases in each category and our calculation of their sample mean overcharge estimates²⁹. They fall within a broadly similar range (23% - 45%), but with an intriguing ranking – price wars being the highest and price-after the lowest.

Table 3: Methods of estimating overcharge

	%	Mean overcharge
Price before	33	29
Price after	11	23
Price war	2	45
Yardstick competition	11	39
Econometric	15	31

²⁹ C&B do not report these means directly in their paper, but we have recovered them from their reported preliminary regression of overcharge solely against dummy variables for each category – in that case, regression coefficients coincide with the sample mean values.

A simple null hypothesis is that each sub-sample (i.e. for each methodology) is drawn from the same population distribution of overcharges (i.e. no systematic tendency for specific methodologies to over- or under-estimate. Abstracting from systematic measurement errors³⁰, if the null is rejected, this implies some systematic difference between methodologies in how they measure the counterfactual. In common with others working in this area, C&B refer to the counterfactual as the 'but for' price. However, this is not without ambiguity: is the 'but for' price the 'competitive' price or the market price that would have obtained under a set of identical conditions (including market structure) except for the existence of the cartel? Contrary to some discussion, the correct answer would appear to be the latter and this leaves open the possibility that a defendant might argue that damages should be only moderate because the 'but for' should be the tacitly collusive price.

To pursue this further, consider the relative magnitudes of the sample means in the Table. For this purpose the 'econometric' category is discarded because it includes a heterogeneous variety of implicit oligopoly models (counterfactuals) across papers. But the remaining categories do lend themselves to the following speculative interpretation. Suppose, merely for the sake of the argument, that the yardstick method typically identifies the 'competitive' outcome as its counterfactual. In that case, we know that on average cartels set a price 39% higher than the competitive level. This then allows us to interpret the counterfactuals identified in the three other categories as follows:

- during price wars the cartel price falls to 45% below the non-price-war cartel price, i.e. 6% (= 39-45) *below* the competitive level.
- the price before cartel is typically 10% (=39-29) *higher* than the competitive level
- the price after cartel is typically 16% *higher* than the competitive level

³⁰ Or that the researchers' choice of estimator is not independent of the actual cartel price

Taken at face value, this implies the following typical time path for price. Even before the cartel episode, price starts from a supra-competitive level, then rises during the cartel period, before falling post-cartel. However, the post-cartel price remains at a supra-competitive level. On the other hand, in the typical price war within the cartel period, price falls substantially below the competitive level (implying harsher punishment than Nash reversion).

Leaving aside whether these sample means are statistically significantly different, these *point* estimates imply two conclusions:

Of course, this is highly speculative - it is based on simplistic interpretation of sample point estimates without attention to statistical significance. This is deliberate because of doubts about the quality of at least some of the estimates in Connor's database (Ehmer and Rosati (2009)). I have re-worked this database to exclude all estimates not meeting a variety of selection criteria, and this leads to a drastic pruning in the sample. Future work is anticipated which will investigate whether the estimates reported in Table 3 OXERA (2009, p.90-92) are robust to such a pruning, and whether or not they differ significantly.

In the meantime,, these exploratory results are of some interest for our understanding of typical cartel time paths. Regarding post-cartel prices, Kovacic et al. (2007) report that "In analyzing prices in the post-plea period, which is a period of potential tacit collusion, we find that vitamin products with two conspirators continue as if the explicit conspiracy never stopped, while products with three or four conspirators return to pre-conspiracy pricing, or lower, quite quickly". Sabbatini (2008, p.501) concludes his study of Italian milk cartels by noting that "Cartels don't break simply because you uncover them as they may continue as 'well established' rules". This possibility, that shared understandings (tacit collusion) might often survive cartel busts is also implied by some of the results emerging from the experimental literature (e.g. Fonseca and Norman, 2010). Harrington (2004) raises a slightly different argument: the parties might moderate price reductions mindful of the signal that this would send to the courts regarding the magnitude of the previous overcharge. Regarding pre-cartel price, Harrington (2006) points to various

instances where cartel formation is preceded by a significant price decline, and suggests that this might be accounted for by, variously, cost/demand shocks, entry and/or capacity expansion, and the breakdown of tacit collusion. However, as Harrington notes, this question has so far attracted little or no theoretical examination.

5. Selection bias and deterrence

5.1 *Some potential sources of selection bias*

Inevitably, evaluation studies focus on documented cases in the public domain, and amongst these, usually the ones where an intervention has actually occurred or been seriously contemplated. As noted frequently in the literature, this raises a variety of doubts about whether such a sample can provide an unbiased estimate of the benefits of competition policy. It also leaves open the problematic issue of how to treat the deterrent effect. This section draws together various strands in a fairly disparate existing literature on selection bias and deterrence into a coherent framework which highlights the most pressing questions for future research.

We start by recalling two conclusions from section 2.6.

- (i) Different evaluation methodologies are less practicable for some types of markets and interventions (e.g. event studies for markets populated by unquoted or conglomerate firms.) This relates to potential selection bias *within* the set of detected cases investigated by CAs.
- (ii) Some policy areas, notably Article 102, have been evaluated much less frequently than others. This may reflect a low incidence of abuse in the population, but it also raises the possibility of selection bias due to lower detection rates, or reluctance by CAs to intervene.

The previous literature contains numerous other instances of why evaluation is susceptible to selection bias problems; the following examples illustrate.

- (i) In the empirical literature on **cartels** (e.g. duration and overcharge) it is widely understood that the samples analysed may be intrinsically

biased because they are drawn exclusively from detected investigated cases, which may not be representative of the unknown population of undetected cases.

- (ii) When the rigour of **merger policy** is evaluated using a sample of *unchallenged* mergers, Carlton (2009) points to an easily overlooked bias. He poses the question: ‘suppose we observe that the mean price increase in a sample of unchallenged mergers is negative, can we deduce that the CA is sufficiently (or even over-) strict?’ The answer is no because, even with a lax CA (inclined to Type I errors), such a sample will include ‘good’ mergers alongside any incorrectly permitted ‘bad’ (price increasing) mergers. He suggests that any such evaluation should more properly compare outcomes with the CA’s predictions at the time of its decisions to estimate the systemic bias in enforcement.
- (iii) It is widely acknowledged that the beneficial **deterrent** effects of competition enforcement are likely to be considerable, probably far outweighing the measurable benefits of the actual caseloads of CAs. It follows that any evaluation of the benefits of policy based only on investigated cases may be a serious underestimate, probably by an order of magnitude; however, we know remarkably little about the magnitudes of this global underestimate.

The remainder of this section introduces a classification scheme designed to structure the various dimensions of potential selection bias and highlight directions for future research.

5.2 ***How much of the iceberg lies below the waterline***

Figure 1 suggests a simple classification scheme to describe the full distribution of all potential competition cases in the population³¹. In the event, some of these are deterred and never occur; amongst the undeterred cases,

³¹ This notion of a well-defined population of potential cases is not unproblematic, given that it is to include all deterred anti-competitive practices. Applying such a classification, analogously, to the law criminalising murder would require one to quantify the number of murders that would be committed were the practice not illegal.

some will be detected but some not; and then, within the detected set, some are investigated and some are not.

This classification is used to illustrate some of the selection problems which would need to be resolved if one is to extrapolate from what we know (the investigated cases) to what we don't (the deterred, undetected and un-investigated parts of the population.) We proceed in two steps, considering first the relative sizes (frequencies of cases) of the classes, and then the potential heterogeneity between the classes (measured by, say, mean expected harm), which reflects any potential selection bias.

Figure 1 A general classification of potential competition cases

Deterred ω	Undeterred $(1-\omega)$	
	Undetected $(1-\phi)$	Detected ϕ
	Uninvestigated $(1-\sigma)$	Investigated σ

5.2.1 Relative frequencies

Denoting the conditional probabilities by: deterrence rate (ω), detection rate (ϕ), and investigation rate (σ), it follows that the sample of investigated cases:

Property 1: represents only a (perhaps very) small proportion, $(1-\omega)*(\phi)*(\sigma)$, of the population of all potentially relevant cases.

Property 2: fails to capture any beneficial deterrent effect.

Property 3: fails to capture the 'missed opportunities' represented by harmful cases that are either wrongly un-investigated or undetected.³²

The first two are trivial, but the third is often overlooked in evaluation studies.

³² We abstract for the moment from possible offsetting effects, e.g. deterrence of welfare-enhancing mergers, and incorrect Type II intervention decisions.

5.2.2 A back of envelope quantification

It remains for future work to establish how rigorously these three properties might be quantified, but to illustrate how information already in the public domain might be employed, we draw on a rare qualitative study by Deloitte (2007), commissioned by the OFT. This involved interviews/telephone surveys of lawyers, economists and companies. The key findings (2007, pp. 7-12) on deterrence are shown in Table 4. From the survey of legal advisers, they suggest that, for each merger blocked or modified by the CA, there were at least another 5 proposed mergers, that were abandoned or modified on competition grounds. There 'multiplier' was slightly smaller for potential Article 102 cases, but higher for commercial agreements (Article 101). According to their survey of the companies themselves, the reported multipliers were all considerably higher.

Table 4: Deterrence multipliers

	Legal survey	Company survey
Mergers	5:1	–
Cartels	5:1	16:1
Commercial agreements	7:1	29:1
Abuses	4:1	10:1

Turning to undetected cases, Deloitte report that the number of 'under the radar' (i.e. undetected by the OFT) mergers was at least as high as the number which are blocked or modified following intervention by the UK competition authorities³³³⁴.

³³ Strictly speaking the survey reports that the number of undetected problematic mergers was at least as high as the number of investigated ones, in which case the total detection rate may be much lower than 50%.

³⁴ Another study, by the Dutch competition authority was designed to estimate the scope of merger deterrence (NMa - Research into the Anticipation of merger Control, October 27, 2005 http://www.nmanet.nl/Images/Rapport%20Research%20into%20the%20Anticipation%20of%200merger%20control_tcm16-86578.pdf). Law firms were asked to report ideas, initiatives, and notifications of mergers. The responses showed that around 58% of the ideas became an initiative, and around 68% of the initiatives progressed into a notification, implying a 60% upper bound for frequency-based deterrence. The report also reveals that according to the law firms' estimates, there is one case per year in each law firm where the merger does not

These estimates should be interpreted with considerable caution (see the qualifications stressed in the report itself), but taken at face value, and arbitrarily assuming a value of 0.5 for σ the investigation rate (in principle, this could be computed from the CA's own records), we can back out estimates of the probabilities as in Table 5, and then the population frequencies as in Table 6.

Table 5: Deterrence and detection probabilities

	ω ³⁵	ϕ
Cartels	0.56	0.17 ³⁶
Mergers	0.56	0.5

The three above properties can then be quantified as: (i) the investigated sample is only about 10% of the full population of potential mergers and less than 4% of potential cartels;³⁷ (ii) there are five (fifteen) times as many deterred as investigated merger (cartels). On the other hand, (iii) there are three (ten) times as many 'missed opportunities' (undetected or un-investigated) as investigated mergers (cartels) cases.

Table 6: Calculated frequencies for categories

	Cartels	Mergers
Deterred	0.556	0.556
Undetected	0.368	0.222
Un-investigated	0.0377	0.111
Investigated	0.0377	0.111

proceed exclusively for its anticipated anticompetitive effects (6% of the total number of cases), and around two cases per year (12%) where a merger proceeds in an amended form. These figures are to be treated with caution bearing in mind the small sample (only 15 firms), but they suggest a much lower deterrence ratio than the OFT study.

³⁵ This uses the Deloitte finding of 5 deterred cases for each investigated case, $\omega = 5(1 - \omega)\phi\sigma$, from which ω can be calculated using the estimations for ϕ , and the arbitrarily chosen 0.5 for σ .

³⁶ Based on Ormosi (2010a) discussed below.

³⁷ Even if we assume that all known cases are investigated (i.e. $\sigma = 1$) these ratios will still only be 22% and 10% respectively.

Needless to say, these estimates are presented for merely illustrative purposes – they show the implications of the Deloitte study in particular, but we reserve judgement on whether or not they might be credible.

5.2.3 *Heterogeneity between case classes (selection bias)*

It should also be stressed that these calculations relate only to the frequencies of each class – they do not quantify the relative *amounts* of deterred harm or missed opportunities. This would only be true if the expected harm of cases in all categories were identical. However, the possibility of selection bias raises suspicion that expected harm will differ systematically between the classes of case in fig. 1. It is to this question that we now turn³⁸.

First, it is important to note that the distinction between categories is not clear cut, and that the frequencies and make-up of the classes are endogenous to the policy decisions of the CA. Consider mergers as an example. In jurisdictions where there is a compulsory pre-merger notification regime (most of the world) a regulatory threshold demarcates the cases that do not have to be notified. Above the threshold there are 3 types of case: (1) those where there is only a trivial preliminary screening, (2) phase I cases; (3) phase 2 cases. Figure 2 positions these subsets on a line which represents increasing intensity of investigation. Although all cases above the threshold theoretically have to be investigated (i.e. the CA has no discretion), it is arguable whether a simplistic screening of the merger (or even a phase I review, as discussed below) should qualify as an investigation. So the question for mergers is: where to draw the line on Figure 2, below which we consider cases as un-investigated.³⁹

³⁸ The OFT (2010c) p.22. speculates about applying a multiplier to account for the deterrent effect of its enforcement activity, where the value of the multiplier is derived from the Deloitte survey. This method implicitly assumes that the distribution of case types in the investigated and deterred classes are similar.

³⁹ In non-mandatory pre-merger notification regimes (such as the UK) the situation is different as the CA has more discretion whether to investigate mergers.

Figure 2 **The level of the investigation (increasing from left to right)**

Under threshold Trivial screening Phase I Phase II

The possibility that CA discretion may have an important distortionary impact on evaluation is not confined to just mergers, and the choices made by the CA are an integral part of its policy and should therefore also form part of any evaluation. The key idea here is that a CA is confronted with a set of potential cases where it *could* assemble sufficient information to undertake a full investigation (and subsequently possibly intervene) but, in the event, it chooses not to pursue them all.

This possibility cannot be excluded as one possible explanation for the previously highlighted scarcity of evaluations of Article 102 cases - perhaps CAs choose to intervene less frequently in such cases because successful prosecution is more uncertain⁴⁰. A related issue is highlighted in the cartel setting by Chang and Harrington (2010) who discuss how the CA's caseload can affect cartel births and deaths. They model the outcome of a cartel fine as the result of three events: (1) detection, (2) decision to investigate, and (3) successful investigation. They assume that detection is exogenous, but posit that the decision to investigate is a matter of CA choice, and success is negatively correlated with the CA's workload, that is, it is determined by (1) and (2).⁴¹

Of course, given finite resources, coupled with a need to substantiate its impact, it is rational for any CA to pursue the 'easy options'; i.e. easier cases at the expense of more difficult cases (for which the probability of 'success' is lower or where there is greater uncertainty.) This may have serious

⁴⁰ Alternatively, it may be that the CA is totally unaware because abuses of Article 102 are simply less transparent.

⁴¹ If this is so, they show how the introduction of a leniency programme will lead the CA to pursue a less aggressive policy towards non-lenency cartels.

implications for selection bias – both for policy in general and in specific parts (e.g. Article 102). We believe that this issue merits further research.

A related danger is highlighted by Neven and Zengler (2008), who suggest that the evaluation programme itself may introduce an additional motive for distortionary discretion in CA conduct: “Faced with simplistic assessment, authorities may be tempted to be overly interventionist, to spend too many resources and to ignore relevant information” (2008, p. 477.) This returns us to the question of what is the purpose of evaluation, and who undertakes it. If evaluation is driven by external accountability (to verify whether the CA delivers its objectives) and especially if undertaken by the CA itself, the CA is prone to fall into the trap identified by Chang and Harrington (2010), that it will not seek to maximise deterrence, but focus on something that is observable/measurable (e.g. the proportion of Art.102 cases that are won, or the number of cartels detected). This can have very important feedback effects, not just for evaluation, but also for success in achieving the ultimate objectives of competition policy.

5.2.4 Existing academic literature on detection and deterrence

The task of calibrating the three parameters and quantifying likely magnitudes of associated selection biases is formidable. In our opinion (Table 7) the investigation rate is potentially the easiest to estimate, and could be computed from the CA’s own records. But detection and especially deterrence are more problematic.

Table 6 The increasing difficulty of estimating the three probabilities

Probabilities of interest		Source of data	Literature examples
Observed	Investigation rate	CA's records	–
Unobserved	Detection rate	Atheoretical empirical methods	Bryant and Eckard (1991) Combe et al. (2008) Ormosi (2010)
		Theory-based empirical methods	Miller (2009) Harrington et al. (2009) Cheng et al. (2010)
		Surveys	OFT (2007) Beckstein and Landis-Gabel (1982) Aaronson (1992) Baros et al. (2010) Clougherty et al. (2009)
Unobserved	Deterrence rate	CA's records	Harrington et al. (2009) Cheng et al. (2010)
		Theory-based empirical methods	
		Surveys	NMa (2005)

The detection rate is not directly observable of course, but a body of emerging research offers some promise that indirect estimation may be possible using relatively atheoretical statistical methods. This research also has implications for the rate of deterrence, but may involve additional theoretical modelling of the creation of mergers, cartels and other business practices. More work is clearly needed – for example, on whether observed CA activity and latent deterrence are complementary or substitutes⁴², More generally, there may be fruitful avenues of research drawing on the wider legal deterrence literature. For instance, the critique by Donohue and Wolfers (2006) of the common mistakes made in the statistical analysis of the deterrent effect of the death penalty might be helpful. Within the existing academic economics literature, a number of papers suggest potentially fruitful strands of future research on both cartels and mergers.

⁴² This might be approached statistically by comparative analysis of high and low deterrence authorities.

5.2.4.1 Cartels

A commonly cited statistic, originating from the early work of Bryant and Eckard (1991) is that, in a given year, 13-17% of cartels are detected. More recently for Europe, Combe et al. (2008) confirm a similar result. However, this statistic should not be confused with what we refer to here as the detection rate: these estimates were of conditional probabilities – *conditional on the cartel eventually being detected*. A discussion paper by Ormosi (2010a) proposes a different method (a capture-recapture analysis) to estimate the detection rate, which does not require the assumption that detected cartels have to be a random sample of all cartels. This method allows the analysis of the detection rate over time, and suggests that the European Commission's detection rate has improved to 15-20% in the most recent years. However, this still leaves open the question of how many cartels are never detected and how many are deterred.

Four recent articles address this question by modelling the unobserved process of cartel formation and breakdown, in order to draw inferences about the deterrent effect of enforcement. Harrington and Chang (2009) (hereafter H&C) and Chang and Harrington (2010) (hereafter C&H) use similar approaches to establish the relationship between what we can observe (detected cartels) and what is hidden. H&C set up a system for cartel formation and death and identify the links between changes in this system and a change in observable factors such as the number of detected cartels. They argue that observing the *number* of discovered cartels can only permit ambiguous speculations about the effect of enforcement⁴³, and that the duration of cartels is a better indicator of the effect of a policy change. This is because a more stringent enforcement policy will result in a set of detected cartels which have longer duration because unstable, and therefore shorter, cartels are now no longer formed in the first place. However, although duration of detected cartels is easily observed, we still face the familiar

⁴³ The rate of cartel formation and detection both decrease with the size of penalties, but changing the effectiveness of enforcement (e.g. the introduction of a leniency programme) has an ambiguous effect.

problem, that the set of detected cartels is most likely not a random subsample of all cartels.

In C&H, the model is changed to allow for the introduction of a leniency programme, which affects not only the probability of detection in cases where one of the firms comes forward, but also the probability of discovery for other cartels as well.⁴⁴ They find that leniency has a deterrent effect because it reduces the cartel rate and also improves the CA's probability of winning non-leniency cases.⁴⁵

C&H also develop the model to incorporate the amount of resources available to the CA, which is a primary determinant of the success of its cartel enforcement. This affects the rate of cartel formation because it changes the probability that firms assign to being caught, investigated, and convicted. This in turn determines the number of cases handled by the CA and thus the probability that they can get a conviction.

As mentioned earlier, an interesting assumption in C&H is that the CA does not seek to minimise the cartel discovery rate because it is not observable, much less verifiable. Rather, it strives to maximise the number of cases won. This would mean that enhanced deterrence is not in the interest of the CA if that were to reduce its caseload – and consequently the number of cases won. Although this might seem a questionable assumption, there is some intuition to support it: deterrence is difficult to measure and therefore include in any performance measure of the CA. If the CA's budget is based on previous years' performance, which excludes deterrence, then the assumption becomes more plausible. However, a less plausible assumption in C&H is that the CA has a discretion which cartels to investigate. In reality, it seems unlikely that the CA is in the position to turn down cartel cases, although it is

⁴⁴ Other literature (such as Hinloopen and Soetevent (2008) and Bigoni et al (2008)) had assumed that enforcement through non-leniency means is unaffected by the introduction of leniency measures.

⁴⁵ The paper also shows that an increasing cartel rate is now possible for the set of industries where cartels are more stable. As the introduction of leniency programmes weakens enforcement for non-leniency cases, industries with more stable cartels will be characterised by increasing cartel rates.

likely that 'too many' cartel investigations might seriously constrain the resources it is able to devote to individual cases. Another problem is that, if non-leniency cases are more likely to remain un-investigated, this might distort firms' incentives to come forward.

Miller (2009) empirical work, which assumes a first-order Markov process in the expected number of colluding or competing industries in the next period, also attempts to measure the impact of the introduction of a leniency programme. He uses a database of indictments and information reports filed in the US between 1985 and 2005. His results show that the total number of cartel discoveries increased around the time of the introduction of the leniency programme but then fell back below pre-leniency levels. He argues that this is consistent with enhanced cartel detection and deterrence. However, following C&H, this could also simply imply that, with leniency in place, the CA expends much less effort in discovering non-leniency cartels, and this will reduce the number of challenged cartels in the long run.

Hyytinen et al. (2010) use a variation of the models proposed by H&C and C&H to exploit the characteristics of a Hidden Markov model which links observed and unobserved processes. Their database is for cartels in Finland between 1951 and 1990, an era when cartels had to be registered but typically were not illegal. The fact that cartels were legal means that the authors only use two dimensions of their proposed model (as regulatory discovery and leniency applications are not applicable). There are two restrictive shortcomings of their model: (1) it only allows leniency applications to occur when a cartel is unstable; (2) the detection probability only shows up in states where the industry was cartelising at the beginning of the same period.

An important limitation of all four models is that markets are assumed to be dichotomously either cartelised or competing. This therefore abstracts from the possibility that a cartelised industry might switch to tacit collusion, or vice versa.

Further work in this area should be fruitful. One way of proceeding may be along the lines of Ormosi (2010a) recent working paper. The key idea of this

paper is to employ a multiple sampling method, similar to those used in ecology to estimate population parameters. This provides estimates of population characteristics (such as survival or detection rate) over time. The change in these time dependent estimates could be used, together with the change in the number of detected cases, to deduce how the deterrence rate changes over time, but this part of that paper is still work in progress.

5.2.4.2 Mergers

The deterrent effect of merger policies is slightly more complicated because it impacts not only on the number of foregone mergers, but also on the design of the mergers (i.e. had there been a less stringent policy in place, a different merger would have been proposed). Recognising this, Aaronson (1992) distinguishes frequency-based and composition-based deterrent effects. On the other hand, estimating the deterrent effect of merger policies should be relatively easier than for cartels because typically fewer mergers go undetected, and the observed change in the number of notified mergers should be informative about the deterrent effects of a given regulatory change or intervention.⁴⁶ This is confirmed by Barros et al. (2010), who find that the number of merger proposals (frequency-based deterrence) is a more appropriate indicator of the effects of underlying changes in merger policy than the relative anti-competitiveness of proposals (composition-based deterrence). Seldeslachts et al. (2009) use the frequency of mergers in their method for explaining merger waves, and then deterrence, as a departure from a wave. They use a longitudinal model where current mergers are explained in terms of various (lagged) CA actions and lagged merger notifications. Clougherty and Seldeslachts (2009) look at US merger deterrence using a similar method. They examine composition based and frequency based deterrence. Their model distinguishes between the start of investigations, the challenging of a merger, and the prohibition of mergers. They find that instigating an investigation and challenging a merger have significant deterrent effects, whereas prohibition does not.

⁴⁶ Except in regimes where there is no pre-merger notification such as in the UK.

Finally, it should be acknowledged that over-stringent policy can sometimes deter welfare *enhancing* practices or proposals. This would include mergers which might lead to strong efficiency effects; vertical practices (perhaps RPM) where the beneficial efficiency gains outweigh any anti-competitive foreclosure effect; or even cases where firms may refrain from fierce price competition because they are deterred by predatory pricing policy. This is rarely even discussed in CA evaluations, although admittedly quantification would be difficult. An interesting example is given by Eckbo (1992), who finds no evidence of Canadian mergers having anticompetitive effects. If so, it follows that any deterred merger must be either a welfare loss, or at least no gain.

6. Conclusions and future research priorities

This paper surveys the existing literature on the evaluation of competition policy and identifies some of the most pressing undeveloped areas and unanswered questions for priority in future research.

The first part of the paper focuses on the most popular quantitative evaluation methodologies, and draws five main conclusions:

1. Simulation, event studies and difference-in-differences are each useful weapons in the evaluation armoury, but they each seem less suited for certain types of markets and interventions than for others. It follows that any evaluation study based on a sample of cases, using only one methodology faces a likely problem of **selection bias** in the make-up of that sample.
2. There has been far less evaluation of **Article 102** cases than of merger and cartel cases. Possible explanations include (i) limitations of the existing methodologies, (ii) greater difficulty in detection (iii) greater deterrence of existing policy, (iv) greater caution by CAs in prosecuting Article 102 cases. Further research on these alternative explanations is a high priority.

3. Most evaluation studies have been applied over only **short-term time periods**. This is partly due to the nature of the methodologies themselves. In any event, it fails to recognise that (i) the effects of a particular policy intervention may not always be exhausted within one or two years, and (ii) intervention in one (e.g. merger) case may well affect the probabilities that subsequent events (mergers) will occur. More longer term studies would help fill an important gap in the literature.
4. Any policy evaluation necessarily requires an assumed **counterfactual**. Likewise, when choosing a particular evaluation methodology for a particular case, in effect, one may be choosing between different implicit counterfactuals. It follows that, where data limitations constrain which methodology is applied, or how it is applied, this will also constrain the choice of counterfactual. This deserves more recognition in the literature.
5. There are few comparative studies applying **alternative methodologies** to the same sample of cases, to identify whether there are systematic differences between methodologies in the quantification of impact. This paper explores one potential route for undertaking such a comparative analysis. It draws on an existing meta-sample of cartels, for which overcharge has been estimated using a variety of methodologies. Preliminary analysis suggests that there may be systematic differences which may be explicable in terms of systematic differences between the methodologies in their implicit counterfactuals. Work on this is ongoing, and the results may be of relevance, not only for evaluation *per se*, but also in understanding the typical time path of price before, during and after cartel episodes, and estimating damages.

The second part of the paper turns to arguably the most daunting problem faced by the evaluation project: how to extrapolate from the cases we know about to those we don't? This is crucial in order to draw any conclusions on the aggregate benefits of competition and competition policy. At this stage we draw four conclusions.

6. The set of investigated cases probably constitutes only a **small sample of the population** of all potential cases with competition implications. First, a

major impact of policy is the deterrence of anti-competitive practices/mergers which would have occurred absent competition policy. This is already widely recognised, but rarely quantified. On the other hand, what is sometimes overlooked is that CAs inevitably fail to detect some anti-competitive cases and wrongly fail to investigate some others of which they are aware. Where this occurs, there is a lost opportunity to secure welfare enhancement by the CA, and in that sense a 'cost'. Both success in deterrence and lost opportunities should be included in any aggregate assessment of competition policy.

7. Further research is clearly required if the extent of **deterrence and detection** are to be better understood and quantified. There are two dimensions to this: assessing (i) the frequency with which deterrence and detection occur and (ii) the expected magnitudes of harm entailed in deterred and undetected cases, relative to investigated cases.
8. Attempts to quantify both dimensions are still at a very preliminary stage, especially on the latter – precisely because of the problem of **selection bias**, which occurs whenever the characteristics of a selected sample (observed investigated cases) may differ systematically from cases that are excluded from the sample. For example, it is insufficient to know merely how many cartels go undetected, we also need to know whether they are systematically more or less harmful than those cartels which are detected and prosecuted.
9. There is an emerging academic literature which, in our opinion, currently offers the most promising prospects for improving an understanding of both detection and deterrence. Given further development, this should help improve how we tackle both problems of selection bias and how we attempt to quantify the full benefits of competition policy.

7. References

- AARONSON, R. (1992) Do companies take any notice of competition policy? *Consumer Policy Review*, 2, 140-145.
- ARISI, E. & ESPOSITO, F. M. (2007) L'attività di advocacy dell'Autorità Garante della Concorrenza e del Mercato: un bilancio dei primi 15 anni (Advocacy by the Italian Competition Authority: an overview over the first fifteen years). *unpublished paper referenced in Sabbatini (2008)*.
- ASHENFELTER, O. & HOSKEN, D. (2008) The effect of mergers on consumer prices: evidence from five selected case studies. *NBER Working Paper*, N° 13859.
- BARROS, P. P., CLOUGHERTY, J. A. & SELDESLACHTS, J. (2010) How to Measure the Deterrence Effects of Merger Policy: Frequency or Composition? *International Journal of the Economics of Business*, 17, 1-8.
- BECHER, D. A. (2009) Bidder returns and merger anticipation: Evidence from banking deregulation. *Journal of Corporate Finance*, 15, 85-98.
- BERGMAN, M. A. (2008) Quis Custodiet Ipsos Custodes? or Measuring and Evaluating the Effectiveness of Competition Enforcement. *De Economist*, 156, 387-409.
- BEVERLEY, L. (2008) Stock Market Event Studies and Competition Commission Inquiries. *CCP Working Paper No. 08-16*
- BIZJAK, J. M. & COLES, J. L. (1995) The Effect of Private Antitrust Litigation on the Stock-Market Valuation of the Firm. *The American Economic Review*, 85, 436-461.
- BOLOTOVA, Y. & CONNOR, J. (2005) Cartel Overcharges: Survey and Meta-Analysis. *Purdue University Paper*.
- BOSCH, J.-C. & ECKARD, E. W. J. (1991) The Profitability of Price Fixing: Evidence From Stock Market Reaction to Federal Indictments. *The Review of Economics and Statistics*, 73, 309-317.
- BRYANT, P. G. & ECKARD, E. W. (1991) Price Fixing: The Probability of Getting Caught. *The Review of Economics and Statistics*, 73, 531-536.
- BUCCIROSSI, P., CIARI, L., DUSO, T., FRIDOLFSSON, S.-O., SPAGNOLO, G. & VITALE, C. (2006) Ex-post review of merger control decisions. *European Commission*.
- CARLETTI, E., HARTMANN, P. & ONGENA, S. (2009) The Economic Impact of Merger Control Legislation. *TILEC Discussion Paper No. 2008-006, FDIC Center for Financial Research Working Paper No. 2008-12*
- CARLTON, D. (2009) Why we need to measure the effect of merger policy and how to do it. *Competition Policy International*, 5, 76-90.
- CHANG, M. H. & HARRINGTON, J. E. (2010) The Impact of a Corporate Leniency Program on Antitrust Enforcement and Cartelization. *mimeo.*, available at: <http://www.econ.jhu.edu/pdf/papers/WP548.pdf>.

- CLARKE, R., DAVIES, S. W. & DRIFFIELD, N. (1998) *Monopoly policy in the UK: Assessing the evidence*, Cheltenham, Edward Elgar.
- CLOUGHERTY, J. A. & SELDESLACHTS, J. (2009) The Deterrence Effects of U.S. Merger Policy Instruments. Available at SSRN: <http://ssrn.com/abstract=1437648>.
- COMBE, E., MONNIER, C. & LEGAL, R. (2008) Cartels: the Probability of Getting Caught in the European Union. *BEER paper* No.12.
- COOPER, Z., GIBBONS, S., JONES, S. & MCGUIRE, A. (2010) Does Hospital Competition Improve Efficiency? An Analysis of the Recent Market-Based Reforms to the English NHS. *CEP Discussion Paper* No 988.
- COX, A. J. & PORTES, J. (1998) Mergers in Regulated Industries: the Uses and Abuses of Event Studies. *Journal of Regulatory Economics*, 14, 281-304.
- DA GRACA, T. (2006) Simultaneous Mergers Probability Event Study
- DAVIES, S. W. (2010) A Review of OFT's Impact Estimation methods. *Office of Fair Trading*, OFT1164.
- DAVIES, S. W. & LYONS, B. (2007) *Mergers and Merger Remedies in the EU: Assessing the Consequences for Competition*, Cheltenham, Edward Elgar.
- DELOITTE (2007) The deterrent effect of competition enforcement by the OFT. *OFT962*.
- DELOITTE (2009) Review of merger decisions under the Enterprise Act 2002. *A review prepared for the Competition Commission, the Office of Fair Trading, and the Department for Business, Enterprise and Regulatory Reform*.
- DIEPOLD, B., FEINBERG, R. M., ROUND, D. K. & TUSTIN, J. (2006) Merger Impacts on Investor Expectations: An Event Study for Australia. *American University, Department of Economics Working Paper Series*, No. 2007-07.
- DOBSON, P. W. & PIGA, C. A. (2009) Mergers and Business Model Assimilation: Evidence from Low-Cost Airline. *Department of Economics, Loughborough University, Discussion Paper Series* No. 2009-2.
- DONOHUE, J. J. & WOLFERS, J. (2006) Uses and abuses of empirical evidence in the death penalty debate. *Stanford Law Review*, 58, 791.
- DUSO, T., GUGLER, K. & YURTOGLU, B. (2005) EU Merger Remedies: A Preliminary Empirical Assessment. *SP II 2005 – 16, Social Science Research Centre, Berlin*.
- DUSO, T., GUGLER, K. & YURTOGLU, B. (2006a) How Effective is European Merger Control? *WZB, Markets and Politics Working Paper* No. SP II 2006-12.

- DUSO, T., GUGLER, K. & YURTOGLU, B. (2006b) How Effective is European Merger Control? *WZB, Markets and Politics Working Paper*, No. SP II 2006-12.
- DUSO, T., GUGLER, K. & YURTOGLU, B. (2006c) Is the Event Study Methodology Useful for Merger Analysis? A Comparison of Stock Market and Accounting Data. *Governance and the Efficiency of Economic Systems*, Discussion Papers, No. 163.
- ECKBO, B. E. (1985) Mergers and the Market Concentration Doctrine: Evidence from the Capital Market. *The Journal of Business*, 58, 325-349.
- ECKBO, B. E. (1992) Mergers and the Value of Antitrust Deterrence. *Journal of Finance*, 47, 1005-1029.
- EHMER, C. & ROSATI, F. (2009) Science, Myth, and Fines: Do Cartels Typically Raise Prices by 25%? *Concurrences* No 4, 1-5.
- EVENETT, S. J. (2006) Competition Advocacy: Time for a rethink. Available at: www.alexandria.unisg.ch/EXPORT/DL/Simon_Evenett/28424.pdf.
- HARRINGTON, J. (2006) The Anatomy of a Cartel Price Path: Theory Meets Practice. IN KEYNOTE SPEECH, E. (Ed.
- HARRINGTON, J. E. (2004) Post-Cartel Pricing During Litigation. *Journal of Industrial Economics*, 52, 517–533.
- HARRINGTON, J. E. & CHANG, M. H. (2009) Modelling the Birth and Death of Cartels with an Application to Evaluating Antitrust Policy. *Journal of the European Economic Association*, 7, 1400-1435.
- HAUSMAN, J. & LEONARD, G. (1997) Economic Analysis of Differentiated Products Mergers using real world data. *George Mason Law Review*, 5, 321-46.
- HYYTINEN, A., STEEN, F. & TOIVANEN, O. (2010) Cartels Uncovered. *NHH Dept. of Economics Discussion Paper*, 10/2010.
- IVALDI, M. & VERBOVEN, F. (2005) Quantifying the Effects from Horizontal Mergers in European Competition Policy. *International Journal of Industrial Organization*, 669-691.
- KEMP, R. & SINDEREN, J. V. (2008) Measuring the economic effects of competition law enforcement. *De Economist*, 156, 365-385.
- KOVACIC, W., MARSHALL, R. C. & RAIFF, M. E. (2007) Lessons for Competition Policy from the Vitamins Cartel. IN GHOSAL, V. & STENNEK, J. (Eds.) *The Political Economy of Antitrust*,.
- KWOKA, J. E. & WHITE, L. (2004) *Antitrust Revolution: Economics, Competition, and Policy*, Oxford University Press.
- LANGUS, G. & MOTTA, M. (2007) The Effect of EU Antitrust Investigations and Fines on a Firm's Valuation. *CEPR Discussion Paper* No. DP6176.
- LEVENSTEIN, M. & SUSLOW, V. (2006) What Determines Cartel Success? *Journal of Economic Literature*, 44, 43–95.

- LÜBBERS, T. (2009) Is Cartelisation Profitable? A Case Study of the Rhenish Westphalian Coal Syndicate, 1893-1913. *MPI Collective Goods Preprint, No. 2009/9*.
- LYONS, B. (2009) *Cases in European competition policy : the economic analysis*, Cambridge, Cambridge University Press.
- MAJORAS, D. P. (2005) A Dose of our Own Medicine: Applying a Cost/Benefit Analysis to the FTC's Advocacy Program. Available at: www.ftc.gov/speeches/majoras/050208currebttopics.pdf.
- MALKIEL, B. G. (2003) *A random walk down Wall Street : the time-tested strategy for successful investing*, New York, London, W. W. Norton.
- MEYER, B. D. (1994) Natural and Quasi- Experiments in Economics *NBER Working Paper Series*, Vol. t0170.
- MILLER, N. H. (2009) Strategic Leniency and Cartel Enforcement. *American Economic Review*, 99, 750-768.
- NEVEN, D. & ZENGLER, H. (2008) Ex-post evaluation of enforcement: a Principal-Agent Perspective. *De Economist*, 156, 477-490.
- NEVO, A. (2000) Mergers with Differentiated Products: the Case of the Ready-to-eat Cereal Industry. *RAND Journal of Economics*, 31, 395-421.
- OFT (2010a) Drivers of Compliance and Non Compliance with Competition Law. *Office of Fair Trading*, OFT1227.
- OFT (2010b) Evaluation of OFT Competition Advocacy. *Office of Fair Trading*, OFT866.
- OFT (2010c) Positive Impact 09/10 Consumer benefits from the OFT's work. *Office of Fair Trading*, OFT1251.
- ORMOSI, P. (2010a) Is it really a tip of the iceberg? A parsimonious way to measure detection rate in cartel enforcement. *mimeo*.
- ORMOSI, P. L. (2010b) Merger Remedies versus Efficiency Defence: An analysis of merging parties' litigation strategy in EC merger cases. *CCP Working Paper Series*, 10-1.
- OXERA (2009) Quantifying antitrust damages Towards non-binding guidance for courts. *Study prepared for the European Commission*, <http://www.oxera.com/cmsDocuments/Quantifying%20antitrust%20damages.pdf>.
- PETERS, C. (2006) Evaluating the Performance of Merger Simulation: Evidence From the U.S. Airline Industry. *Journal of Law & Economics*, 49, 627-649.
- PINSKE, J. & SLADE, M. E. (2004) Mergers, Brand Competition, and the Price of a Pint. *European Economic Review*, 48, 617-643.
- PRAGER, R. A. (1989) Using stock price data to measure the effects of regulation: the Interstate Commerce Act and the railroad industry. *RAND Journal of Economics*, 20, 280-290.

- R.A. CONNOR, FELDMAN, R. D. & DOWD, B. E. (1998) The effects of market concentration and horizontal mergers on hospital costs and prices. *International Journal of Economics and Business*, 5, 159-180.
- SABBATINI, P. (2008) Assessing the impact of antitrust intervention by the Italian Competition Authority. *De Economist*, 156, 491-505.
- SELDESLACHTS, J., CLOUGHERTY, J. A. & BARROS, P. P. (2009) Settle for Now but Block for Tomorrow: The Deterrence Effects of Merger Policy Tools. *The Journal of Law and Economics*, 52, 607-634.
- SHAW, R. W. & SIMPSON, P. (1986) The Persistence of Monopoly: An Investigation of the Effectiveness of the United Kingdom Monopolies Commission. *The Journal of Industrial Economics*, 34, 355-372.
- SIMPSON, J. D. & HOSKEN, D. (1998) Are Retailing Mergers Anticompetitive? An Event Study Analysis. *Federal Trade Commission*, WP-216.
- SIMPSON, J. S., DAVID (2008) Difference-in-Differences Analysis in Antitrust: A Cautionary Note. *Antitrust Law Journal*, 75, 623.
- SLADE, M. (2008) The effects of vertical restraints: an evidence based approach. *The pros and cons of vertical restraints*. Stockholm, Swedish Competition Authority.
- SYMEONIDIS, G. (2002) *The effects of competition : cartel policy and the evolution of strategy and structure in British industry*, Cambridge, Mass., MIT Press.
- TENN, S. & YUN, J. M. (2010) The success of divestitures in merger enforcement: Evidence from the J&J–Pfizer transaction. *International Journal of Industrial Organization*, Forthcoming.
- VERBOVEN, F. & VAN DIJK, T. (2009) Cartel damages claims and the passing-on defence. *Journal of Industrial Economics*, 57, 457-491(35).
- WARREN-BOULTON, F. R. & DALKIR, S. (2001) Staples and Office Depot: An Event-Probability Case Study. *Review of Industrial Organization*, 19, 469-481.
- WERDEN, G. J. (2000) Expert Report in United States v. Interstate Bakeries Corp. and Continental Baking Co. *International Journal of the Economics of Business*, 7, 139-148.
- WERDEN, G. J. (2008) Assessing the Effects of Antitrust Enforcement in the United States. *De Economist*, 156, 433-451.
- WHINSTON, M. D. & COLLINS, S. C. (1992) Entry and Competitive Structure in Deregulated Airline Markets: An Event Study Analysis of People Express. *The RAND Journal of Economics*, 23, 445-462.