



# The Impact of Competition Policy: What are the Known Unknowns?

**Stephen Davies**

School of Economics, University of East Anglia

**Peter Ormosi**

Norwich Business School, University of East Anglia

**CCP Working Paper 13-7**

## **Abstract**

Evaluations of competition policy are increasingly common and typically establish that consumer benefits from detected cases easily outweigh the costs of competition authorities (CA). However, such assessments are often driven by data availability and only capture a small part of the total impact because they sidestep the difficult issue of how to evaluate deterrence. Similarly, they ignore the fact that policy does not root out all anti-competitive cases. This paper suggests a broader framework for evaluation which encompasses these unobserved impacts. Calibration is difficult precisely because we cannot rely on empirical observations on cases which have been observed to make deductions about cases which have not (because they are deterred or undetected). It thereby confronts the classic sample selection problem which is endemic in all studies based on data from CA decisions. Drawing on insights from economic theory, it argues that selection bias is likely to be substantial because the unobserved cases could well be those which are most harmful. If so, the deterrence of anti-competitive mergers may have a much greater positive impact, but the effects of non-detected cartels may be more serious than is usually supposed.

# The Impact of Competition Policy: What are the Known Unknowns?\*

Stephen Davies<sup>†</sup>

School of Economics, University of East Anglia, UK

Peter L. Ormosi<sup>‡</sup>

Norwich Business School, University of East Anglia, UK

June 13, 2013

---

\*The support of the Economic and Social Research Council and the ESRC Centre for Competition Policy is gratefully acknowledged.

<sup>†</sup>School of Economics and ESRC Centre for Competition Policy, University of East Anglia, Norwich Research Park, NR4 7TJ, Norwich, United Kingdom, email: s.w.davies@uea.ac.uk

<sup>‡</sup>Norwich Business School and ESRC Centre for Competition Policy, University of East Anglia, Norwich Research Park, NR4 7TJ, Norwich, United Kingdom, email: p.ormosi@uea.ac.uk

## Abstract

Evaluations of competition policy are increasingly common and typically establish that consumer benefits from detected cases easily outweigh the costs of competition authorities (CA). However, such assessments are often driven by data availability and only capture a small part of the total impact because they sidestep the difficult issue of how to evaluate deterrence. Similarly, they ignore the fact that policy does not root out all anti-competitive cases. This paper suggests a broader framework for evaluation which encompasses these unobserved impacts. Calibration is difficult precisely because we cannot rely on empirical observations on cases which have been observed to make deductions about cases which have not (because they are deterred or undetected). It thereby confronts the classic sample selection problem which is endemic in all studies based on data from CA decisions. Drawing on insights from economic theory, it argues that selection bias is likely to be substantial because the unobserved cases could well be those which are most harmful. If so, the deterrence of anti-competitive mergers may have a much greater positive impact, but the effects of non-detected cartels may be more serious than is usually supposed.

**Keywords:** competition enforcement, impact assessment, selection bias, cartels, mergers

**JEL Classification codes:** K21, L44, H11

# 1 Introduction

*A policeman happens upon a drunk crawling on his hands and knees under a street-lamp on a darkened stretch of sidewalk. When he inquires, the drunk explains that he's searching for his lost keys. The policeman promptly kneels down and begins searching as well. After they've scoured the nearby pavement and gutter several times over the policeman thinks to ask, "Are you sure this is where you lost your keys?" The drunk replies, "Oh no, I lost my keys about fifty yards back." He gestures at the darkened street behind him. "Well, then, why are you searching here?" asks the policeman. The man replies, "The light here is so much better."*

With the increasing prominence of competition enforcement throughout the world, competition authorities (CA) and academics have begun to devoting increasing effort to estimating its impact on economic welfare. Although there are some dissonant voices,<sup>1</sup> most evaluations are very positive, suggesting that the costs of competition policy are more than balanced by the benefits it brings.<sup>2</sup>

Very often, empirical assessments take the form of ex-ante impact evaluation in which the CA estimates, using conservative assumptions, the consumer benefits of the cases it has intervened during the year. Although fairly rough and ready, these evaluations suggest that the benefits to consumers from the activities of the CA more than outweigh their cost, usually by an order of magnitude. For example, impact evaluation for the EU in 2010 suggests that consumers benefited from the removal of overcharges by cartels that were detected and prohibited by the European Commission (EC) to the tune of €7.2 billion; and the benefits of remedying/blocking anti-competitive mergers were €4.2-6.3 billion. The Commission's costs (DGCOMP) were around €100 million in that year.

This paper starts from the premise that competition policy probably does provide good value for money and that impact evaluation of the sort just described are

---

<sup>1</sup>For example, Crandall and Winston (2003) stridently reject any positive impact of competition policy. The balance is redressed by Baker (2003) in the same journal issue, who paints a far more positive picture.

<sup>2</sup>See for example the EC's 2007-2010 Annual Activity Reports, the USDOJ Antitrust Division's Performance Budget Congressional Submissions for 2008-11, or the OFT's Positive Impact for 2012.

valuable, not only for advocacy purposes but also because it encourages the monitoring of activity and prioritisation within the authorities. However such studies are based only on the cases that the CA prosecutes/intervenes: cartels which it busts, anti-competitive mergers which it blocks or remedies, and abuses of dominance it removes - we call these the observed cases. Inevitably however this raises the obvious question: what about the cases that go unobserved, because they are deterred? After all a major function of any law is to deter antisocial behaviour, and this is hopefully the most important function of competition policy. It is often suggested that the magnitude of deterred harm probably far exceeds that from harm removed by direct intervention,<sup>3</sup> but this is rarely if ever quantitatively factored into policy evaluations. But it also raises a second question: what about a second type of unobserved case - those which the CA fails to detect, even although they involve anti-competitive harm? In effect, these represent a foregone opportunity, or even failure of policy; again, this is never quantified, although the magnitudes involved also could be substantial.

The starting motivation for the current paper is that ideally policy should be evaluated not only on the basis of the 'observed' (the detected and intervened), but also taking account of the 'unobserved' (the deterred and the undetected). This leads to the thought that a more encompassing approach to policy evaluation would be to ask first "how much potential for anti-competitive harm is there out there in an economy?" and then assess the success of policy by asking "how much of that potential harm is avoided and/or removed by the presence of competition law and the activities of competition authorities?" The first question is reminiscent of an old literature, provoked by Harberger (1954) which attempted to quantify the "social costs of monopoly", although our approach is very different from that literature. Undoubtedly such an approach is ambitious and speculative - ultimately, we can never know how many cases remain unobserved because they are deterred or not detected. Moreover, even if we know how many cases are unobserved, there is a potential sample selection issue - the characteristics of the unobserved may differ significantly from the observed. Equally however the potential insights on how policy

---

<sup>3</sup>See for example Werden (2008), Geroski (2004, p.3) and Baker (2003, p.40).

impacts are far wider reaching than from merely asking, as in a conventional impact assessment, whether policy generates benefits which exceed its costs.

The paper presents a conceptual framework for assessing the full impact of competition policy in this way, by taking into account the deterred and undetected as well as the detected. It marshals what we do ‘know’ - stylised facts from existing academic and policy literatures, alongside what it is reasonable to infer from theory about what we do not know - in particular, the nature of the underlying population of potential cases, and how samples are drawn from it, in the form of cases detected by the CA. Simple numerical simulations are conducted within this framework and these suggest that, while the benefits of deterrence, especially for merger enforcement, are considerable. It is also likely that considerable harm may remain unremedied due to the non-detection of cartels. Moreover, in both respects simple estimates of deterrence and detection rates, based on estimating or guessing ‘how many’ cases are involved may seriously under-estimate the ‘magnitudes of harm’ because many of the cases we do not observe will be far more harmful than those we observe. We emphasise that these simulations are indeed very speculative and just a first step. Nevertheless, they help focus on what should be the priorities in future research.

## 1.1 Previous literature

The policy evaluations mentioned above have taken a variety of forms. Some by the agencies themselves, have provided estimates of the total impact of all cases they have enforced, others focus more specifically on types of policy, e.g.. just mergers or cartels, and others estimate the impact of individual interventions (individual mergers or cartels). Quantitative methods vary widely, including difference-in-differences, event studies, simulations and surveys. See Ormosi (2012) and Davies and Ormosi (2012) for summaries of these various literatures. However, all these works are constrained by the same limitation, driven mostly by data availability inferences are made on the basis only of the sample of cases in which the CA has intervened, and fail to take account of unobserved cases.

There are also valuable empirical contributions in the wider literatures on both

deterrence and detection. For mergers, one strand in the academic literature makes inferences on the rate of deterrence by examining the serial correlation in the number of intervened cases over time. In general, this approach is problematic because the number of detected cases is an ambiguous indicator of deterrence - a reduced number could imply either increased deterrence or a reduction in the detection rate. This is less of a problem for mergers if the rate of detection can be assumed constant, which is likely for regimes with compulsory pre-merger notification rules. Clougherty and Seldeslachts (2013), Barros et al. (2010), and Seldeslachts et al. (2009) exploit this feature to identify the deterrent effect of merger policy in the US and EU, and confirm that lagged detection does impact on current deterrence, although not necessarily in the composition (degree of harm) in those cases.

In the cartel literature, a commonly cited statistic, originating from the early work of Bryant and Eckard (1991) is that, in a given year, only 13-17% of cartels are detected. More recently for Europe, Combe et al. (2008) confirm a similar magnitude using the same method.<sup>4</sup> However, these estimates are derived from information on cartel duration – if we know that cartels last on average for 7 years, it follows that in any one year only one seventh of cartels will be detected. This statistic therefore only estimates the conditional probability – conditional on the cartel eventually being detected. Ormosi (2013) proposes an alternative method which draws an analogy with capture-recapture analysis in ecological science. His work provides estimates of how detection rate changes over time, which - used together with the change in the number of detected cases - may tell us something about the rate of deterrence. However, this approach also relies on cartels which are eventually caught, and still leaves open the question of how many and what types of cartels are never seen because there are deterred or never detected.

Most immediately relevant to the aims of the current paper are some survey studies by the CAs themselves which are designed to estimate deterrence and sometimes detection rates. Deloitte (2007), commissioned by the OFT in the UK, involved interviews/telephone surveys of competition practitioners (lawyers and economists)

---

<sup>4</sup>Lande and Connor (2012) provides an exhaustive survey of all estimates including those derived from surveys.

and of companies. From the survey of legal advisers, it reports 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. On detection, the survey suggests that the number of 'under the radar' mergers (i.e. undetected by the OFT) was at least as high as the number which are blocked or modified following intervention.<sup>5</sup> On the other hand a study by the Dutch CA (NMa), der Noll et al. (2011), based on surveys of companies, reports that only 5 per cent of all notified mergers had been modified prior to notification, and that 13 per cent were deterred completely in anticipation of the NMa's intervention. For cartels the Deloitte (2007) study reports an analogous "multiplier" of 5:1. In a follow-up larger survey along similar lines, also commissioned by OFT, London Economics (2011) reports much higher deterrence multipliers, 28 deterred cases for every detected case.<sup>6</sup> Similar surveys conducted by der Noll et al. (2011) from the NMa found that 60% of cartels were deterred and about a third of the undeterred cases was detected.

At first sight, multipliers derived from such surveys provide a straightforward route for converting the estimated magnitudes of impact from observed CA interventions into estimates of both the magnitudes of deterred and undetected cases - although survey results based on subjective assessments are inevitably open to doubt, especially when, as here, estimates differ so noticeably between the UK and the Netherlands.<sup>7</sup> But even accepting such multipliers at face value, there remains the potentially serious possibility of sample selection bias. Even if we know that for every cartel detected there are another 5 which are deterred a 5 which go undetected, we can not deduce that the magnitudes of undetected and deterred cartel harms are each 5 times greater than the CA's estimates of the savings it achieves from the

---

<sup>5</sup>Pre-merger notification is not compulsory in the UK. Below, we generally assume that the CA detects all anti-competitive mergers.

<sup>6</sup>However, this figure does not account for non-detection. For example if only, say, 1 in 5 cartels are detected, then for every cartel formed there are 'only' 5.6 deterred.

<sup>7</sup>Some of the difference may be due to a difference in the interpretations of 'deterrence'. For example the Dutch survey presented participants with a specific scenario, assuming among other things a merger which would increase price by 10%, and respondents had to decide whether they would be deterred. The UK survey was not so specific, and thus deterrence might be interpreted as including more extreme possibilities, such as a 2-to-1 merger proposal.



cartels it detects, because the observed sample may not be representative of the full population.

This paper confronts this sample selection problem, and, in that sense, is of wider relevance than just the narrowly defined policy evaluation literature. There is a significant body of empirical research in Industrial Organisation which employs datasets of competition cases investigated and intervened by CAs, and such data have provided a rich source of information on cartels (typical overcharge, duration, structural characteristics etc.), mergers (price effects, efficiency, likelihood of unilateral and coordinated effects etc.), and other areas of policy. While many studies in that literature acknowledge the possibility of sample selection bias, one can only speculate on just how important it might be for many of the conventional wisdoms that are derived from that literature, for example, on cartel duration and overcharge or the price raising effects of anti-competitive mergers.

## 1.2 Structure of the paper

Section 2 defines terms and introduces a simple framework which relates the number of detected cases to an underlying population of all potential anti-competitive cases that would occur in the absence of competition policy. Using simple multipliers of the type discussed above, this provides rough estimates of the size of the observed sample as a proportion of the population, when measured by the number of cases. But these may be seriously misleading as estimates of measures of the magnitudes of total population harm, unless the sample is representative of the population. The remainder of the paper argues that this is unlikely. Section 3 draws on previous literature to propose that the population of potential anti-competitive harm will be positively skewed; if so, non-random selection could lead to quantitatively very serious bias. Section 4 surveys existing theoretical literature to propose some likely properties of the samples of mergers and cartels which are observed in CA interventions. Section 5 translates into a simple sampling problem in which there is differential sampling, first non-parametrically and then in the special case where the population is lognormal. As an alternative, an Appendix posits an alternative skewed distribution, the

Pareto curve. Section 6 calibrates using, as an illustration, data on merger and cartel enforcement by the European Commission. Section 7 explores the likely impact of allowing for Type 1 errors in merger control. Section 8 concludes and points to directions of our further work.

## 2 Framework

Figure 1 describes a stylised classification of all potential, privately profitable, cases in a given economy for a given point in time (say year). These represent all cases in which firm behaviour would cause consumer harm, and which would actually occur in the absence of competition enforcement.<sup>8</sup> These include cartels, anti-competitive mergers and abuses of dominance, although for simplicity in this paper we consider only mergers and cartels, leaving abuse as the subject of ongoing work. There is a total population of  $N$  such cases.

Figure 1: The population of cases

Deterred $\omega$	Undeterred $(1 - \omega)$	
	Undetected $(1 - \sigma)$	Detected $\sigma$

Although all these cases would be privately profitable in the absence of competition policy, some do not occur because they are deterred by the existence of competition law and enforcement. The probability that a case is deterred, and therefore does not occur, is given by  $\omega$ . For simplicity at this stage, we abstract from ‘deterrence of composition’, i.e. where the parties choose to modify their behaviour in order to avoid detection (for example, a cartel choosing to limit the scale of its overcharge, or merger parties choosing to exclude some markets in which the merger may otherwise

---

<sup>8</sup>This counterfactual is over-simplification to the extent that private enforcement might anyway deter some such cases. It also abstracts from the possibility that other policy, such as import liberalisation or deregulation, might also deter behaviour.

have an anti-competitive impact). We return to the composition effect at the start of section 4.

Within the undeterred subset, there are some cases that are discovered by the CA, but others go undetected; the conditional probability that an undeterred case is detected is  $\sigma$ . Since we have defined the population to include only anti-competitive cases, any investigated, but cleared, case is excluded from the population. For the moment, Type 1 and Type 2 errors are excluded from this framework, but these are discussed below in section 7.

In general, both probabilities and  $N$  may be time variant and interdependent; for example, higher detection probabilities may lead, *ceteris paribus*, to greater deterrence in the future. However, this is irrelevant for the purposes of the current paper.

Since the probability that a potential case is not deterred and detected is  $(1-\omega)\sigma$ , the expected number of detected cases is:<sup>9</sup>

$$C = (1 - \omega)\sigma N \tag{1}$$

This leads to the straightforward observation that, at any given time, the number of detected and the number of deterred cases are inversely related, although it should be stressed that this observation merely refers to the inverse static tautological relation between  $C$  and  $\omega N$  in Eqn.(1).<sup>10</sup> It ignores the possibility of a positive dynamic causal relation running from  $\sigma$  to  $\omega$ : success by the CA in detecting might deter firms from attempting to contravene the law in the future. Nevertheless, it highlights an important qualification which should always be remembered in evaluation – a low number of detected cases does not necessarily signal a lazy or inept CA, it might also be consistent with an Authority which is particularly effective in deterring.

**Proposition 1** *Detected cases are likely to be only a small proportion of the population*

---

<sup>9</sup>Throughout the paper, to avoid superfluous notation, any expectation operators are omitted.

<sup>10</sup>This follows obviously because as deterrence increases, this leaves fewer cases as detectable.

To substantiate this and provide a rough illustration, we return to the surveys briefly described in Section 1: suppose for every merger intervened, there are 5 others in the population deterred, and for every detected cartel, there are 5 others deterred, but 5 others undetected. We assume that all anticompetitive mergers above the compulsory notification requirement (which applies in most countries) are detected, and that anticompetitive mergers below have trivial effect. Table 1 backs out estimates of the two parameters and  $C$  from these figures. This suggests that ‘small’ in Proposition 1 is 9% and 17% for cartels and mergers respectively.

Table 1: A back of the envelope calibration

	Cartels	Mergers
Deterrence multiplier	5	5
Non-detection multiplier	5	0
$\omega$	0.45	0.83
$\sigma$	0.17	1
$C/N$	0.09	0.17

These figures show that previous works - based on detected cases only - focus on a small segment of anticompetitive harm, the total population harm is likely to be much larger. If detected cases are taken randomly from the population then this can be derived by a simple multiplication.<sup>11</sup> For example suppose the CA estimates the aggregate consumer harm it has avoided by its interventions to be £100, then if the sample is random, the magnitudes of these might be estimated by applying these multipliers respectively to £100, i.e. £500 for both deterrence and non-detection for cartels, and £500 for deterrence of anticompetitive mergers.<sup>12</sup>

However, the assumption of randomness is questionable, and underlying the rest of this paper is the possibility that it is in fact implausible. Especially if some

---

<sup>11</sup>In our notation, the harm by deterred cases could be derived by using the multiplier  $\omega/[(1 - \omega)\sigma]$ , and the harm from non-detected cases by multiplying observed harm by  $[(1 - \sigma)(1 - \omega)]/[(1 - \omega)\sigma] = (1 - \sigma)/\sigma$ .

<sup>12</sup>The OFT discusses the potential use of the deterrence multiplier along these lines in its current Positive Impact evaluation (2012).

unobserved cases involve substantially higher than average harm, the magnitudes of aggregate harms involved may be substantially higher than implied by the above calculations.

Viewed in this way, the paper addresses a classic potential problem of selection bias. While the context here is the evaluation of the impact of CAs, this question has much wider implications for any empirical research in IO where conclusions are drawn from databases based on those mergers and cartels intervened by CAs. For example Connor and Bolotova (2006) found the average cartel overcharge to be between 31 and 49 per cent, and Boyer and Kotochoni (2011) later claim to derive a ‘less biased’ estimate of around 13-17 per cent. But neither of these works acknowledge the possibility that these detected cartels may be systematically different from undetected cases. If the undetected subset is more likely to include higher overcharge cases, then these estimates will undershoot the true average overcharge for the population of all cartels. For mergers, Crandall and Winston (2003) estimate the impact of public merger interventions on price cost margins between 1984 and 1996 for 20 manufacturing industries, and argue that a merger challenge has a very small and statistically non-significant negative effect on the price-cost margin. They conclude that this is because the authorities are not successful in distinguishing bad from good mergers. But, even if this is the case, this type of analysis ignores the fact that in addition to the intervened mergers, there will be other anticompetitive mergers that have been successfully deterred. An outright anticompetitive merger is unlikely to be proposed in the first place, therefore the true beneficial impact of competition policy may potentially far exceed the impact estimated from the selected set of intervened mergers.

To investigate the likely nature of such selection bias, the next two sections draw on previous literatures to suggest working assumptions about the nature of (i) the population distribution of harm, and (ii) samples drawn from it, i.e. cases detected and intervened by CAs.

### 3 The population distribution

Denote the net harmful (price increasing or quality reducing) effect of any anti-competitive merger or cartel by the random variable  $h$ , with a density function  $f(h)$  and finite mean:  $E(h) = \int_0^\infty hf(h)dh$ .

**Proposition 2**  *$h$  has a population size distribution which is positively skewed.*

In justification, note first that harm is the product of  $\Delta p$ , the price change (or quality degradation) caused by the anticompetitive behaviour  $Q$ , the size of the affected market, and  $t$  the avoided duration of the anticompetitive effect. In general,  $Q$ , like most non-negative size variates will have some form of positively skewed distribution; this seems intuitively likely given that CA impact evaluations are often dominated quantitatively by a few high-profile very large markets (e.g. grocery supermarkets, petrol, banking). The duration of the conduct  $t$  is also likely to behave like any other duration variable, with a long right-skew. Numerous studies on the duration of cartels support this assumption.<sup>13</sup>  $\Delta p$  merits further discussion for mergers and cartels separately.

**Mergers** In a market of  $N$  firms, the possible permutations of bilateral horizontal mergers is:  $\frac{N(N-1)}{2}$ . On a priori grounds, it is likely that the mergers involving the largest firms are more likely to be harmful. This assumption is still embedded in most authorities' merger guidelines, and a simple formalisation can be easily derived for the Cournot model, albeit for a special case. Given linear demand, Davies and Lyons (2007, p.127) show that the price rise following merger will be:  $\Delta P^M = \frac{P_1 - P_0}{P_0} = \frac{1}{N} \frac{s}{\varepsilon}$ , where  $P_1$  and  $P_0$  are the post and pre-merger prices respectively,  $s$  is the pre-merger market share of the acquired firm,  $\varepsilon$  is the market demand elasticity and  $N$  is firm numbers. Since the distribution of market share is typically positively skewed, so too will be the potential price rise from merger. With differentiated products this effect will be less pronounced to the extent that diversion ratios need not necessarily map directly on to market shares.

---

<sup>13</sup>See for example Levenstein and Suslow (2011).

**Cartels** In principle, the upper bound on the cartel mark-up (harm) is the full monopoly price - the reciprocal of the demand elasticity. We have little hard evidence on how market elasticities are distributed across industries, but assuming, as a default, a uniform distribution, then its reciprocal would take some sort of positively skewed distribution. While the incentive compatible cartel price will rarely reach this upper bound, and the counterfactual price (what would happen absent the cartel) is rarely marginal cost pricing, unless the difference between upper bound and actual price is systematically and strongly negatively related to the elasticity, the actual overcharge will also follow a positively skewed distribution.<sup>14</sup>

## 4 Sample selection

This section examines what can be inferred from previous literatures about how the probability of selection (non-deterrence and detection) varies with case-harm, i.e. how  $(1 - \omega)$  and  $\sigma$ , vary with  $h$ . It surveys some models in the existing theoretical literatures on cartels and mergers to draw out their implications for how deterrence and detection may vary according to how harmful a case is. Key inferences from these literatures are presented and discussed.

To avoid any illusion of spurious precision, we choose not to impose too much structure on the inferences we make. Therefore we merely adopt a stylised trichotomy, in which the population is broken down into three segments, low, medium and high harm cases ( $h_L$ ,  $h_M$ , and  $h_H$ ) - leaving open more precise delineation of these categories for later research. Thus the purpose is to draw inferences about how detection and deterrence rates,  $\sigma_L^{M,C}$ ,  $\sigma_M^{M,C}$ ,  $\sigma_H^{M,C}$  and  $\omega_L^{M,C}$ ,  $\omega_M^{M,C}$ ,  $\omega_H^{M,C}$ <sup>15</sup> might differ between the three segments, for cartels and mergers respectively. Although crude, this trichotomy allows for the possibility of non-monotonicity and is also carried forward into the calibration section below.

---

<sup>14</sup>It might be noted that most meta-analyses of cartel overcharge (e.g Connor and Bolotova, 2006) report positively skewed distributions, but this is inconclusive for our purposes since such studies are heavily biased towards cartels detected under prohibition regimes.

<sup>15</sup>To avoid notation overload, the M (mergers) and C (cartels) superscripts are omitted hereafter.

As a preliminary, we first return to the issue of composition deterrence mentioned earlier. We now allow that, for any given merger or cartel, there is a potentially large number of forms and therefore harms ( $h$ ), it might take. For example two firms may merge their activities in different ways, depending on how much overlap they want to create; and cartels may decide between different collusive prices. For each case there is then a continuous distribution of potential harm, and the choice that firms make depends on their assessment of the probability of detection. When the prospect of enforcement deters a cartel altogether, then a larger proportion of  $h$  is deterred than when the cartel merely decides to coordinate on a lower price to lessen the chance of detection.<sup>16</sup> Similarly, some mergers may be deterred in any form, while others are still consummated, albeit with divestments of overlapping assets offered. In this way both frequency and composition deterrence are consistent with our analysis.

## 4.1 Cartels

Anti-cartel enforcement has two arms: ex officio detection and leniency applications, and previous literature has implications for how each impacts on deterrence and detection.

### 4.1.1 Detection

**Inference 1a (leniency):**      *The probability of detection declines with case harm.*

This draws from both the theoretical and experimental literatures, such as Chang and Harrington (2012) and Bigoni et al. (2012), which suggest that less stable cartels are more likely to apply for leniency. Leniency applications are therefore strongly linked with the incentive compatibility of the cartel, and often only made once the cartel breaks down. As our population is defined to include only privately profitable cartels (incentive compatible if there were no enforcement) this implies that more profitable cartels are less likely to break down when reduction of fines is offered under

---

<sup>16</sup>See for example Block et al. (1981), where the threat of detection does not necessarily dissuade the cartel from forming, but may persuade it to reduce the markup.



leniency. This is formally proven in the Appendix (A.1.1) applying the framework used by Motta and Polo (2003).

**Inference 1b (ex-officio detection):**      *There are no strong reasons for expecting ex-officio detection to vary with case harm.*

Our reading of the literature suggests that there are no convincing arguments for rejecting the hypothesis that the probability of ex-officio detection is invariant with respect to the magnitude of case harm. This is contrary to Block et al. (1981), who assume that the detection probability increases with the size of markup because higher prices are more likely to be spotted by customers or the CA. However, neither the empirical and experimental evidence offers any support for this assumption.<sup>17</sup> Rather, there is a growing understanding that it is high price changes, rather than levels, that create suspicions in the minds of customers and the CA.<sup>18</sup>

**Inference 1c:**      *The probability of detection by leniency is greater than the probability of ex-officio detection.*

There is widespread evidence that in countries with leniency programmes the proportion of leniency cases far exceeds the proportion of ex officio cases; in the EU the ratio is approximately 2.

If these inferences are correct, then it follows that the aggregate probability of detection, the weighted average of the leniency and ex-officio detections, must decline with case harm:

**Inference 1**       $\sigma_L > \sigma_M > \sigma_H$

Note that this would be robust to a weak tendency for ex-officio detection to increase with harm.

---

<sup>17</sup>Bigoni et al. (2012).

<sup>18</sup>Harrington (2005) and Harrington (2004).

### 4.1.2 Deterrence

Conventionally the literature assumes that rational profit-maximising firms will only form a cartel if the expected payoff exceeds the payoffs from competing/deviating after allowing for the probability of detection and regulatory sanctions. It follows that, if the expected fine is invariant to the amount of harm, or has a significant fixed component, cartels with sufficiently small price raising effects are more likely to be deterred. For instance, Chang and Harrington (2009) defines a marginal industry, in which firms are indifferent between collusion and competition (given a competition authority), and a toughening of policy raises the critical level defining the marginal industry, and the probability of deterrence increases. This is confirmed empirically by the earlier mentioned NMa survey, which reports that increasing expected fines would result in a larger proportion of deterred cases.

More formally, there are many theoretical models that analyse how the incentive compatibility of cartels changes given CA enforcement. Some of these, Motta and Polo (2003), Chang and Harrington (2009), Harrington (2004 and 2005) are summarised in the Appendix. They lead us to:

**Inference 2**      $\omega_L > \omega_M > \omega_H$

Note that in models that do not allow leniency and assume that the probability of detection increases with harm, the results would be the contrary, see for example Block et al. (1981) or Houba et al. (2012) in which cartels with larger harm are less IC. However, given the prominence of leniency in anti-cartel enforcement and the absence of evidence for the assumption that higher harm cartels are more likely to be detected, we prefer not to follow this route.

## 4.2 Mergers

### 4.2.1 Detection

**Inference 3**     *All anticompetitive mergers are detected ( $\sigma = 1$ )*

In most jurisdictions, notification of all proposed mergers above some threshold is compulsory, in which case, literal non-detection is confined to the lower tail of the  $h$  distribution. An exception is the UK, but even there it is unlikely that any large merger would go unnoticed by the OFT.

#### 4.2.2 Deterrence

Assuming that the harm caused by the merger is directly proportional to its profitability, Barros et al. (2010) argue that a merger will be deterred if (a) it is too restrictive (i.e. mergers in the upper tail), or (b) the expected payoff does not exceed the cost of the merger approval (mergers in the lower tail). Regarding (a) assume that in the absence of merger control all privately profitable mergers take place. In this case one would observe that markets would merge to monopoly. However, this is not the case, and this is presumably because firms rarely propose mergers which are clearly anti-competitive because they anticipate that they are likely to be prohibited by the CA. Many theoretical papers follow this intuition. For example Barros et al. (2010), Selde-slachts et al. (2009) and Garrod and Lyons (2011) all assume that a CA prohibits mergers with harmful effects above a CA's reservation point,  $h > \bar{h}$  (where  $\bar{h}$  depends on the rigour of merger control). Merger guidelines send out clear messages to firms on the types of cases where a merger is likely to be found anti-competitive (e.g. in the EC mergers with  $HHI > 2000$ ). Therefore highly anti-competitive mergers are less likely to be proposed as they carry a very high chance of prohibition.

The NMa (2011) survey described above investigated firms' reactions to both (a) and (b) in The Netherlands, and found that an increase in merger administration fees increased deterrence. Also, while the firms themselves were less inclined to abort plans for mergers with higher joint market shares, their legal advisors advised to the contrary – higher market shares would be interpreted by the CA as more likely to be considered anti-competitive and less likely to be cleared.

This leads to Inference 4. Formal proof is also provided in A.2.2 of Appendix A.

**Inference 4**       $\omega_L > \omega_M < \omega_H$

Table 2 summarises the inferences of this section. It shows that selection (non-deterrence or detection) is unlikely to be random over the harm caused by a cartel or anticompetitive merger. Also there is a strong chance that by looking at detected cases we may ignore those cases where potentially most anticompetitive harm lies.

Table 2: Deterrence and detection probabilities

	Cartel	Merger
Deterrence	$\omega_L > \omega_M > \omega_H$	$\omega_L > \omega_M < \omega_H$
Detection	$\sigma_L > \sigma_M > \sigma_H$	$\sigma_L = \sigma_M = \sigma_H = 1$

## 5 Selection bias with differential sampling

Against this backdrop, this section considers a simple sampling problem in which the purpose is to estimate the magnitude of total harm for a population which comprises three segments, each of which is sampled randomly, but with different sampling proportions. It shows that, in these circumstances, the "simple multiplier estimate" - of the type described above in section 2, in which the magnitude of aggregate sample harm is multiplied by the inverse of the proportionate sample size - is typically biased. It establishes what other information is needed to provide an unbiased estimate. This provides the methodology for the next section in which the magnitudes of deterred and undetected harm are simulated under alternative scenarios based on the above inferences.

### 5.1 Notation

A population comprises three, not necessarily equal, segments ( $L, M, H$ ), ordered by the size of anticompetitive harm. The lower tail of least harmful cases accounts for a proportion  $P_L$  of the population, and proportion  $H_L$  of population harm; the upper tail for  $P_H$  and  $H_H$ , and the middle segment for  $P_M = (1 - P_H - P_L)$  and  $H_M = (1 - H_H - H_L)$ , respectively.

Each segment is sampled randomly, but with different sample proportions  $\lambda_i$  ( $i = L, M, H$ ). For notational convenience, write  $\lambda_L = \delta_L \lambda_M$  and  $\lambda_H = \delta_H \lambda_M$ .<sup>19</sup> The parameters  $\delta_L$  and  $\delta_H$  will be referred to as the sampling differentials and are employed in the next section to capture different scenarios about differentials in deterrence and detection rates between the three segments.

Suppose that the aggregate proportionate sample size,  $\lambda = \sum_i \lambda_i P_i$ , and magnitude of sample harm,  $H^s$ , are both known. Denoting the magnitude of population harm by  $H$ , then the "simple multiplier estimate",  $H^s/\lambda$ , is an unbiased estimator of  $H$ , if the sample is random overall.<sup>20</sup> However, in this case, it is unlikely that the sample will be random.

In the above notation:

$$\lambda \equiv \lambda_L P_L + \lambda_M P_M + \lambda_H P_H \equiv \lambda_M [1 - (1 - \delta_L) P_L - (1 - \delta_H) P_H] \quad \text{and} \quad (2)$$

$$H^s/H \equiv \lambda_L H_L + \lambda_M H_M + \lambda_H H_H \equiv \lambda_M [1 - (1 - \delta_L) H_L - (1 - \delta_H) H_H] \quad (3)$$

Thus

$$H^s/H \equiv \lambda + \lambda_M [(1 - \delta_L)(P_L - H_L) + (1 - \delta_H)(P_H - H_H)] \quad (4)$$

where  $\lambda_M = \lambda \{1 - (1 - \delta_L) P_L - (1 - \delta_H) P_H\}^{-1}$

---

<sup>19</sup>Thus we approximate continuous relationships between the sampling rate and case harm with a simple three-step function, in which the magnitudes of the  $\delta$  are sensitive to how broadly defined are the sizes of the different segments ( $P_i$ ). Although analytically crude, this device is nevertheless sufficiently flexible for present purposes - the relative magnitudes of the  $\delta$  can capture monotonicity or not, and concavity, linearity or convexity in the relationship between sampling and harm.

<sup>20</sup>In conventional notation, for a sample  $n$  with sample mean  $x$ , drawn from a population  $N$  with mean  $\mu$ , the magnitudes of sample and population harms are  $nx$  and  $N\mu$  respectively. If the sample is random,  $E(x) = \mu$ , the ratio of the magnitude of sample harm to proportionate sample size has  $E(nx)/(n/N) = N\mu$ .

Thus

$$H = \frac{H^s}{\lambda + \lambda_M[(1 - \delta_L)(P_L - H_L) + (1 - \delta_H)(P_H - H_H)]} \quad (5)$$

**Proposition 3** *With random sampling across the three segments ( $\delta_L = \delta_H = 1$ ), the "simple multiplier"  $H^s/\lambda$  is an unbiased estimator of aggregate population harm. With differential sampling across segments, the estimator will be typically biased, and the direction and magnitude of bias will depend on (i) the sampling differentials and (ii) the relative sizes of mass in the two tails of the population:*

$$H^s/\lambda \begin{cases} \geq \\ \leq \end{cases} H \text{ as } (1 - \delta_L)(P_L - H_L) \begin{cases} \geq \\ \leq \end{cases} (1 - \delta_H)(H_H - P_H) \quad (6)$$

The impact of the sampling differentials is obvious and intuitive: for a given population (described by  $P_L, H_L, P_H$  and  $H_H$ ), upward bias is more likely the larger is sampling of the upper tail ( $\delta_H$ ) relative to sampling of the lower tail ( $\delta_L$ ). Intuition on  $(P_L, H_L, P_H$  and  $H_H)$ , is less clear and Figure 2 introduces a traditional Lorenz curve to aid interpretation. Recall that, if a population is ranked in ascending order of size (here, harm) along the horizontal axis, the Lorenz curve shows the proportion of aggregate harm accounted for by those cases: i.e. how the proportion of the harm accounted for by the proportion of least harmful cases  $x$  increases with  $x$ <sup>21</sup>. Both axes are therefore normalised to the range, zero to 1, the 45 degree diagonal provides the symmetric benchmark where all cases involve identical harm. The curve, which must be concave to the diagonal because cases are ranked in ascending order of size, lies further below the diagonal the more asymmetric is the harm size distribution.

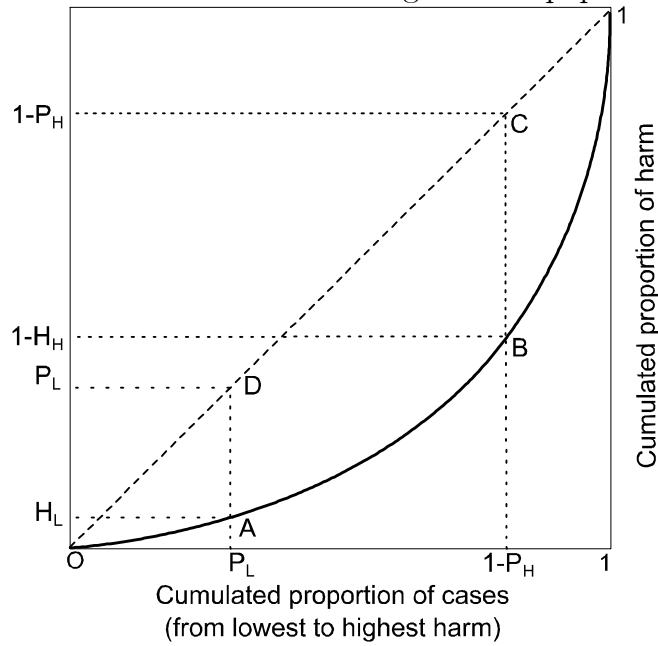
Noting that, in the Figure,  $(P_L - H_L) = AD$  and  $(H_H - P_H) = BC$ , it follows that

$$H^s/\lambda \begin{cases} \geq \\ \leq \end{cases} H \text{ as } AD/BC \begin{cases} \geq \\ \leq \end{cases} (1 - \delta_H)/(1 - \delta_L) \quad (7)$$

---

<sup>21</sup>The Lorenz curve often proves helpful in the analysis of positively skewed distributions. Traditionally, 'size' might be personal income in studies of income distribution or firm size in studies of industrial concentration; see Lorenz (1905), or Gastwirth (1972).

Figure 2: The Lorenz curve with three segments of population and harm



The magnitudes of  $AD$  and  $BC$  depend on the relative sizes of the tails and the degree of population asymmetry, represented by the distance of the Lorenz curve from the diagonal, but the direction of bias also depends on the precise shape of the Lorenz curve, and therefore the form of the underlying population distribution. A specific paramaterisation of the Lorenz curve proves to be very convenient for present purposes.

## 5.2 The symmetric Lorenz curve (lognormal distribution)

If the Lorenz curve is symmetric either side of a downward sloping main diagonal, then, for any sized lower tail  $P_L$  accounting for harm  $H_L$ , there is an upper tail of size  $P_H = H_L$  which accounts for harm of size  $H_H = P_L$ , e.g. if the 30% lowest harm cases account for 10% of total harm, then the 10% most harmful cases will account for 30% of harm. This leads to:

**Proposition 4** *If the Lorenz curve is symmetric, the simple multiplier will produce*

upward (downward) biased estimates of population harm if the upper tail is more (less) heavily sampled than the lower tail; and, for a given total sample size, the magnitude of bias will be greater the more asymmetric is the population size distribution.

**Proof.** For a symmetric Lorenz curve,  $AD = BC$ , i.e.  $(P_L - H_L) = (H_H - P_H)$  and (15) simplifies to:

$$H = \frac{H^s/\lambda}{1 + (\delta_H - \delta_L)(P_L - H_L)(\lambda_M/\lambda)} \quad (8)$$

and  $H^s/\lambda \gtrless H$  as  $\delta_H \gtrless \delta_L$ , and for given  $(\delta_H - \delta_L)$ , it is absolutely increasing in asymmetry,  $(P_L - H_L)$ . ■

Lorenz-symmetry is clearly analytically convenient here, but two points should be noted. First, it requires that we define the tails in a relative way: the size of the upper tail is defined by the size of the lower tail (in the sense that  $P_H = H_L$ ). However, this is less arbitrary than it might appear, since the trichotomous segmentation is itself only an approximation to capture, in a simple way, underlying continuous, but not necessarily monotonic, relationships between  $\omega$  and  $\sigma$  and harm. In that sense, there are no 'true' definitions of the lower and upper tails; although, for a given underlying relationship, different delineations of the tails clearly imply different values for the  $\delta$ .

Second, and more generally, most size distributions need not generate symmetric Lorenz curves. However, there is at least one distribution which does: the lognormal (Aitchison and Brown, 1957, p.113). Since this has a strong pedigree and is found to be widely applicable, at least as a first approximation, to many real world business size-related distributions - we employ this as our base assumption in the calibrations below.

As a sensitivity test we also use Pareto's curve - another equally plausible positively skewed distribution but which does not necessarily generate symmetric Lorenz curves (see Appendix 2).



## 6 Estimating deterred and undetected harm

This section applies Eqn.(5) to simulate the magnitudes of undetected and deterred harms, assuming that estimates are available for (i) the proportionate sample size ( $\lambda$ ), which is derived from extraneous estimates of  $\omega$  and  $\sigma$ , available from surveys such as those discussed in Section 2, and (ii) the magnitude of sample harm detected and rectified by a CA ( $H^s$ ), as reported in its annual impact evaluation. This is illustrated by using the European Commission's (EC) own estimates of remedied harm, and calibrating with parameter values in line with the expectations derived earlier in Table 2.

### 6.1 Notation and methodology

The purpose is to estimate population harm from detected harm (i.e. sample harm above). Denoting by  $H^{UDT}$ ,  $H^{DR}$ ,  $H^{UDR}$  the undetected, deterred and undeterred harms respectively, then aggregate population harm  $H$  is related to detected harm  $H^{DT} = H^s$  by:

$$H \equiv H^{UDT} + H^{DT} + H^{DR} \quad (9)$$

Note that each is defined as absolute magnitudes, rather than the proportions denoted previously by  $H_L$ ,  $H_M$  and  $H_H$ .

We proceed in three stages. Eqn.(5) is first applied to estimate total population harm and second to estimate undeterred harm. Then, deterred and undetected harms can be backed out as residuals from Eqn.(5).

*Stage 1: estimating total population harm ( $H$ )*

If population harm in Eqn.(5) is defined as the harm of all potential anticompetitive cases, and sample harm is detected harm,  $H^{DT}$ , then the proportionate sample size is  $\lambda = (1 - \omega)\sigma$ . The differential sampling proportions each have two constituents - differentials in the undeterred and detection rates:

$$\delta_L = \delta_L^{UDR} \delta_L^{DT} \text{ and } \delta_H = \delta_H^{UDR} \delta_H^{DT}$$

$$\text{where } \delta_L^{UDR} = \frac{(1 - \omega_L)}{(1 - \omega_M)}, \delta_H^{UDR} = \frac{(1 - \omega_H)}{(1 - \omega_M)}, \delta_L^{DT} = \frac{\sigma_L}{\sigma_M}, \delta_H^{DT} = \frac{\sigma_H}{\sigma_M}$$

*Stage 2: estimating undeterred harm ( $H^{UDR}$ )*

If the population is now re-defined to refer only to all undeterred cases, then  $H$  is replaced in Eqn.(5) by aggregate undeterred harm  $H^{UDR}$  and the sample harm is again  $H^{DT}$ , but detected cases are now a sample of proportionate size  $\lambda = \sigma$ . In this case, the  $\delta$  refer only to differentials in detection rates between the low and medium,  $\delta_L^{DT}$ , and medium and high segments,  $\delta_H^{DT}$ .<sup>22</sup>

*Stage 3: estimating deterred and undetected harms*

Deterred harm can now be calculated as the residual ( $H - H^{UDR}$ ), and undetected harm as the residual ( $H^{UDR} - H^{DT}$ ).

## 6.2 Calibration: an illustrative case study

This is illustrated by employing estimates from the EC of the magnitudes of harm it publishes as avoided/rectified from of its cartel and merger interventions in 2010:  $H^{DT} = 7.2$  and 5.0 billion EUR respectively. The EC explains that these estimates are based on very conservative estimates. For values of  $\sigma$  and  $(1 - \omega)$  we use those derived earlier in Table 1.

If these samples were random, then total potential population harm would be  $H = H^{DT}/[\sigma(1 - \omega)]$ . For comparative purposes these are shown for cartels and mergers in Table 3 as the  $H^S/\lambda$ : 79 and 30 billion EUR respectively. These are disaggregated into deterred, accounting for 36 and 25 billion EUR, and undetected, accounting for 36 billion and 0 respectively.

**Assumptions for calibrated parameters** In order to simulate using Eqn.(5), calibrations are needed for the  $\delta$ ,  $P_L$  and  $H_L$ .

(i) *Asymmetry of population distribution ( $P_L$  and  $H_L$ .)*

---

<sup>22</sup>Given that the 'population' is now confined to only the undeterred cases, the segment proportions ( $P_i$ ) now refer to the proportions of undeterred cases  $((1 - \omega_i)P_i)$ .

For a symmetric Lorenz curve, the extent of asymmetry in the population is reflected by the mass in the lower tail - more precisely, the difference between the proportion of cases and proportion of harm it accounts for. As base values, these are set at 0.4 and 0.05: for a lognormal distribution, this corresponds to a standard deviation of log harm of approximately 1.39<sup>23</sup>, which is a fairly typical estimate from previous empirical firm size studies (see for example Cabral and Mata (2003), and is appropriate if the size of infringing firms is a main driver of  $h$ ). Two alternatives to this base case are also considered: the more symmetric ( $P_L = 0.4$  and  $H_L = 0.1$ ) and less symmetric ( $P_L = 0.5$  and  $H_L = 0.05$ ), corresponding to standard deviations of log harm of 1 and 1.645 respectively.

*(ii) Sampling differentials ( $\delta$ )*

The calibrations here are based on the expectations summarised earlier in Table 2. For cartels, in the base case, the probabilities of both deterrence and detection both decline monotonically with harm - but at a declining rate for detection, which reflects the expectation that detection via leniency is considerably more likely for low harm cartels than for medium and high harm cartels. As alternative, we also consider (i) a linear decline in detection between the three segments, and (ii) constancy of the deterrence rate across the segments. For mergers, a U-shaped deterrence probability is assumed, being highest for the low and high harm tails. Detection is assumed to be certain.

To aid interpretation, Table 3 also reports the segment values of  $\omega$  and  $\sigma$  implied by each pair of hypothesised  $\delta$  values.<sup>24</sup>

### 6.2.1 Results (i) Symmetric Lorenz curves (lognormal)

Table 3 reports calculated population harm, in aggregate and then disaggregated into undetected and deterred harms, for each scenario. Different scenarios are presented

---

<sup>23</sup>A lognormal variate, mean  $\mu$  and variance  $\sigma^2$ , generates a first moment distribution which is also a lognormal, with mean  $\mu + \sigma^2$  and variance  $\sigma^2$ , see Aitchison and Brown (1957). It follows that  $z(P_L) = z(H_L) + \sigma$ , when  $z$  denotes the standard normal transformation.

<sup>24</sup>Note that the range of permissible  $\delta$  values are constrained by the necessity for all segment  $\omega$  and  $\delta$  to lie in the range 0 to 1.

in each column and the final columns report what total population harm would have been using the 'biased' multiplier  $H^S/\lambda$ .

Table 3: Simulated results for symmetric Lorenz curve

	<b>CARTELS</b> ( $\omega=0.45, \sigma=0.17$ )						<b>MERGERS</b> ( $\omega=0.83, \sigma=1$ )				
	(1)	(2)	(3)	(4)	(5)	(6)	$H^s/\lambda$	(7)	(8)	(9)	$H^s/\lambda$
<b>Magnitude of harm</b>											
Population	149	181	94	79	133	172	<b>79</b>	41	39	47	<b>29</b>
Detected	7	7	7	7	7	7	<b>7</b>	5	5	5	<b>5</b>
Deterred	52	82	43	28	49	54	<b>36</b>	36	34	42	<b>24</b>
Undetected	89	92	44	44	77	110	<b>36</b>	0	0	0	<b>0</b>
Deterred pro-competitive mergers								1.73	3.29	1.96	
<b>Parameters</b>											
$P_L$	0.4	0.4	0.4	0.4	0.4	0.5		0.4	0.4	0.5	
$H_L$	0.05	0.05	0.05	0.05	0.1	0.05		0.05	0.1	0.05	
$\delta_L^{DR}$	0.75	1	1	0.75	0.75	0.75		0.8	0.8	0.8	
$\delta_H^{DR}$	1.25	1	1	1.25	1.25	1.25		0.1	0.1	0.1	
$\delta_L^{DT}$	4	4	1.25	1.25	4	4		1	1	1	
$\delta_L^{RR}$	0.5	0.5	0.75	0.75	0.5	0.5		1	1	1	
<b>Implied parameters</b>											
$\omega_L$	0.55	0.45	0.45	0.55	0.56	0.54		0.84	0.84	0.84	
$\omega_M$	0.40	0.45	0.45	0.40	0.41	0.39		0.81	0.81	0.80	
$\omega_H$	0.25	0.45	0.45	0.25	0.26	0.23		0.98	0.98	0.98	
$\sigma_L$	0.34	0.31	0.19	0.20	0.35	0.30		1.00	1.00	1.00	
$\sigma_M$	0.09	0.08	0.15	0.16	0.09	0.07		1.00	1.00	1.00	
$\sigma_H$	0.04	0.04	0.11	0.12	0.04	0.04		1.00	1.00	1.00	

**Cartels** Under *scenario 1*, the preferred base case, total potential population harm is estimated to be nearly double that derived from applying simple (biased) multipliers, and this difference comes from the estimate that undetected cases generate much larger anticompetitive harm. This total is even more pronounced when (*scenario 2*) deterrence is assumed to be invariant with respect to harm; i.e. the probability of deterrence in higher harm cases is larger than in the first scenario. In this case, the

difference between undetected and undeterred harms narrows significantly but is not removed. The two are more or less in balance when (*scenario 3*) the detection rate is less sensitive to case harm, but even here aggregate harm remains higher than implied by the simple multiplier. *Scenario 4* is included largely for pedagogic reasons - in this case deterrence increases and detection fall identically with case harm, and the net effect is that total population harm is exactly as predicted by the simple multiplier. *Scenarios 5 and 6* return to the base case, but now posit, respectively, less and more asymmetry in the population distribution. Predictably, total population harm is much higher, especially the undetected part, than under simple multipliers, but even with a relatively symmetric population distribution (*scenario 5*) this is substantially greater than with simple multipliers.

Inevitably, numerical simulations of this type, based on what is at best informed guess work, should not be taken literally. However, some qualitative conclusions seem fairly robust across scenarios, and these all derive from not unreasonable underlying assumptions that (i) the population distribution of cartel harm is positively skewed and (ii) that more harmful cartels are at least weakly less likely to be deterred or detected - put simply: very profitable (harmful) cartels are less likely to be deterred by policy, and insofar as they are more stable and less vulnerable to leniency programmes, also likely to avoid getting caught.

**Result 1** *The potential harm from cartels in an economy may be noticeably greater than usually supposed. Simple multipliers, based on estimates of how many cartels are undetected or deterred may underestimate this potential by as much as 50%. A significant part of this potential harm never actually happens, because it is deterred, and this is a positive finding for CAs. On the other hand, deterred harm is typically outweighed by the magnitude of undetected harm - at least in the scenarios considered here. In that very particular sense, the net balance for CAs is negative. This result is qualitatively robust with respect to the degree of population asymmetry, and the extent to which deterrence and detection vary with respect to harm.*

**Mergers** For mergers, we consider fewer scenarios because it is assumed that detection is certain and that the overall deterrence rate is high (5 out of 6 mergers are deterred), and this reduces the scope for significant differences related to harm.<sup>25</sup> In the preferred base case (scenario 7), deterred harm is nearly 50% higher than under simple multipliers. This derives from the assumption that virtually all highly harmful mergers are deterred, and this is not offset by a greater tendency for low harm cases to be deterred than medium harm. The two other scenarios (scenarios 8 and 9), which vary the degree of population asymmetry reveal that, as might be expected, this result is accentuated the greater is the degree of population asymmetry.

**Result 2** *The deterrence of anti-competitive mergers has an impact that is typically much larger than has previously been supposed - for example, as implied by simple multipliers.*

The intuition is fairly obvious if we pose the question: how often do mergers to monopoly occur, despite the fact that these would be potentially very profitable? Under our plausible assumptions, such mergers rarely occur because they are deterred by an effective, and well-recognised framework of competition policy and related institutions. This will be true to a greater or lesser extent no matter how asymmetric is the underlying population distribution of potential mergers. The effects from deterring pro-competitive mergers are discussed in Section 7.

### 6.2.2 Results (ii) Non-symmetric Lorenz curves (Pareto)

Although these two results are conditional on symmetric Lorenz curves, the assumption is driven largely by presentational convenience and is easily relaxed. In further simulations we have used Pareto's curve as an alternative. Like the lognormal, this comes with a strong pedigree from previous literatures of size distributions, but unlike the lognormal, does not necessarily imply Lorenz-symmetry.

---

<sup>25</sup>An important difference between mergers and cartels is that, while many cartels may go undetected, for those that are detected, the CA's decision is typically correct. For mergers, while detection is less of an issue, incorrect decisions are more likely. This is pursued in the next section.

However, a drawback of Pareto’s curve is that it is typically found to provide inaccurate fits for the lower tails of most real world positively skewed size distributions. We therefore assume that it describes only that part of the distribution above some critical value,  $h_0$ , which lies somewhere within our lower tail. In other words, harm is distributed in the medium segment and the upper tail by Pareto’s curve. We need not make any assumption about how harm is distributed within the lower tail, but for comparability with the base line scenarios in Table 3, it is assumed that the lower tail accounts for 40% of the population and 5% of the harm. The remaining 60% of the population, accounting for 95% of the harm, follows Pareto’s curve. Again as in the base case above, the upper tail accounts for 5% of the population, but its share of harm is not constrained to be 40%, but is simulated according to Pareto’s curve. Appendix B derives the relationship between  $H_H$  and  $P_H$  for the upper tail of a Pareto curve. Using these figures the harm in the upper segment is given by:

$$H_H = 0.95(0.05/(0.6))^{(\alpha-1)/\alpha} \quad (10)$$

As can be seen,  $H_H$  now varies with the value of Pareto’s inverse inequality parameter  $\alpha$ . Table 4 calibrates for three alternative values of  $\alpha = 1.1$ , and 2, denoting respectively very asymmetric and a more symmetric distribution. A third alternative,  $\alpha = 1.54$ , is also shown because this generates  $H_H = 0.4$ , i.e. Lorenz-symmetry. In this case, the Pareto and lognormal simulations are identical, but with  $\alpha = 1.1$ , there is a relatively greater mass, and  $\alpha = 2$  a relatively smaller mass, in the upper tail than with Lorenz-symmetry.

Comparing Tables 3 and 4, the base case simulations with Lorenz-symmetry generated total harms of 148 and 42 for cartels and mergers respectively, but now, when the mass in the upper tail is greater than under Lorenz-symmetry, these rise sharply to 214 and 85 respectively; while, when the mass in the upper tail is smaller than under Lorenz-symmetry, the harms fall, if less sharply, to 124 and 35. Disaggregating, undetected harm is the main driver of these differences for cartels, but deterred harm is for mergers.

Table 4: Magnitude of harm if harm has Pareto distribution

	<b>CARTELS</b>			<b>MERGERS</b>		
	(1)	(2)	(3)	(4)	(5)	(6)
$\alpha$	2	1.54	1.1	2	1.54	1.1
Detected	7	7	7	5	5	5
Deterred	46	52	63	30	37	80
Undetected	71	89	143	0	0	0
Population	124	148	214	35	42	85
Deterred pro-competitive mergers				1.50	1.76	3.61

## 7 Type 1 and 2 errors on mergers

So far it has been assumed that the CA does not make mistakes, by either incorrectly deterring or intervening in pro-competitive cases (Type 1), or by failing to deter or intervene in welfare-reducing cases (Type 2). This section shows how these errors can be incorporated in the framework and quantified within the methodology.

For deterrence, the only new issue is Type 1 errors for mergers. Type 2 errors are already represented in the framework as the undeterred harmful cartels/mergers. Type 1 deterrence errors cannot occur for cartels if one assumes that beneficial cartels can never occur. But the possibility of business chilling cannot be ignored, i.e. where pro-competitive mergers are deterred (Type 1). It might be argued that the quantitative impact is unlikely to be important. For example, the Deloitte (2007) survey reports that respondents generally perceived this as rarely, if ever, occurring, and Baker (2003) also argues that business chilling is minimal in practice. He draws this conclusion from a survey of the limited empirical research literature which evaluates ex-post the impact of mergers on efficiency, and concludes that mergers typically fail to achieve the efficiency gains firms anticipate or claim ex-ante. In our opinion this is an extravagant claim, implying that there is no such thing as a pro-competitive merger. Our approach is more cautious and poses the question 'what is the maximum, or upper bound, on the magnitude of business chilling?' As argued earlier, one important source of deterrence is the costs the firm anticipates it would incur



if it engages with merger control (administrative, legal etc). Where these costs exceed its private benefits, the merger will not be proposed. But this will be equally true whatever the source of the benefits - whether enhanced efficiency or increased market power - so long as engagement costs of approval are the the same for pro- and anti-competitive mergers.<sup>26</sup> If so, an upper bound on the benefits lost from pro-competitive mergers deterred will be the magnitude of deterred harm in the relatively less profitable harmful mergers in the lower segment of our framework.<sup>27</sup> In terms of our methodology, this can be quantified as the magnitude of lower tail harm,  $\omega_L H_L$ , which is easily recovered from the estimated harms in Tables 3 and 4: as reported in the last row of harms in the tables, namely EUR 1.5-3.6bn for the illustrative case study. These magnitudes are not at all trivial since they represent between 30% and 70% of the benefits claimed by the CA for its detected cases. However, they do not materially alter the two main results of the previous section.

For intervention, again this can be largely discounted for cartels. Type 1 is impossible assuming all cartels are anticompetitive per se, and Type 2 would imply that the CA does not prosecute a cartel of which it is aware<sup>28</sup>. For mergers the potential for both types of error cannot be excluded. Conceptually, Type 2 can be incorporated quite easily by introducing a further conditional probability into Eqn(1). Define  $\theta$  as the conditional probability that a case is intervened, given that it is undeterred and detected. Then, the number of intervened cases is:

$$C = (1 - \omega)\sigma\theta N \tag{11}$$

Thus  $1 - \theta$  is the probability of Type 2 error, and in order to simulate how this impacts on the above estimates, we would need to know its aggregate magnitude

---

<sup>26</sup>In fact this assumption is overly cautious if the firm anticipates that an anti-competitive proposal would probably require higher legal costs and a longer investigation stage than one which was based on highly plausible anticipated efficiency gains. To the extent that this is true, our approach is consistent with an upper bound.

<sup>27</sup>This assumes that mergers with anticipated efficiency gains which are so great as to exceed anticipated costs are never deterred.

<sup>28</sup>Strictly speaking, there may be cases where a CA does not intervene because it judges that the evidence is not sufficiently strong to rebut any appeal. In our framework, such cases would be included in the estimates of undetected harm.

and how it varies with case harm. We do not have such information, and this is a subject for future research. But to give an approximate idea of the magnitudes which might be involved, assume that the overall rate of Type 2 errors is relatively low ( $\theta = 0.9$ ) and invariant with case harm. Observed sample harm, as recorded by the CA (5.0 in Table 3) should now be interpreted as the magnitude of intervened detected harm. However, the ‘true’ harm of detected cases is  $5.0/0.9 = 5.6$ , of which the CA fails to recognise 0.6 due to Type 2 error. This should be subtracted from the previous estimates of deterred harm in Tables 3 and 4. Of course, the correction would be larger (smaller) if Type 2 errors are more likely for high (low) harm mergers. Turning to possible Type I errors, these would be cases where the CA has intervened and has calculated an estimate of harm detected, and therefore avoided. If, on the contrary, there is no net harm, then the starting estimates of detected harm are biased upwards and in our methodology, aggregate undetected and undeterred harms, which are derived as multiples of detected harm, will also be upward biased by the same proportion. Having said this, as already mentioned, CA estimates of positive impact are always based on very conservative estimates, and it is our view that the main source of Type 1 errors lies in deterrence, as discussed above, and further work should give this the top priority.

## 8 Concluding remarks

Phenomena following the Pareto principle - or its special case, the so called 80-20 rule - impact on every aspect of our lives. Very often a small proportion of causes (e.g. 20 per cent of cases) generate a very large proportion of effects (e.g. 80 per cent of the effects). For example the largest few websites generate the vast majority advertising income on the internet, or the largest few customers generate most revenues for businesses. In these cases the need to focus on the few cases that lead to most effect is well understood. We have shown above that very likely the same should be true when assessing competition policy. The largest harm cases may account for most of the anticompetitive harm, yet often no special attention is directed towards these cases.

The introduction raised two ambitious questions. Subject to some very important caveats, the above calibrations provide some very rough answers. First, how much potential harm is 'out there' in an economy? Pooling the simulations of Tables 3 and 4 if the CA detects (and removes) 7.2 billion of harm with its anti-cartel activities, simulated total potential population harm lies in the range 79 and 214 billion euros; and for anti-competitive mergers, detected cases of 5 billion suggest a population harm ranging from 30 to 85 billion euros. Second, how successful is the CA at combating this harm? Judged in these terms, success is the sum of detected and deterred harms, which ranges from 43-89 billion euros for cartels, and for mergers, by assumption, all potential harm is either detected or deterred.<sup>29</sup> In other words, the aggregate *potential* for anti-competitive harm dwarfs what is detected and recorded by a CA, by a magnitude of between 10 and 30 times for cartels and 6 to 17 for mergers; and very approximately, half of the cartel harm and all of the merger harm is combatted by the CA. The lower bounds in these ranges apply if observed (detected) cases are a random sample of the population, while the upper bounds give an indication of just how considerable selection bias might be in this context.

However, the caveats should not be understated. At best, these are a first stab at answering these ambitious questions, presented merely to illustrate what is possible. They are clearly sensitive to the assumptions and particular numerical calibrations of what is, at this stage, a highly stylised and simplified framework. The magnitudes of all estimates are dependent on the starting point values of sample detected harm and simple detection and deterrence rates. The latter in particular are speculative, being derived from largely survey interview data. The estimates of detected harm produced by CAs are typically derived using very conservative assumptions in order to avoid accusations of self-justification. This cautions against attaching over-reliance on the absolute magnitudes of our estimates, although the main qualitative results would be robust to using alternative plausible estimates.

A more conceptual limitation to the methodology is that it abstracts from Type 1 or Type 2 errors by the CA. While, in principle, Type 2 errors can be incorporated fairly easily as shown in Section 7, little is known about the precise magnitudes

---

<sup>29</sup>Alternatively, some 'failure' occurs to the extent that the CA makes Type 2 errors.

involved. The possibility of a perverse deterrent effect (Type 1) for welfare enhancing mergers is more problematic. This falls outside the framework because the population is defined to include only anti-competitive mergers, but this clearly merits further research.

However, and putting to one side these very real issues, the primary intended contribution of this paper lies not in the results of the calibrations, but is more methodological. It asks what sort of information would be needed to quantify the unknowns. It has highlighted two key building blocks. The first is a better understanding of the nature of the population distribution of harm. While the assumption of some form of positively skewed distribution seems highly likely, and the precise nature of the distribution (here lognormal versus Pareto) is probably of second-order importance, our choice of parameter values has been necessarily largely arbitrary. The second is the question of how the rates of detection and deterrence (and intervention) vary with case harm. Here, existing theory provides some leads about *a priori* expectations, but these are largely unsubstantiated empirically, and at this stage have been modelled using a simplified trichotomy between low, medium and high harm cases.

This provides some obvious priorities for future research. Empirically, further work is required to replace some fairly arbitrary calibrations and assumptions with defensible estimates. For example, for cartels, insights into the differences between the population and samples may be revealed by careful comparison of existing historical cartel databases which include cases drawn from different regimes, under some of which cartels were legal while in others they were illegal. Also laboratory experiments which explore how deterrence, derived from fines under leniency, varies with case harm could also be informative. For mergers, a better understanding of the potential population might be gained from a meta-analysis of the previous empirical literatures on the price raising effects of mergers. This may help identify the critical rate of cartel price increase which typically activates CA intervention, and how active merger enforcement affects the distribution. Existing research (as discussed in section 1) already provides some empirical insights into the magnitudes of deterrence, measured simply in terms of numbers of mergers, but further work might focus more on whether

deterrence impacts differentially on different merger types (delineated for example by unilateral versus coordinated effects, and by the market shares of merging parties). Finally, the present paper has been confined to mergers and cartels, and future work will examine how far the proposed methodology might also be applied to cases of monopoly abuse - where, arguably, the possibilities of Types 1 and 2 errors are even more pronounced.

## References

- Aitchison, J., Brown, J. A. C., 1957. *The Lognormal Distribution: With Special Reference to Its Uses in Economics*. Cambridge University Press, Cambridge.
- Baker, J. B., 2003. The case for antitrust enforcement. *Journal of Economic Perspectives* 17 (4), 27–50.
- 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), 1–8.
- Bigoni, M., Fridolfsson, S.-O., Coq, C. L., Spagnolo, G., 2012. Fines, leniency and rewards in antitrust. *RAND Journal of Economics* 43 (2), 368–390.
- Block, M. K., Nold, F. C., Sidak, J. G., 1981. The deterrent effect of antitrust enforcement. *Journal of Political Economy* 89 (3), 429–445.
- Boyer, M., Kotochoni, R., 2011. The econometrics of cartel overcharges. CIRANO Scientific Series 2011s-35.
- Bryant, P. G., Eckard, E. W., 1991. Price fixing: The probability of getting caught. *The Review of Economics and Statistics* 73 (3), 531–536.
- Cabral, L. M. B., Mata, J., 2003. On the evolution of the firm size distribution: Facts and theory. *The American Economic Review* 93 (4), 1075–1090.

- Chang, M.-H., Harrington, J. E., 2009. Modelling the birth and death of cartels with an application to evaluating antitrust policy. *Journal of the European Economic Association* 7, 1400–1435.
- Chang, M.-H., Harrington, J. E., 2012. The impact of a corporate leniency program on antitrust enforcement and cartelization. mimeo.
- Clougherty, J., Seldeslachts, J., 2013. The deterrence effects of u.s. merger policy instruments. *Journal of Law, Economics and Organization* Forthcoming.
- Combe, E., Monnier, C., Legal, R., 2008. Cartels: The probability of getting caught in the european union. BEER Paper 12.
- Connor, J. M., Bolotova, Y., 2006. Cartel overcharges: Survey and meta-analysis. *International Journal of Industrial Organization* 24 (6), 1109–1137.
- Cowell, F. A., 1977. *Measuring Inequality*. Phillip Allan, Oxford.
- Cramer, J. S., 1971. *Empirical Econometrics*, 2nd Edition. North-Holland Pub. Co.
- Crandall, R. W., Winston, C., 2003. Does antitrust policy improve consumer welfare? assessing the evidence. *Journal of Economic Perspectives* 17 (4), 3–26.
- Davies, S. W., Ormosi, P. L., 2012. A comparative assessment of methodologies used to evaluate competition policy. *Journal of Competition Law and Economics* 8 (4), 769–803.
- Deloitte, 2007. The deterrent effect of competition enforcement by the OFT. Office of Fair Trading OFT962.
- der Noll, R. V., Baarsma, B., Rosenboom, N., Weda, J., 2011. Anticipating cartel and merger control. Report commissioned by the NMa.
- Economics, L., 2011. The impact of competition interventions on compliance and deterrence. Office of Fair Trading OFT1391.

- Garrod, L., Lyons, B., 2011. Early settlements and errors in merger control. CCP Working Paper.
- Gastwirth, J. L., 1972. The estimation of the lorenz curve and gini index. *Review of Economics and Statistics* 54 (3), 306–316.
- Harberger, A. C., 1954. Monopoly and resource allocation. *The American Economic Review* 44 (2), 77–87.
- Harrington, J. E., 2004. Cartel pricing dynamics in the presence of an antitrust authority. *The RAND Journal of Economics* 35 (4), 651–673.
- Harrington, J. E., 2005. Optimal cartel pricing in the presence of an antitrust authority. *International Economic Review* 46 (1), 145–169.
- Houba, H., Motchenkova, E., Wen, Q., 2012. Competitive prices as optimal cartel prices. *Economics Letters* 114, 39–42.
- Lande, R. H., Connor, J. M., 2012. Cartels as rational business strategy: Crime pays. *Cardozo Law Review* 34 (2), 427–490.
- Levenstein, M. C., Suslow, V. Y., 2011. Breaking up is hard to do: Determinants of cartel duration. *Journal of Law and Economics* 54, 455–492.
- Lorenz, M. O., 1905. Methods of measuring the concentration of wealth. *Publications of the American Statistical Association* 9 (70), 209–219.
- Lyons, B., Davies, S. W., 2007. *Mergers and Merger Remedies in the EU: Assessing the Consequences for Competition*. Edward Elgar Publishing Ltd.
- Motta, M., Polo, M., 2003. Leniency programs and cartel prosecution. *International Journal of Industrial Organization* 21, 347–379.
- Ormosi, P. L., 2012. Evaluating the impact of competition law enforcement. OECD Working Papers DAF/COMP/WP2(2012)5.

Ormosi, P. L., 2013. A tip of the iceberg? the probability of catching cartels. *Journal of Applied Econometrics* Forthcoming.

Seldeslachts, J., Clougherty, J. A., Barros, P. P., 2009. Settle for now but block for tomorrow: The deterrence effects of merger policy tools. *Journal of Law and Economics* 52 (3), 607–634.

Sorgard, L., 2009. Optimal merger policy: Enforcement vs. deterrence. *Journal of Industrial Economics* 57 (3), 438–456.

Werden, G. J., 2008. Assessing the effect of antitrust enforcement in the united states. *De Economist* 156 (4), 433–451.

## A Literature review to Section 4

### A.1 Cartels

#### A.1.1 Motta and Polo (2003)

**Detection** Motta and Polo (2003) assume perfectly symmetric firms with three possible strategies: not collude (NC), collude and reveal (CR) and collude and not reveal (CNR). They derive ICs which establish which strategies firms choose given a CA with leniency programme. These are directly relevant for our purposes as this allows a comparison of the CR and the CNR equilibria and therefore provides an insight into how the probability of detection by leniency varies with cartel harm. They show that CNR dominates CR if the probability of prosecution ( $p$ ) is sufficiently small.

$$p < \frac{(1 + \delta)(\Pi_M - \Pi_N + R)}{\delta(\Pi_M - \Pi_N + F)} \quad (\text{A1})$$

Where  $\Pi_M$  and  $\Pi_N$  are the collusion and the competitive profit respectively,  $F$  and  $R$  are the normal and the leniency-reduced fine, and  $\delta$  is the discount factor. Denote  $h = \Pi_M$  and differentiate WRT  $\Pi_M$ :



$$\frac{(1 + \delta)(\Pi_M - \Pi_N + F) - (1 + \delta)(\Pi_M - \Pi_N + R)}{\delta(\Pi_M - \Pi_N + F)^2} \quad (\text{A2})$$

This is always non-negative because  $F \geq R$ . Therefore increasing  $h$ , the ICC in Eqn.(A1) relaxes, and firms are more likely to chose not to apply for leniency. Therefore when leniency is allowed, the probability of detection is a non-increasing function of  $h$ , i.e.  $\sigma_L > \sigma_M > \sigma_H$ .

**Deterrence** This framework also provides an insight into how deterrence changes with  $h = \Pi_M$ . Denote the deviation profit by  $\Pi^D$  and compare the NC and the CR and CNR equilibria. CR is sustainable over NC if the probability of investigating a cartel ( $\alpha$ ) is sufficiently small:

$$\alpha < \frac{\Pi_M - (1 - \delta)\Pi_D - \delta\Pi_N}{\Pi_M - \Pi_N + R} \quad (\text{A3})$$

Similarly, CNR dominates NC if  $\alpha$  is sufficiently small:

$$\alpha < \frac{(1 + \delta)(\Pi_M - (1 - \delta)\Pi_D - \delta\Pi_N)}{\delta p(\Pi_M - \Pi_N + F)} \quad (\text{A4})$$

Differentiating the RHS of both conditions WRT  $\Pi_M$  gives:

$$\frac{(\Pi_D - \Pi_N) + R + \delta(\Pi_N - \Pi_D)}{(\Pi_M - \Pi_N + R)^2} \quad (\text{A5})$$

and

$$\frac{(1 + \delta)(\Pi_M - \Pi_N + F) - (1 + \delta)(\Pi_M - (1 - \delta)\Pi_D - \delta\Pi_N)}{(\Pi_M - \Pi_N + F)^2} \quad (\text{A6})$$

As  $\Pi_D > \Pi_M > \Pi_N$  both expressions show that an increase in  $h = \Pi_M$  relaxes the ICC of collusion, therefore cartel formation (the complement of deterrence) is an increasing function of  $h$ , and  $\omega_L > \omega_M > \omega_H$ .

### A.1.2 Chang and Harrington (2009)

Arguably one of the most compelling theoretical models that analyses cartel deterrence in the presence of a CA and leniency programme is by Chang and Harrington (2009). The model endogenises the level of intervention by the CA as a function of their resources and workload. In their paper, the incentive compatibility condition is:

$$(1-\delta)\pi+\delta[(1-\sigma)h+\sigma(W-\gamma(h-\alpha\mu))] \geq (1-\delta)\eta\pi+\delta[W-\min\{\sigma,\theta\}\gamma(h-\alpha\mu)] \quad (\text{A7})$$

Where  $\delta$  is the discount factor,  $\pi$  is collusive profit,  $\sigma$  is the detection probability,  $W$  and  $h$  are the non-collusion and the collusion payoffs respectively (our notation),  $\gamma$  is a penalty multiplier,  $\alpha\mu$  is the mean competitive profit and  $\theta$  is the leniency multiplier. Analysing (A7), the LHS of the expression shows that if firms collude they realise profit  $\pi$ . With probability  $(1 - \sigma)$ , the cartel is not caught by the CA, and, given the industry is in the cartel state, the future payoff is  $h$ . With probability  $\sigma$ , the cartel is caught and convicted - which leads to a penalty of  $\gamma(h - \alpha\mu)$  and a future profit of  $W$ . On the RHS the profit from cheating is  $\eta\pi$ , which, by definition, causes the cartel to collapse and earn a future payoff of  $W$ . A deviating firm will apply for leniency iff the penalty from doing so,  $\theta\gamma(h - \alpha\mu)$ , is less than the expected penalty from not doing,  $\sigma\gamma(h - \alpha\mu)$ .

To establish how the ICC differs between high-harm and low-harm cartels, we take the differential of both sides WRT  $h$  keeping in mind that  $\sigma$  is also a function of  $h$  (assuming that  $\theta > \sigma$ , which follows from their discussion):

$$\begin{aligned} LHS & : \left\{ 1 - \frac{\partial\sigma}{\partial h}[h - W + \gamma(h - \alpha\mu)] - \sigma(1 + \gamma) \right\} \Delta h \\ RHS & : -\theta\gamma\Delta h \end{aligned} \quad (\text{A8})$$

Where, in the brackets,  $h - W$  is the incremental value of being in a cartel and  $\gamma(h - \alpha\mu)$  is the antitrust fine. Therefore increasing  $h$  the ICC relaxes (i.e. deterrence

lessens) if:

$$1 - \frac{\partial \sigma}{\partial h} [h - W + \gamma(h - \alpha\mu)] - \sigma(1 + \gamma) \geq -\theta\gamma \quad (\text{A9})$$

which always holds if  $\frac{\partial \sigma}{\partial h} < 0$ , which we assume to be true as discussed in Section 4.1.1. Thus  $\omega_L > \omega_M > \omega_H$ .

### A.1.3 Houba et al. (2012)

An example where leniency is ignored is Houba et al. (2012), who look at a linear oligopoly model where firms face CA enforcement without leniency but harm dependent detection probability and fines, notably they assume that cartel detection is a non-decreasing function of  $h$ , i.e.  $\sigma_L < \sigma_M < \sigma_H$ , which results in the following ICC (p.40):

$$\frac{1}{n} \frac{1 - \sigma k h}{1 - \delta + \gamma \delta \sigma h} h(2 - h) > h(2 - h) \quad (\text{A10})$$

where their  $p$  and  $\beta$  have been replaced by our  $h$  (harm) and  $\sigma$  (detection probability) for consistency with the notation used in the current paper. By simple differentiation it can be shown that an increase in  $h$  tightens the ICC. Therefore, in contrast with our Proposition, in the absence of leniency programmes, an increase in  $h$  would deter more, i.e.  $\omega_L < \omega_M < \omega_H$ .

### A.1.4 Harrington (2005) and Harrington (2004)

Neither of these models allows for leniency, rather, they assume that it is the over-time price variation that triggers detection and not the level of cartel price itself.<sup>30</sup> This leads to very different results from Houba et al. (2012). In Harrington (2004) firms choose a collusion price by balancing between maintaining cartel stability and avoiding the suspicions of the CA. In Harrington (2005) firms face an intertemporal tradeoff as higher price in the current period increases current profit but also increases future detection probability. The ICC in Harrington (2004) is rather complex and

---

<sup>30</sup>p.654 in Harrington (2004) and p.152 in Harrington (2005).

a closed form FOC for the ICC is difficult to reach; however it is clear that the IC loosens for a higher cartel price as long as the resulting increase in profit exceeds the increase in the probabilistic magnitude of damages.<sup>31</sup> In Harrington (2005, p.160) the ICC leads to the same conclusion on the relationship between cartel deterrence and  $h$ . Both of these models would therefore confirm that  $\omega_L > \omega_M > \omega_H$ .

## A.2 Mergers

### A.2.1 Baros et al. (2009)

Baros et al. (2009) assume that a merger may lead to pro-competitive and anti-competitive effects,  $\Pi^R$  and  $\Pi^C$  respectively, where  $\frac{\partial \Pi^R}{\partial h} > 0$  and  $\frac{\partial \Pi^C}{\partial h} > 0$  and to avoid corner solutions  $\frac{\partial^2 \Pi^C}{\partial h^2}, \frac{\partial^2 \Pi^R}{\partial h^2} < 0$  For every merger there is an endogenously chosen the level of harm  $h$ , that has distribution and density functions  $F(h)$  and  $f(h)$ . The CA clears the merger if  $h < \bar{h}$ , remedies the merger if  $\bar{h} < h < \bar{h} + \alpha$  and blocks the merger if  $h > \bar{h} + \alpha$ , where  $\bar{h}$  is the CAs reservation restriction level and  $\alpha$  denotes the extra restrictiveness level the authority is willing to accept as long as settlements are imposed.  $K$  measures the fixed cost of building up a merger proposal.

In this setting firms chose  $h$  such that:

$$V = 0 \times F(h - \alpha) + \Pi^R(h)[F(h) - F(h - \alpha)] + \Pi^C(h)[1 - F(h)] > K \quad (\text{A11})$$

Eqn.(A11) shows that if  $h$  is sufficiently small, the merger will not happen, i.e.  $\omega_L^M \approx 1$ .

As  $h$  increases  $V$  changes by:

$$\Pi^{C'}[1 - F(h)] - \Pi^C f(h) \quad (\text{A12})$$

Therefore mergers are less likely to be profitable at increasing  $h$  if:

---

<sup>31</sup>p.657 in Harrington (2004).

$$\ln(\Pi^C)' = \frac{\Pi^{C'}}{\Pi^C} < \frac{f(h)}{1 - F(h)} = \lambda(h) \quad (\text{A13})$$

Because  $\frac{\partial^2 \Pi^C}{\partial h^2} < 0$ , as long as it is true for the hazard rate that  $\lambda''(h) > 0$  (a likely proposition) then the above inequality holds for larger  $h$ , therefore  $\omega_H > \omega_M$ .

### A.2.2 Proving Inference 4

This section provides a proof for Inference 4 when the probability of regulatory intervention is a function of  $h$ . The proof is loosely based on Sorgard (2009). Let the probability of merger prohibition be  $\gamma(h)$ , where the CA is more likely to prohibit highly harmful mergers,  $\gamma'(h) > 0$ . The anticompetitive profit from merging is given by  $\Pi_h(h)$ , where  $\Pi'_h(h) > 0$ , and we assume concavity,  $\Pi''_h(h) < 0$ . Also, denote any increase in profit from efficiency gains by  $\Pi_e$ . Because we are only interested in anti-competitive cases, we can assume that  $\Pi_h > \Pi_e$ . If the proportion of anticompetitive effects that are remedied by the CA is  $\theta$ , (determined by the CA's characteristics), and the cost of the approval procedure is  $C$ , the merger will be proposed if:

$$\Pi = (1 - \gamma)(\theta \Pi_h + \Pi_e - C) - \gamma C > 0 \quad (\text{A14})$$

It is easily shown that this becomes less likely as  $h$  increases if:

$$\frac{(1 - \gamma)\Pi'_h}{\gamma'(h)} - \frac{\Pi_e}{\theta} < \Pi_h \quad (\text{A15})$$

Which is true for sufficiently high  $\Pi_h$ , therefore  $\omega_H > \omega_M$ .

Eqn.(6) also shows that if the anticompetitive profit is very small ( $\Pi_h \approx 0$ ) mergers are deterred if the profit from merging does not exceed the cost of approval. We appeal to this characteristic to argue that  $\omega_L \approx 1$ .

## B Alternative population assumption: Pareto's curve

Assume as an alternative to the lognormal parametrisation of the population distribution in the main text, Pareto's curve. This also has convenient analytical

properties, but does not generally generate symmetric Lorenz curves. Because it is typically found to provide inaccurate fits for the lower tails of most real world positively skewed size distributions, it is assumed here to fit only the medium and upper tails of the harm distribution.

Formally, suppose that all cases in the population with harm no less than  $h_0$  are distributed according to Pareto's Law. A convenient feature of the Law is that it also applies with any truncation point above  $h_0$  so long as the population is redefined accordingly. Therefore, we only require here that  $h_0$  should lie within our lower tail. In that case, with harms above this lower bound normalised by the lower bound,  $h^* = h/h_0$ , the cumulative distribution function, denoting the proportion of cases with harm less than  $h^*$  is:<sup>32</sup>

$$F(h^*) = 1 - (h^*)^{-\alpha} \quad \forall h^* > 1 \text{ and } \alpha > 1. \quad (A16)$$

Defining the three segments of the population, as in Figure 2, by two critical levels of harm:  $h_L$ , the upper limit of the lower tail, and  $h_H$  the lower limit of the upper tail, such that  $h_H > h_L > h_0$ ; then for  $h^* = h_H/h_L$ :

$$\frac{P_H}{1 - P_L} = \left(\frac{h_H}{h_L}\right)^{-\alpha} \quad (A17)$$

The equation of the Lorenz curve, depicting the proportion of harm caused by all cases with harm less than  $h$ , is:<sup>34</sup>

$$1 - L(F) = [1 - F]^{(1-1/\alpha)} \quad (A18)$$

Then, for  $h^* = h_H/h_0$ , where  $1 - L(F) = (H_H/(1 - H_L))$  and  $1 - F = P_H/(1 - P_L)$ :

$$\frac{H_H}{1 - H_L} = \left(\frac{P_H}{1 - P_L}\right)^{(1-1/\alpha)} \quad (A19)$$

In the present base case, for which  $P_L = 0.4$ ,  $H_L = 0.05$ ,  $P_H = 0.05$  (a lower tail accounts for 40% of the population and 5% of the harm, and an upper tail accounts

---

<sup>32</sup>We follow Cramer's notation and terminology (1971, pp. 51-58).

<sup>34</sup>Cowell (1977, p.156).

for 5% of the population), the upper tail share of harm (exclusive of the lower tail) is:

$$\frac{H_H}{0.95} = \left(\frac{0.05}{0.6}\right)^{(1-1/\alpha)} \quad (\text{A20})$$