A tip of the iceberg? The probability of catching cartels*

Peter L. Ormosi†

April 2, 2013

Abstract

Reliable estimates of crime detection probabilities could help design better sanctions and improve our understanding of the efficiency of law enforcement. For cartels, we only have limited knowledge on the rate at which these illegal practices are discovered. In comparison to previous works, this paper offers a more parsimonious and simple-to-use method to estimate time-dependent cartel discovery rates, whilst allowing for heterogeneity across firms. It draws on capture-recapture methods that are frequently used in ecology to make inferences on various wildlife population characteristics. An application of this method provides evidence that less than a fifth of cartelising firms are discovered.

JEL Classification codes: C18, D43, K21, L41

Keywords: crime detection rate, capture-recapture analysis, cartels, deterrence rate

---

*I am grateful for the useful comments to Iwan Bos, Stephen Davies, Joseph Harrington, Morten Hviid, Bruce Lyons, Franco Mariuzzo, Matthew Olczak, the participants of ZEW 2011 June and ETH Zurich Seminar Series and three anonymous referees. A special thanks goes to Ali Massadeh for his assistance with the dataset. The usual disclaimer applies. The support of the Economic and Social Research Council and the ESRC Centre for Competition Policy is also gratefully acknowledged.

†ESRC Centre for Competition Policy, University of East Anglia, NR4 7TJ, Norwich, United Kingdom, email: p.ormosi@uea.ac.uk

1 Introduction

Cartels are widely regarded as the most heinous form of anti-competitive practice resulting in serious harms to society in the form of higher prices, lesser quality, and/or reduced choice. Presently, we only have limited knowledge on the magnitude of this harm mainly because of insufficient information on what proportion of cartels are discovered. This reduces our ability to design sufficiently deterrent fines and to adequately evaluate the effectiveness of anti-cartel policies. Without knowing the probability of cartel detection an increased number of observed cases could mean either an increase in the detection rate of cartels or a decrease in the deterrence rate of anti-cartel enforcement, or both.[1] A change in the number of observed cases can only be meaningfully interpreted if, at the same time, the change in detection rate is also known.[2] The focus of this paper is on providing a method for estimating the probability of cartel detection, which has the potential to become a useful tool for future policy analysis on the effectiveness of cartel enforcement and the design of an effective anti-cartel enforcement regime.

Estimating the probability of cartel detection is a daunting task because, by definition, we have no information on cases that are undetected. Research which attempts to quantify this magnitude is still in its infancy, but we do have some estimates. Some of these works follow the seminal article by Bryant and Eckard (1991), who estimated that in the period between 1961 and 1988 the annual probability of cartel detection did not rise above 13-17% in any given year. Using the same method Golub et al. (2008) estimate US cartel detection rates for a later period and found similar figures. For the EU, Combe et al. (2008) derive a cartel detection rate of 12.9-13.3% for the period between 1969-2007. Using a different method Miller (2009) shows that the introduction of the 1993 leniency programme increased the rate of detection and strengthened deterrence.

One of the main limitation of previous estimates is that they provide a constant

[1] For the purposes of this paper detection rate is used synonymously to the probability of cartel discovery.

[2] For example an increase in the number of cases from year t to year $t+1$ paired with constant detection rate would mean reduced deterrence.

(time-invariant) detection rate and as such are not useful for deriving rates of change over time. This paper offers a way to estimate time-dependent cartel detection and survival probabilities. The applied method, most commonly referred to as capture-recapture (CR) analysis, is widely used and has been rigorously developed in ecology and epidemiology for estimating population parameters such as population size, or capture and survival rates. In the most simple form of CR analysis, the population size of a species is estimated by taking two successive random samples from the same population. The first sample is marked and replaced into the population. If the population does not change between the two independent sampling occasions, and all individuals are equally likely to be captured, then the proportion of marked subjects in the second sample should be an unbiased estimate of the ratio of all marked individuals to population size. To account for a more realistic scenario (e.g. continuously changing population, heterogeneity across individuals, time-dependence) a number of robust CR methods have been developed for estimating dynamically changing population characteristics in wildlife.

This paper relies on these CR methods to obtain estimates for the detection rates of cartelising firms. Parallels between wildlife and economic agents are drawn and it is argued that the assumptions used in ecological studies can be similarly interpreted for economic agents. By requiring only a limited number of assumptions, this approach provides a method more parsimonious than previous ones, one that is simple to use, and has only minimal data requirements. Finally, the methods discussed in the paper could be imported into other areas of law enforcement that are also characterised by high proportions of undetected behaviour (e.g. corruption cases, drug offenses, tax evasion, or drink driving).

The paper makes two key contributions. Firstly, it adds a new methodology to research in many areas of social sciences. Secondly it provides estimates of a key parameter of anti-cartel enforcement and shows that cartel detection stays between 10 and 20 percent most of the time between 1985 and 2009. The paper provides evidence that the introduction of the cooperative agreement between EU and US competition authorities had a significant impact on increasing detection rate - larger than the introduction of leniency programmes. It is also found that firms have a high 'death' rate (probability of exiting the market, or stopping to collude or improving collusion techniques so there is no subsequent capture) in the year immediately following capture. This may be a positive sign even if it means that firms are becoming better in avoiding detection as any such change would make maintaining collusion harder and increase the chances of breakdown.

The paper commences with a brief introduction of capture-recapture methods and a discussion of the assumptions. This is followed by a comparative discussion of this method in comparison to previous estimates. An application is then given on EC cartels detected between 1985 and 2010. Finally, the effect of various policy changes on the probability of cartel detection is estimated using different CR methods and the results are discussed.

2 The proposal

2.1 Capture-recapture methods

The paper proposes using capture-recapture (CR) methods to analyse cartel population characteristics. Since these techniques may not be familiar to the readers of economic literature, a brief introduction is sketched out here.[3] The idea of CR methods in ecology is to trap animals, mark and release them, and recapture a sample again on a number of occasions. By looking at the proportion of recaptured animals, inferences can be made on population parameters, such as population size, capture and survival rate.

The two main branches of CR methods are closed population (in which the population does not change through birth, death, or migration), and open population models. This paper uses a variation of open population methods, the Cormack-Jolly-Seber model (CJS).[4] CJS models are restricted in the sense that they rely solely on the recapture of previously marked specimens, and use maximum likelihood meth-

[3] For a detailed introduction to CR methods see: Amstrup et al. (2005), Williams et al. (2002) or Burnham and Anderson (2002).

[4] Cormack (1964), and Lebreton et al. (1992)

ods to estimate survival and capture probabilities (but not the population size).[5] Because it is conditioned on the previous capture of an individual, one of the limitations of this method is that - similarly to Bryant and Eckard (1991) - it only provides estimates for the subpopulation of marked individuals (and not for the total cartel population). However, the population of undetected individuals is - by definition - never detected, therefore the mean detection rate in the subpopulation of individuals never captured should be smaller or equal to the subpopulation of marked individuals, and detection rate estimates can be interpreted as an upper-bound estimate for the whole population. The following section provides a short introduction to a simple CJS model.[6]

2.2 A general CJS model

Let ϕ_t denote the probability of an individual surviving time $t = 1, 2, ...$, which is the conditional apparent survival from year t to year $t+1$, given that the same individual is 'alive' at the beginning of year t. It is apparent survival because an individual that has migrated outside the sampling area cannot be distinguished from one that has died. Denote the probability of an individual being captured at sampling occasion $t = 1, 2, ...$ by p_t.

To record data for a CR analysis, the information on the timing of captures is organised into an $n \times K$ matrix $\mathbf{X}$ (n is the number of individuals captured, K is the number of samples taken, $m \in n$, and $t \in K$), where $x_{mt} = 1$ if individual m was captured at sampling occasion t, and $x_{mt} = 0$ otherwise. Row m of $\mathbf{X}$ is the capture history of individual m. To each capture-recapture history, a multinomial probability with the parameters ϕ_t and p_t may be assigned, and the parameters can be estimated using a maximum likelihood method. The construction of the likelihood functions follows a very simple logic, explained through the following example. Take a time interval between t and $t+3$, where sampling is done at $t, t+1, ...$ An individual that

[5] Another frequently used CR model is the Jolly-Seber model, which gives estimates of population size as well, but assumes that captures of marked and unmarked individuals are equally probable; see: Jolly (1965), Seber (1965), Seber (1982).

[6] To understand the intuition behind CJS models, a simple explanation is given in Appendix A.

was captured at t and at $t+2$ but not seen at $t+1$, and $t+3$, will have a capture history of $CH_m = (1, 0, 1, 0)$. The probability of observing this pattern m is given by:

$$\Pr\{CH_m \mid \text{release at } t\} = \phi_t(1-p_{t+1})\phi_{t+1}p_{t+2}\left[(1-\phi_{t+2}) + \phi_{t+2}(1-p_{t+3})\right] \quad (1)$$

Note that although the individual is not seen at $t+1$, it survived, because it is later seen at $t+2$. Note also that, when creating the likelihood functions, no capture probability is associated to the first release, because the probability of any capture history is always conditional on capture and release at the first occasion.[7] This is why the estimated parameters can only be interpreted for individuals that have been captured at some point.[8] The expression in the squared brackets denotes the probability of not seeing the given individual after $t+2$ (i.e. there is no information on whether the individual survived after $t+2$ or not).

The probability of observing any other capture history can be similarly derived. By simple combinatorics the total number of possible capture histories can be calculated as $2^K - 1$ (where 1 is deduced because the combination consisting of zeros only is never observed), so for the above example there are a maximum of 15 distinct capture histories (this does not mean that all possible combinations are necessarily observed).

Following Pledger et al. (2003), denote the time of the first capture of individual m as f_m, the last capture as l_m, and the departure ('death' or migration) from the sample as $d_m (> l_m)$. Conditional on release at first capture, Equation 1 can be written in a general form (summing up for all possible departure times d_m, which is necessary as d_m is typically not observed):

$$\Pr(CH_m \mid f_m) = \sum_{d_m=l_m}^{K} \left\{ \left(\prod_{t=f_m}^{d_m-1} \phi_t \right) (1-\phi_{d_m}) \times \left(\prod_{t=f_m+1}^{d_m} p_t^{x_{mt}} (1-p_t)^{(1-x_{mt})} \right) \right\} \quad (2)$$

[7] This can be thought of as the individual being inserted into the sample at this time.

[8] Due to the reversibility of CR models, the likelihood could be conditioned on any capture occasions.

This parametric form assumes that both capture and survival rates are time-dependent but it is possible to use a model where these are assumed to be constant over time.

Using the individual capture history likelihoods and provided that all individuals are independent, the likelihood of observing all capture histories is therefore a product of the individual probabilities:[9]

$$L = \prod_{m=1}^{2^K-1} \Pr(CH_m \mid f_m) \tag{3}$$

Once capture histories are recorded for all captured individuals, the log of L can be used to find the parameters p_t and ϕ_t that maximise the likelihood of observing the recorded capture histories.

2.3 Applying CR methods to economic agents

The following thought experiment is intended to illustrate that the analogous use of ecological methods in economics is not far-fetched at all. In a simple scenario ecologists trap some of the individuals, put a tag on them, release them and allow time for them to mingle with uncaptured individuals. When they trap another group of the same species, the proportion of tagged animals in the second captured group is used to make inferences on capture and survival probability. This is very similar to the situation in anti-cartel enforcement: the competition agency collects samples from the market, takes a record of the firms that are involved in the cartel and then these firms are 'replaced' back in the population (following an investigation) and new samples are taken next year in which there are returning firms (firms that have been involved in another cartel) and firms without previous capture history.

There are many analogies between wildlife and economic agents, which can be put to use to find the right CR methods to be applied for analysing cartels. For example one has to allow for the fact that firms may start to collude, break down,

[9] The observed capture history is therefore an observation of a multinomial distribution of all possible capture histories.

or migrate during the period of analysis.[10] Another issue is that individuals are not homogeneous. CR methods allow the distinction between two subpopulation of colluding firms, one that is never caught (for example because their evasion skills are better than the others) and one where individuals are capturable.[11] The capturability of colluding firms may also vary within the sub-population of detected individuals (e.g. by size, industry, duration), and therefore the chosen CR method has to be able to control for firm-specific characteristics.[12] Another source of heterogeneity is called trap-response in CR literature: the probability of detecting colluding firms can increase (trap-happy) or decrease (trap-shy) as a result of previous capture.[13] These issues are all covered in the discussion below.

2.4 Interpretation of CR parameters for colluding firms

The two key parameters of CR methods are detection and survival probability, p_t and ϕ_t. Both of these parameters require a slightly different interpretation when used for economic agents as opposed to members of wildlife. p_t is the probability of exposure of a colluding firm at time t, where exposure can happen by dawn-raid, leniency application or a complaint. It is important to highlight that the analysis is firm-based not cartel-based.

Survival rate ϕ_t is an apparent survival estimate as it is not known whether the firm 'dies' because it does not exist any more, because it refrains from collusion in the future, or because it joins the subpopulation of those colluding firms that are never captured (i.e. they carry on colluding but are never captured again, for example because they improve their techniques to avoid regulatory detection). It is important that in none of these cases does the same firm enter into a captured cartel again. The temporary break-down of collusion does not mean the death of the cartelising firm in a CR setting - although certain assumptions will be made in this regard. Therefore survival in this case means that the cartelising firm is still in the

[10] See Amstrup et al. (2005), Chapter 3.
[11] Pollock et al. (1990), Chapter 4.
[12] As discussed by Pledger et al. (2003) for example.
[13] Crespin et al. (2008).

capturable subset (i.e. it still exist, it did not stop collusion forever, and it did not join the subpopulation of firms never captured).

2.5 A methodological comparison

This paper claims that using a CR method is a more parsimonious way to acquire cartel detection probability estimates. To see whether this is the case the following discussion focuses on the assumptions required for CR analysis in comparison to previous methods.

2.5.1 Stochastic cartel formation and breakdown

Similarly to this paper, Bryant and Eckard (1991) and Miller (2009) assume that cartel formation and breakdown follows a Markov process, within which firms or cartels can migrate between two states (compete and collude) according to some transition probabilities. Bryant and Eckard (1991) rely on Karlin (1966) who show that in a continuous time two-state Markov process the duration spent in either state is exponentially distributed with mean λ^{-1}, where λ is the probability of moving from the collusive state to competition. Bryant and Eckard (1991) uses this postulate to estimate the rate at which cartels break down from the reciprocal of the mean duration of observed cartels. This approach assumes that cartel breakdown is always caused by detection, which is not true even for the sample of detected cases let alone the whole population. For this reason the statement that the estimated sample mean λ_s is an upper bound for the population detection rate λ is misleading as the population λ also includes cartels that break down without detection. As we have no prior on this latter subset of cases, it cannot be concluded that $\lambda \leq \lambda_s$.

Harrington and Chang (2009) and Chang and Harrington (2010) design a theoretical framework based on a similar stochastic process, which suggests that the duration of cartels could be used for making inferences on cartel formation-breakdown parameters. The authors design a more complex cartel formation-breakdown model, and derive a theoretical distribution of cartel duration. They argue that cartel duration may be a good indicator of the effect of a policy change, and more effective policy changes the distribution of cartel duration by leading to detected cartels having longer duration (unstable, and therefore shorter cartels are not formed and are therefore not detected) in the short run and shorter duration in the long run (due to increased deterrence).[14] An advantage of this method is that inferences can be drawn on the total population of cartels (and not just on the observed ones), however it comes at a price of having to make more assumptions for the model to work.[15]

The key difference is that this paper does not directly derive detection estimates from the properties of Markov chains. In the most simple setting CJS models make the assumption that temporary migration between the two states (compete - collude) does not exist. In this case CJS provides unbiased estimates even if temporary migration exists but individuals are homogeneous with respect to migration probabilities and migration happens completely randomly (i.e. the number of individuals migrating to and from the population is the same). This seems like a strong assumption (may even be stronger than the ones required in previous methods) but is not needed when different versions of CR methods (robust design method) are used that allow for temporary migration. Later in the paper both CJS and robust design estimates are reported.

2.5.2 Time-dependent detection rate

Bryant and Eckard (1991) provide detection rate estimates that are constant over time. Technically it would be possible to get annual detection rates using the same method but it would be based on very small samples, which would fundamentally question the precision of these estimates.

As far as Miller (2009) is concerned, it does not provide estimates for the magnitude of detection probability, its focus is on the sign of change in detection rate as a result of the introduction of leniency programmes. Miller's theoretical framework assumes a homogenous, non-absorbing, first-order Markov chain, but - unlike Bryant and Eckard (1991) - this Markov chain allows for a distinction between cartels

[14] Zhou (2012) provides supporting evidence to this based on EC cartel data.

[15] Hyytinen et al. (2010) rely on this framework to estimate the probability of being in a cartel in Finland between 1951 and 1990, an era when cartels had to be registered and were not illegal.

breaking down naturally and as a result of detection.

A limitation of all of the previous works relying on Markov processes is that without homogeneity (i.e. with time-dependent transition parameters) convergence to a steady state cannot be achieved and therefore the model would have limited use for estimating effects of policy changes over time.

CR methods mean that transition parameters are not necesarily stationary, and detection (capture) and survival rate estimates are allowed to vary over time. These rates however do not have to be time-dependent. A test for time-dependence, comparing models with time-dependent and constant parameters, will be conducted below.

2.5.3 Homogeneity across firms and markets

Both Bryant and Eckard (1991) and Miller (2009) assume homogeneity across cartels, which derives from the use of the properties of Markov chains.

In its most simple form CJS methods also assume homogeneity in the following way:

1. The probability of any firm $m = 1, 2, ..., n$ being captured by the competition authority (CA) at sampling occasion t is given by p_t (provided that it has been captured at least once and that it has survived until t).

2. Any firm $m = 1, 2, ..., n$ surviving sampling occasion t has equal probability ϕ_t of survival to $t + 1$.

In these cases the estimated parameters can only be interpreted as an aggregate for all marked cartelising firms. However, in practice, the homogeneity assumption is rarely satisfied. An appealing feature of modern open population CR methods is that we can go beyond this and control for differences between the individual firms.[16] Two main sources of heterogeneity is addressed here: (1) trap-response; (2) firm/market characteristics.

[16] Pollock (2002) gives a complete overview of these methods.

Trap-response. Heterogeneity caused by 'trap-dependence' relates to the response of survival and capture parameters to previous captures. Trap-response could be permanent (marked firms showing different capture/survival rates to the ones never captured) or temporary (within the marked sub-population, parameters directly following capture are different). As far as permanent trap-response is concerned, it is an important premise of CJS models that captured firms are inherently different from the uncaptured sub-sample. As the proposed model only provides estimates for captured individuals, the homogeneity assumption is reduced to all marked cartelising firms having the same capture/survival probability (and not that marked and unmarked firms have equal capture and survival probabilities).

Temporary (or short-term) trap-response may exist for various reasons. Some firms might adjust their behaviour following capture to make future capture less likely or impossible (the latter case would mean the firm joining the subpopulation of firms never captured). Other firms may decide never to collude again. In these cases the survival probability immediately following capture would be lower.

Temporary trap-response is tested by estimating a model that allows 1-year, and 2-year dependence. Depending on whether the parameters are time-dependent or constant, there are numerous possible models. For example the likelihood function of a model with constant and temporarily (1 year) trap-dependent survival rate (i.e. survival rate is different, but only in the year directly following capture) and time-dependent capture probability is given by Equation 4 (note that in this case we only estimate two survival parameters, the one following capture, and ϕ_1 and one for all other years ϕ_2).

$$L = \prod_{m=1}^{2^K-1} \left\langle \sum_{d_m=l_m}^{K} \left\{ \begin{array}{c} \left[\prod_{t=f_m}^{d_m-1} \left(\phi_1^{x_{mt}} \phi_2^{(1-x_{mt})} \right) \right] (1-\phi_1)^{x_{md_m}} (1-\phi_2)^{(1-x_{md_m})} \\ \times \left(\prod_{t=f_m+1}^{d_m} p_t^{x_{mt}} (1-p_t)^{(1-x_{mt})} \right) \end{array} \right\} \right\rangle \quad (4)$$

Heterogeneous firms and markets. Firm/market specific characteristics can also violate the homogeneity assumption. The simplest way of addressing this would be to stratify the dataset based on the relevant characteristics. Assume that there are C distinct categories of cartelising firms, and the probability of belonging to either category is given by π_C. Each category has capture probability p_{ct} and survival probability ϕ_{ct} and the likelihood function is given by:

$$L = \prod_{m=1}^{2^K-1} \left\langle \sum_{c=1}^{C} \left[\sum_{d_m=l_m}^{K} \left\{ \left(\prod_{t=f_m}^{d_m-1} \phi_{ct} \right) (1-\phi_{cd_m}) \times \left(\prod_{t=f_m+1}^{d_m} p_{ct}^{x_{mt}} (1-p_{ct})^{(1-x_{mt})} \right) \right\} \right] \right\rangle \tag{5}$$

For cartels one could cluster firms based on industry, diversified or non-diversified firms, bid-rigging or price fixing cartels, leniency and non-leniency cases, growing or declining industries, etc.

Another way to account for heterogeneity would be to treat the parameters of interest to be random variables that are the function of certain covariates. A simple link function (such as logit, sin or loglog) could be used to express the relevant parameters as a linear function of a vector of explanatory variables. For example using the logit function: $p_t, \phi_t = \exp(\alpha_t' \mathbf{X}_t)/\left[1+\exp(\alpha_t'\mathbf{X}_t)\right]$, where $\mathbf{X}$ is an $n \times k$ matrix of k covariates for n firms. This allows controlling for time-dependent capture probabilities that are a function of time-dependent covariates. The problem with this method is that it largely increases the number of parameters, and reduces the model degrees of freedom.[17]

2.5.4 Firm or cartel as unit of analysis

Both Bryant and Eckard (1991) and Miller (2009) assume that cartels form an indivisible unit and that at any given time the entire industry either colludes or competes. This simplification has been used in economic literature on cartels for a long time. Bos and Harrington (2010) are among the first who account for the possibility of

[17] In a time-dependent model with l covariates affecting both survival and capture rates, there would be an additional $2K \times l$ coefficients to estimate.

industries to be non-inclusive and firms to individually leave/join the cartel and provide a more realistic treatment of the problem.

Due to the fact that it does not require the researcher to design an underlying theoretical model of cartel formation and breakdown, CR methods provide a flexible solution, where the researcher only has to decide on what the analysed conduct (individual) is. For example the analysis could equally focus on cartels as individual units or take a smaller unit than firms and treat all divisions within a firm as separate units for the purposes of the analysis.

2.5.5 Independence

The independence assumption states that capture and survival probabilities of individuals are independent of each other (independence is only needed for the marked subpopulation). Although this assumption is necessary for estimates in earlier works such as Bryant and Eckard (1991), the potential violation of this assumption, which may result in biased estimates, has not been addressed. As far as CR methods are concerned, one of the main sources of overdispersion (extra-binomial variation) is the violation of the independence assumption, leading to underestimated variances.[18] In the case of collusion it is unlikely in many instances for two cartels to be independent from each other.[19] Moreover, the same regulatory action (e.g. a dawn raid) may discover more than one cartels. The goodness of fit tests in Section 3.3.1 look at the amount of overdispersion caused by the violation of this assumption, and a treatment is also offered.

3 Application to European cartels

3.1 Institutional background

The last two decades has seen significant changes in the institutional design and the activity of the European Commission's anti-cartel enforcement units. The increase in

[18] See Anderson et al. (1994) for an analysis of overdispersion caused by lack of independence.

[19] A wildlife analogy would be species whose members exist in schools or flocks.

the number of prosecuted cartels has been due to two key factors: an improvement in cross-Atlantic antitrust cooperation, and the introduction of leniency programmes. The EU signed its first cooperational agreement with the two US competition authorities, the FTC and the DOJ in 1991. As part of this agreement US authorities notify the EU about the cartels they are investigating. If these cartels affect the EU market the Commission's investigation follows. Leniency programmes offer immunity from cartel fines in exchange for vital information for the detection and investigation of these cartels. The US introduced its leniency programme in 1978 but it did not produce the desired effects until a major revision of the programme in 1993. The EU introduced its programme in 1996 and modified it in 2002 in an extensive reform that increased the predictability of the way the programme is applied by reducing discretional elements in the programme and providing certain reductions in some cases. A surge in the number of EU cartel investigations followed these institutional changes and leniency programmes since then have been pronounced as the single most effective investigative tool in anti-cartel enforcement efforts.[20] Some of the more cautious opinions point out that this increase in enforcement activity may equally be caused by an increase in cartel activity.[21] Knowing how the probability of detection changed throughout this period could help to finally answer this question and to cast away doubts about the efficacy of leniency programmes.

One of the reasons CR methods may be applicable to colluding firms is the fact that many of these firms are repeatedly involved in cartels and can be captured on more than one occasion. EU guidance on penalties increases the amount of fine imposed on recidivist cartel members by up to a 100 per cent of the basic fine, which may reduce willingness to re-offend (and thus the possibility of recapture). However, up until the introduction of the current penalty guidelines in 2006 this has only been applied sporadically in a few cases.[22] For this reason it is unlikely to have had a significant impact on the period analysed in this paper (1985-2009), not least because in the period immediately following the 2006 penalty guidelines it is more

[20]See for example: Scott D. Hammond, "Cracking Cartels With Leniency Programs," OECD Competition Committee, Paris, France, October 18, 2005.

[21]See for example Spagnolo (2008) or Hinloopen and Soetevent (2008).

[22]Connor (2010).

likely that those cartels are detected that had been formed earlier (i.e. when the old penalty guidance was in force). Nevertheless, for future research if the penalty increments have the desired deterrent effect then this will have to be accounted for in studies of similar type.

Another issue with respect to repeat offences is that their detection may not be independent from previous detections. For example Akzo was caught in a cartel in 1997 (Sodium Gloconate, informed by US DOJ), 1998 (Choline Chloride, leniency by Bioproducts), 1999 (Monochloroacetic Acid, leniency by Clariant), 2000 (Organic Peroxides, leniency by Akzo), 2002 (Hydrogen Peroxide, leniency by Degussa), 2003 (Heat Stabilisers, leniency by Chemtura), and 2007 (Calcium Carbide, leniency by Akzo). In the brackets the cartelised product and the mode of detection are shown. Almost all of these cases were initiated by a leniency application, two of which were submitted by Akzo. This suggests that at least in some of the cases the detection may have been triggered by an earlier detection. However, there is no evidence that suggests that Akzo would not have come forward with leniency evidence had they not been detected in other cartels, and even if such evidence exists, it is very much sporadic. Neverthless this paper conducts a formal test to measure the level of overdispersion caused by potential dependence between cases. It also examines whether detection rate was significantly different in years immediately following an earlier detection. These results suggest that even if there is interdependence between detection observations it does not significantly reduce the reliability of CR estimates.

The remainder of this paper examines the possibility of applying CR methods to estimate the probability of cartel detection in the European Union. The model below will also test for the impact of institutional changes on the rate of detection.

3.2 The data

An appealing characteristic of CR methods is that the only data needed for estimating a general model such as CJS, is the date of the regulatory exposure of cartels. For this reason, the method can be just as well applied in the US, where the proportion of plea bargaining cases is high and information is therefore less easily available.[23]

The cartels included in the sample are the ones 'captured' by the European Commission (EC). For the purposes of data collection the case reports from the EC's cartel database are used. The first capture in the sample was in 1984, and the last one in 2009. In this time period 128 cases with 592 colluding firms were detected. Only illegal cartels were included in the sample, therefore estimates can only be interpreted for the (sub)population of illegal cartels. The data contains 842 offences, including 116 recidivist firms, and 250 recidivist offences or repeated captures. Table 1 shows the average length of time (in number of years) between recaptures in the sample.

Table 1: Average length before recaptures

	All recaptures	First recapture	Second recapture	Third recapture
mean	3.068	3.581	2.913	2.267
std.dev.	0.286	0.423	0.793	0.556
N	250	116	54	32

To account for the expansion of the EU in the study period it is assumed that the whole geographical area of study is sampled with equal intensity and if new areas are added to the sampling area, they have randomly chosen characteristics of the initial study area. It was assumed that marked individuals do not lose their marks. Although this assumption is typically more relevant to ecological studies, where animals are physically tagged, firms may also change their name during the period of analysis (e.g. as a result of mergers), which was accounted for when data was collected for the empirical analysis.

The data is organised in a $n \times K$ matrix $\mathbf{X}$, (n is the number of individual colluding firms captured, K is the number of years in the analysis, $m \in n$, and $t \in K$), where $x_{mt} = 1$ if an investigation in the t-th sampling year discovered that firm m

[23] Plea bargaining was not an issue for EC cases in the time of this analysis (pre-2009) but will be for future cases.

was colluding, otherwise $x_{mt} = 0$.[24] Row m of $\mathbf{X}$ is the capture history of firm m. As the emphasis is on capture probability, the capture history matrix does not distinguish between captures that revealed one cartel and captures where the investigation discovered more than one cartels as in the latter cases all the consequently discovered cartels are conditional on the one discovery made by the authority. Table B.1 in the Appendix contains an example of how capture history data is organised for the ML estimation.

3.3 Model fit

3.3.1 Goodness of fit

Before finding the best fitting model, the goodness of fit of the general model in Equation 2 is tested. The most frequently used measure of goodness of fit in CR estimations is the deviance of the general model from the saturated model, where the saturated model can be loosely defined as the model in which a parameter is estimated for each observation (the fit of which therefore should be 'perfect'). The deviance of the model is calculated following Lebreton et al. (1992), and the deviance statistic, which is χ^2 distributed, is used in the GOF test as: $\widehat{c} = \frac{deviance}{df}$. For the saturated model $\widehat{c} = 1$, therefore any upward deviation from 1 would imply extra multinomial noise in the tested model. This overdispersion is typically caused by the violation of the independence and the homogeneity assumptions.[25]

Another way of estimating overdispersion is by bootstrap GOF testing, which involves simulating capture histories for the tested model and comparing the simulated deviances with the deviance of the general model. This way the probability of observing the general model deviance is acquired, which is used to test the fit of the model. A measure of overdispersion in this case is given by $\widehat{c}_{bootstrap} = \frac{\text{model deviance}}{\text{mean of simulated deviances}}$.

$\widehat{c}$-statistics were acquired for the most general model in Equation (2) (where survival and capture parameters are time-dependent and captured individuals are homogeneous) and Equation 4 (where capture can have an immediate effect on the

[24]Firms are identified if they have at least once been captured by the Commission.
[25]Anderson et al. (1994).

otherwise constant survival rate - trap-response). Table 2 shows the measures of overdispersion for these two general models. The lower $\widehat{c}$-statistics imply that the model, which allows for an immediate trap-response (i.e. allows for some level of heterogeneity) provides a better fit than the other general model, and it produces less overdispersion.

Table 2: GOF of the most general models

	$\widehat{c}$	$\widehat{c}_{bootstrap}$
Equation 2	2.321	1.931
Equation 4	1.822	1.139

Lebreton et al. (1992) suggest that $\widehat{c} < 3$ indicates a tolerable level of overdispersion. Although the statistics in Table 2 all fall under this threshold (therefore a significant violation of the independence or the homogeneity assumptions can be rejected), to ensure the robustness of the estimates, the measured overdispersion was corrected by multiplying the covariance matrix by the dispersion parameter, as proposed by Cooch and White (2010).[26]

3.3.2 Model choice

The parametrisation of the general model in Equation 2 can vary depending on our assumptions about real life and they differ in how well they fit the data. For example, the fit of the model, where survival and/or detection rate is assumed to remain constant throughout the analysed period can be compared to a model where parameters are time-dependent. Finally, models, where heterogeneity as a result of trap-response is allowed, can also be tested against other models.

The notations used below follow a very simple logic: $\phi(t)$ and $p(t)$ denote time-dependent survival and capture parameters, and $\phi(.)$ and $p(.)$ refer to models with constant survival and capture probabilities respectively. $\phi(t/t)$ denotes a model

[26] White (2002) shows that in most circumstances $\widehat{c}$ performs better than $\widehat{c}_{bootstrap}$, therefore the likelihood terms provided below are adjusted by $\widehat{c}$.

where it is assumed that survival rates are influenced by a temporary trap-response (i.e. survival rate immediately following capture is different from survival rates in subsequent years), therefore two different survival estimates are assumed, both of which are time-dependent. The model denoted as $\phi(./t)p(t)$ assumes a trap-response effect for survival rates (the survival rate immediately following capture is constant throughout the whole period of analysis, and the subsequent survival rates are time-dependent), and that capture probability is time-dependent. The examples given in Equations (2) and (4) are denoted as $\phi(t)p(t)$ and $\phi(./.)p(t)$ respectively. 30 different parametrisations of the likelihood function in Equation 2 and Equation 4 have been estimated (the best fitting 15 are reported in Table B.2 in the Appendix), and the most efficient one was chosen using Akaike's Information Criterion (AIC). Low-AIC models are parsimonious, i.e. they fit the data well with a relatively small number of parameters The test statistics are presented in Table B.2, where AIC_c is the corrected AIC, $Delta\ AIC$ is the difference in comparison to the model with the lowest AIC.[27] Q implies that the likelihood terms were adjusted for overdispersion, and quasi-likelihood adjusted AIC is reported.[28]

The $QAIC_c$ *weight* in Table B.2 shows that given the dataset, model $\phi(./.)p(t)$ is around 2.8 times more likely than the second most efficient model.[29] The model likelihood column captures the same information, the relative likelihood of model $\phi(././.)p(t)$ to model $\phi(./.)p(t)$ is 0.393. Therefore the model with two different but individually constant survival rates (post-capture and later) and time-dependent capture probabilities is preferred. What this means for cartels, is that according to the best fitting model, detection rate varies over time and survival rate (i.e. the rate of not leaving the capturable subpopulation) is constant, however, the survival

[27] The AIC is given by: $AIC = -2\ln(\mathcal{L}) + 2M$, where $\mathcal{L}$ is the model likelihood, and M is the number of parameters estimated. As this AIC may be biased in certain circumstances, an unbiased version was given by Hurvich and Tsai (1989) and was used in this paper: $AIC_C = -2\log(\mathcal{L}) + 2M\left[n/\left(n - M - 1\right)\right])$.

[28] The software Mark was used for the estimations to follow: http://warnercnr.colostate.edu/~gwhite/mark/mark.htm

[29] Burnham and Anderson (2002) establish a set of rules of thumb, according to which, $\Delta AIC_c < 2$ supports evidence of no difference between two models, $2 < \Delta AIC_c < 7$ gives reasonable support for differnce between the two models, and $\Delta AIC_c > 7$ supports strong evidence that the two models are different.

rate of cartelising firms in the year immediately following capture is different from survival rate in subsequent years.

3.4 Parameter estimates and discussion

The primary purpose of this paper is to offer a method for estimating detection rates over time. For this reason, the economic and policy discussion of the results is limited, and the focus remains on the applicability of CR methods to economic phenomena. Further, and more in-depth interpretation is left to future works.

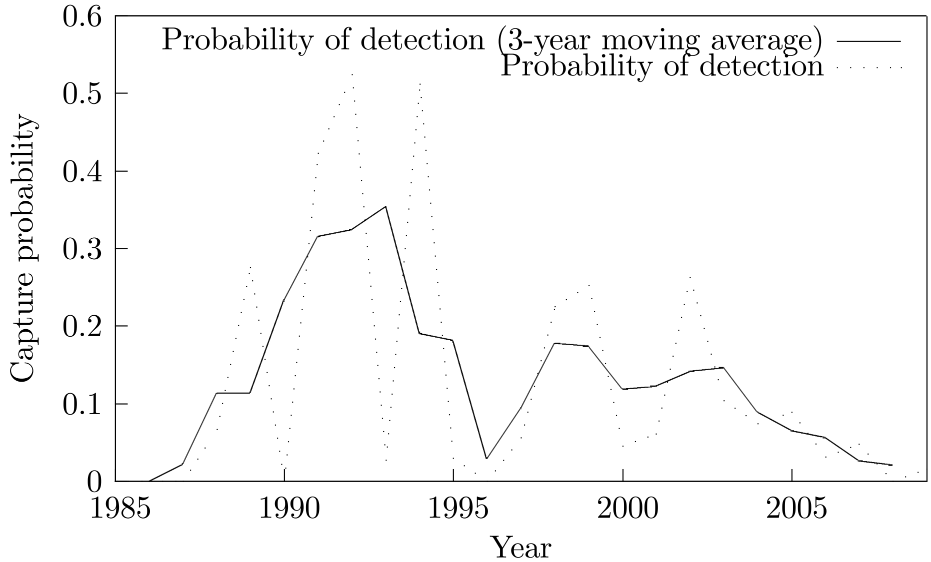
3.4.1 The detection rate of cartelising firms

Figure 1 plots the maximum likelihood estimates for $p(t)$ using the $\phi(./.)p(t)$ model as shown in Equation (4).[30] The estimates are also plotted as 3-year (mid-point) moving averages. As explained in Section 3.1, these estimates may be considered as an upper bound of detection rates for the total population of cartelising firms, much in the same way as in Bryant and Eckard (1991). The mean (over-time) estimate is similar to the constant detection rate estimated by Combe et al. (2008). The comparative advantage of the CR-method is that the change of capture rate over time can now be estimated, which provides a convenient tool for more profound policy analysis.

Figure 1 shows that the detection rate of cartelising firms stays under 20% most of the time (with the exception of a peak around 1993). The moving average line on Figure 1 reveals three peaks in cartel detection rate. One around 1993, one around 1998 and one around 2002. A possible reason for the first one is the 1991 EU/US cooperation agreement. The introduction of EC leniency programmes (1996) and the establishement of the Cartel Unit within the Commission (1998) and the 2002 reform of leniency programmes may explain the other - more moderate - peaks. The decline post-2005 is likely to be due to the fact that not all of the cases detected in this period have had their investigations closed by the time of this study.

[30] The table of estimations can be found in Table B.3 in the Appendix. The variance of these estimates is somewhat inflated due to the adjustment made in order to correct for overdispersion.

Figure 1: The estimated probability of cartel detection (1985-2009)



Although having time-dependent capture estimates may seem a useful achievement by its own merit, a more important question is whether the variation over time is explained by an underlying economic reason. An obvious reason for example for an increased detection rate would be explained by the agreement between EU and US competition authorities and/or the introduction of leniency programmes. The model below tests whether the difference between detection rates before, during and after these events is statistically significant from what we observed previously. This is done by using the best fitting model but now limiting the values that p can take to be constant over 3 and 5 year periods. This allows a distinction between periods before the 1991 EU/US Competition Cooperation Agreeent, after the agreement, following the introduction of EU leniency programmes, and following the 2002 leniency reform. The pre-1991 era was treated together for the relatively small number of detections, and post 2006 was also held together as estimates in the last period are always dependent on where the cutoff point was and full information was not available to all cases detected in 2006 onwards. These final year estimates may also be biased if a detection happened but it is not reported as the investigation has not finished before writing this paper.

Table 3 shows that the EU/US cooperation agreement contributed to a significant improvement in EU detection rates. This is likely to have been further boosted by the 1993 reform of US leniency programmes, which led to an increased number of cases in the US as well. After the introduction of the EU leniency programme detection rates remained higher than pre-1991 level. These findings are consistent irrespective of whether the window was taken as 3 or 5-year long.[31] Finally, in the post-leniency period detection rate remains above pre-1991 levels. How much of this is due to the US/EU agreement or the introduction of leniency programmes cannot be estimated from this data. The most curious finding though is that the impact of policy changes seem to wear off over time. This may suggest that firms adjust to policy reforms and become better at avoiding detection as they gain more experience about these reforms. Further research would be required to confirm this.

3.4.2 The survival of cartelising firms

The estimated survival parameters (and the 95% CI) are: $\phi_1 = 0.256$ [0.195; 0.327] for the year immediately following capture, and $\phi_2 = 0.905$ [0.841; 0.945] otherwise. It is not clear whether the subpopulation of uncaptured colluding firms 'live' longer than those that are eventually captured. For this reason survival rates can only be interpreted for the latter subpopulation, for example as the propensity to recidivist behaviour.[32]

The estimates show that there is a high chance of 'death' immediately following capture, i.e. colluding firms may cease to exist, or decide never to be in a cartel again, or simply become part of the subpopulation that is never captured again, for example because they become more wary of the EC's investigations.[33] If colluding

[31] The fit of these models is given in Table B.2 denoted as $\phi(./.)p(3yrs)$ and $\phi(./.)p(5yrs)$.

[32] According to the USDOJ (2002) price fixers tend to be recidivists. This is also confirmed by Connor (2010), who finds a recidivist rate that has grown to around 20% by 2009, this is roughly in line with the survival rates of this paper ($0.25 \times 0.9 = 0.225\%$).

[33] As explained earlier, these death rates do not include the temporary breakdown of cartels.

Table 3: Effect of leniency programmes on detection rate

3-year windows			5-year windows		
year	mean	[95% CI]	mean	[95% CI]	year
pre-1991	0.035	[0.018 - 0.067]	0.036	[0.018 - 0.068]	pre-1991
1991					1991
1992	0.271	[0.184 - 0.380]			1992
1993			0.275	[0.201 - 0.363]	1993
1994					1994
1995	0.178	[0.111 - 0.273]			1995
1996					1996
1997					1997
1998	0.192	[0.121 - 0.290]	0.118	[0.075 - 0.179]	1998
1999					1999
2000					2000
2001	0.130	[0.078 - 0.209]			2001
2002					2002
2003			0.116	[0.076 - 0.175]	2003
2004	0.101	[0.057 - 0.152]			2004
2005					2005
2006					2006
2007	0.034	[0.018 - 0.067]	0.036	[0.019 - 0.069]	2007
2008					2008
2009					2009

firms 'survive' the year following capture, then they have a much better chance of survival. To put it differently, many firms may figure out how to avoid detection following their capture, or may decide never to form a cartel again. This would mean that they leave the capturable subpopulation immediately following capture. Those that remain capturable are the ones who fail to do so, and the high survival rate following the year after capture shows that these firms are possibly the notoriously recidivist ones.

The problem is that formally 'death' cannot be interpreted in an unambiguous way.[34] The welfare consequences of 'death' because of avoiding future detection

[34] 'Death' means not seen again in the sampling period. It is possible that the same cartels would show up on the radar again in the future.

forever is clearly different from 'death' because of refraining from collusion in the future. Which one of the two dominates cannot be estimated from this model, because we do not observe whether a firm is not captured because it does not collude or because it colludes too well.[35] Intuitively, however, avoiding future detection would require higher efforts. This would make sustaining collusion harder and would be more likely to result in the breakdown of the cartel. This would suggest that even the situation where firms are not captured again because they become better at avoiding detection would have some positive effect. Therefore higher 'death' rates could always be interpreted as a desirable outcome.

3.5 Allowing for temporary cartel breakdown

The above estimates assume that entry into and exit from the population happens only once. This would mean that once a firm 'enters' the sample (by being caught) it remains in the population (goes on colluding) although may not be caught. The problem is that firms may temporarily leave the sampling area (i.e. the pool of colluding firms) and therefore become uncapturable. For example cartels typically break down after some time of operation and then may re-form later on. If the population is homogenous with respect to migration and this movement between collusive and competing states is completely random - i.e. at any time the same number of firms leave and enter the sample - then the above CJS model provides unbiased estimates (Burnham, 1993). Otherwise there are two ways to deal with the problem:

1. **Multi-state models:** assuming that movement between competition and collusion follows a Markovian process as originally described by Arnason (1972, 1973), and later by Schwarz et al. (1993). The problem is that using this method would require observations from both states but in reality we can never be sure that a firm observed as not being in a cartel is truly competing.

[35] Connor (2010) also points out that recidivism rate is very likely higher than the one observed, i.e. some firms carry on cartelising but as part of the uncaptured subpopulation.

2. **Robust design (RD) models**: Robust design models allow for situations where parts of the population may be temporarily unavailable for capture. Variations of RD models allow the estimation of both capture and migration parameters without prior knowledge on the unobserved subset, which makes it - at least conceptionally - suitable for the problem at hand.

Under CJS we assumed that a repeated encounter is a product of survival and capture, therefore return rate was expressed as $R = \phi \times p$. In Section 3.4.2 we acknowledged that survival is an apparent rate as it is a product of true survival and 'fidelity' (i.e. that the individual did not migrate permanently). Pollock's (1982) robust design (RD) models recognise that capture probability p is also a product of two factors, the probability of no temporary migration and the probability of true capture, which means that return rate is given by:

$$R = (S \times F) \times ([1 - \psi] \times p^*) \tag{6}$$

where S is true survival rate, F is the probability that the individual remains in the super-population (does not permanently migrate), ψ is the probability of temporary migration, and p^* is true detection probability (given the presence of the individual). Burnham (1993) discusses the possibility of using live encounter/dead recovery data to decompose ϕ. However, for the purposes of this discussion our focus is on capture probability, therefore the remaining discussion uses ϕ as apparent survival rate without further decomposition.

Apparent capture probability is given by $p = (1-\psi)p^*$. From this alone we cannot express $\gamma = 1 - (p/p^*)$ and $p^* = p/(1 - \psi)$ at the same time, so more information is needed. RD models provide this extra bit of information by distinguishing between primary and secondary sampling periods. To understand how this works assume that a capture history $(1, 0, 0, 0, 1, 1)$ is observed, where there are three secondary sampling periods, for example $(1, 0)$, $(0, 0)$, and $(1, 1)$. Within these respective secondary sampling periods the population is assumed to remain static. The sampling history at primary level is $(1, 0, 1)$ because the individual was captured in the first and the last secondary periods but not in the second one. At the primary level (between the secondary sampling sessions) the population is open to death and migration.[36]

As a closed population is assumed for secondary sampling periods, true capture rate p^* can be estimated from it.[37] At primary level open population is assumed and thus p can be estimated as explained in Section 2.2. From these the migration parameter is derived using the relationship $\psi = 1 - (p/p^*)$. The RD model is therefore a combination of open and closed-population models, where the full likelihood functions are products of the likelihood functions from the two sampling levels. At primary level Markovian migration between two states (A - collude and B - compete) is assumed (i.e. the probability of migration depends on the previous state). The movement parameters are ψ^{AB} and ψ^{BA}. The probability expressions of the above capture history at primary sampling level $(1, 0, 1)$ is given by:

$$\phi_1 \psi^{AB} \phi_2 \psi^{BA} p_3^* + \phi_1 (1 - \psi^{AB})(1 - p_2)\phi_2 (1 - \psi^{AB}) p_3^* \tag{7}$$

Although standard RD models allow migration between secondary samples, they assume closed population during secondary samples - i.e. firms are not allowed to leave or join cartels throughout these secondary periods.[38] To relax on this restriction, multi-state open robust design (MSORD) models allow entry and exit within secondary periods.[39] Although very powerful, these methods raise the issue of dimensionality as it introduces two more parameters migration to and from the available population in secondary periods.[40]

To make the model estimable on a sparse dataset parameters were assumed to be constant over secondary sampling periods. Even so, some of the parameters are

[36] When running these studies in wildlife the sampling sessions are designed to more realistically fit these assumptions, with sufficiently short intervals between secondary sampling periods and sufficiently long periods between primary sampling periods. For cartels as the sampling has been done previously by the CA this distinction is not possible, which reduces the reliability of RD methods.

[37] See for example Kendall et al. (1997).

[38] When used in wildlife secondary sample periods are sufficiently short for this assumption.

[39] For a detailed explanation of MSORD models the reader is referred to Kendall and Bjorkland (2001), and Kendall and Nichols (2002).

[40] Assuming full time dependence, it would require the estimation of 151 parameters for the cartel data at hand.

either confounded (and are thus not estimable separately) or sparsity of the data did not allow their estimation. Tables B.4 and B.5 in the Appendix report the estimates assuming 5 and 3-year long secondary periods.

The figures for detection probability are similar to the ones derived from using a CJS method. This seems to imply that the assumption behind the use of CJS models (i.e. that availability is homogeneous and completely random) may not be far from the reality of cartel formation and breakdown. What may also be interesting here is assuming migration between states survival probability increases close to unity. This is not surprising as the dataset controls for firms 'disappearing' due to mergers and thus the only true deaths here are firms that were completely liquidated from a market. The interpretation of the migration parameters is more difficult given that MSORD models distinguish between migration between primary states and migration throughout secondary states, which does not seem to be a meaningful distinction for cartels at first glance but further research on this would be required.

These preliminary results show that the CJS methods can provide reliable estimates of detection probability. However it seems important to check for any potential bias caused by temporary migration using RD or MSORD models even if sparse data requires simplifying assumptions on the model parameters.

4 Further robustness and sensitivity issues

The paper assumed that the CA engages in cartel enforcement (CR sampling) in discrete annual sampling periods $t = 1, 2, ...$ The relative advantage of using the 1-year model stems from the fact that only one capture per period is allowed in CR models. For this reason using the 2 or 3-year model would mean losing some information (2 or 3 captures in a 2 or 3-year model would only be recorded as 1 single capture). The 2 and the 3-year models also mean a larger deviation from the assumption of instantaneous sampling. Regarding the 6-month model, the number of captures for each period would be too small, which would reduce the precision of the estimates, and we would have twice as many parameters which may also cause a problem with small datasets.

The paper uses firms as a unit of analysis to allow for the possibility of individual firms joining or leaving cartels. However, repeat offences by large diversified firms could also be a result of decisions brought at the level of individual divisions, therefore the right unit of observation may be the individual divisions of firms. The problem is that using divisions as observation would create serious obstacles to empirical work firstly because of the lack of division-level data, secondly because of the difficulty in being able to decide whether the behaviour of divisions is determined at firm level, in which case division-level analysis would not be the most appropriate one. As for the interpretation of the results, the probability of capturing any individual division should by definition be smaller than (or equal to) the probability of capturing firms that embodies the given division. This means that the sign of this bias is the same (positive) as the sign of the bias caused by the fact that only the capturable subpopulation is used, therefore the estimated capture probabilities can still be interpreted as an upper bound of the real capture probability.

Finally, the paper acknowledges the possibility of heterogeneity accross firms with respect to cartel detection. One aspect of heterogeneity (trap-response) has been tested and results suggest very small overdispersion caused. Nevertheless, further work could improve the model fit by introducing firm and/or market specific heterogeneity. A simple way of doing this is to stratify the data into a small number of distinguishable groups of cases and use a model proposed in Equation (5). A source of heterogeneity could be caused by the expectation that more diverse firms are more likely to be recaptured (simply because they may be involved in a cartel in more than one industry). The sample could also be broken down into bid-rigging and price-fixing cases as these represent very different behaviour, however it is hard to say anything about how detection and survival parameters are expected to differ. Finally, stratification could also be introduced by industries. Further work is needed to explore these possibilities.

5 Concluding remarks

Building on the similarity between the wildlife trapping of animals and the discovery of illegal cartels, the paper uses capture-recapture models, - a method widely used for population studies in ecology - to study cartel detection and survival rates. Although previous estimates of the probability of cartel discovery exist, CR methods offer a more parsimonious way to estimate cartel detection rate, whilst having the following comparative advantages: (1) the method is atheoretical and thus requires fewer assumptions on cartel formation and break-down; (2) in its simplest form CR methods only need data on capture (encounter) times; (3) it allows the estimation of time-dependent parameters; (4) it offers the possibility of heterogeneous firms and markets; (5) analysis is firm-based, which allows the possibility of individual firms quitting or rejoining the cartel.

EU cartels are used to illustrate CR methods at work in economics. Estimates show in most years between 1985 and 2009 less than a fifth of firms in EU cartels are detected and that the EU/US cooperation agreement triggered a larger increase in detection rate than leniency programmes. In the year following capture around 75% of cartelising firms may cease to exist, or decide never to be in a cartel again, or become part of the subpopulation that is never captured again, for example because it becomes more wary of the EC's investigations. For firms that remain capturable later on, apparent survival increases to almost 90%. An important limitation of these results is that - similarly to Bryant et al. (1991) - they can only be interpreted as an upper bound given that the analysis, by definition, relies on colluding firms that have been detected at least once.

Given that the above estimates may be biased upwards (i.e. they can only be interpreted as an upper bound), it is possible that real detection rates are significantly smaller. Nevertheless, even if the estimates are biased, so long as the magnitude of this bias remains constant – and there is no a priori reason to think otherwise – time-dependent estimates could still be used to measure the change in detection probability over time. This means that it can become a tool for directly measuring the impact of policy changes on the probability of cartel detection or in any other area of social sciences that is characterised by a potentially high proportion of unobserved behaviour.

References

Amstrup S, McDonald T, Manly B. 2005. *Handbook of Capture-Recapture Analysis*. Princenton University Press.

Anderson DR, Burnham KP, White GC. 1994. AIC model selection in overdispersed capture-recapture data. *Ecology* **75**: 1780–1793.

Arnason NA. 1972. Parameter estimates from mark-recapture experiments on two populations subject to migration and death. *Researches on Population Ecology* **13**: 97–113.

Arnason NA. 1973. The estimation of population size, migration rates and survival in a stratified population. *Researches on Population Ecology* **15**: 1–8.

Bos I, Harrington JE. 2010. Endogenous cartel formation with heterogeneous firms. *RAND Journal of Economics* **41**: 92–117.

Bryant P, Eckard W. 1991. Price fixing: The probability of getting caught. *Review of Economics and Statistics* **73**: 531–540.

Burnham KP. 1993. *A Theory for Combined Analysis of Ring Recovery and Recapture Data*. Basel, Switzerland: Birkhauser Verlag.

Burnham KP, Anderson DR. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, 2nd edition.

Chang MH, Harrington JE. 2010. The impact of a corporate leniency program on antitrust enforcement and cartelization. *Mimeo.* Http://www.econ.jhu.edu/pdf/papers/WP548.pdf.

Combe E, Monnier C, Legal R. 2008. Cartels: The probability of detection in the european union. *BEER paper* **No. 12**.

Connor JM. 2010. Recidivism revealed: Private international cartels 1990-2009. *Competition Policy International* **6**.

Cooch E, White G. 2010. *Program Mark: A Gentle Introduction.* Publisher, 9th edition.

Cormack RM. 1964. Estimates of survival from the sighting of marked animals. *Biometrika* **51**: 429–438.

Crespin L, Choquet R, Lima M, Merritt J, Pradel R. 2008. Is heterogeneity of catchability in capture-recapture studies a mere sampling artifact or a biologically relevant feature of the population? *Population Ecology* **50**: 247–256.

Golub A, Detre J, Connor JM. 2008. The profitability of price fixing: Have stronger antitrust sanctions deterred? *available at: http://ssrn.com/abstract=1188515* .

Harrington JE, Chang MH. 2009. Modelling the birth and death of cartels with an application to evaluating antitrust policy. *Journal of the European Economic Association* **7**: 1400–1435.

Hinloopen J, Soetevent AR. 2008. Laboratory evidence on the effectiveness of corporate leniency programs. *RAND Journal of Economics* **39**: 607–616.

Hurvich CM, Tsai CL. 1989. Regression and time series model selection in small samples. *Biometrika* **76**: 297–307.

Hyytinen A, Steen F, Toivanen O. 2010. Cartels uncovered. *NHH Dept. of Economics Discussion Paper* **10/2010**.

Jolly G. 1965. Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika* **51**: 225–247.

Karlin S. 1966. *A First Course in Stochastic Processes.* New York: Academic Press.

Kendall WL, Bjorkland R. 2001. Using open robust design models to estimate temporary emigration from capture-recapture data. *Biometrics* **57**: 1113–1122.

Kendall WL, Nichols JD. 2002. Estimating state-transition probabilities for unobservable states using capture-recapture/resighting data. *Ecology* **83**: 3276–3284.

Kendall WL, Nichols JD, Hines JE. 1997. Estimating temporary emigration using capture-recapture data with pollock's robust design. *Ecology* **78**: 563–578.

Lebreton JD, Burnham KP, Clobert J, Anderson DR. 1992. Modelling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs* **62**: 67–118.

Miller NH. 2009. Strategic leniency and cartel enforcement. *American Economic Review* **99**: 750–768.

Pledger S, Pollock KH, Norris JL. 2003. Open capture-recapture models with heterogeneity: I. cormack-jolly-seber model. *Biometrics* **59**: 786–794.

Pollock KH. 1982. A capture-recapture design robust to unequal probability capture. *Journal of Wildlife Studies* **68**: 1–13.

Pollock KH, Nichols JD, Brownie C, Hines JE. 1990. Statistical inference for capture-recapture experiments. *Wildlife Society Monographs* **107**.

Schwarz CJ, Schweigert JF, Arnason NA. 1993. Estimating migration rates using tag-recovery data. *Biometrics* **49**: 177–193.

Seber G. 1965. A note on multiple-recapture census. *Biometrika* **52**: 249–259.

Seber G. 1982. *The Estimation of Animal Abundance and Related Parameters*. New York: Macmillan, 2nd edition.

Spagnolo G. 2008. *Leniency and Whistleblowers in Antitrust*. Cambridge, Mass.: MIT Press.

White GC. 2002. The use of auxiliary variables in capture-recapture modeling: An overview. *Journal of Applied Statistics* **29**: 103–106.

Williams BK, Conroy MJ, Nichols JD. 2002. *Analysis and Management of Animal Populations*. Academic Press.

Zhou J. 2012. Evaluating leniency with missing information on undetected cartels: Exploring time-varying policy impacts on cartel duration. *Tilburg University Discussion Paper* **No.353**.

Appendix A : Open population CR models - an intuitive explanation

Although both Amstrup et al. (2005) and Williams et al. (2002) provide a very comprehensive introduction to CR methods, this literature may not be familiar to the economist reader, therefore the main intuition behind the JS and the CJS models is given here. Consider the following notations:

p_t - the probability of capture in period t;

ϕ_t - the probability of survival from t to $t+1$;

M_t - the marked population size just before period t.

m_t - the number of individuals captured at sampling occasion t that are marked;

n_t - the total number of individuals captured at sampling occasion t;

R_t - the total number of individuals captured at sampling occasion t that are released (in the analysis of cartels all cartelising firms are assumed to be released after capture, therefore $R_t = n_t$);

r_t - the number of members of R_t that are captured again later;

z_t - the number of members of the marked population not captured at sampling occasion t, that are captured again later.

Consider an open population with the following two individual groups: $M_t - m_t$, which is the number of marked individuals that are not captured at t, and R_t (see above). Assuming equal capturability of individuals, the recapture rates of two distinct groups in the sample should be equal; therefore:

$$\frac{z_t}{M_t - m_t} \approx \frac{r_t}{R_t}$$

where the LHS is the recapture rate of marked individuals that are captured, and the RHS is the recapture rate of those captured From this M_t can be expressed, and used to derive estimators of survival and capture probabilities: $\phi_t = \frac{M_{t+1}}{M_t+R_t-m_t}$, and $p_t = \frac{m_t}{M_t}$, where $R_t - m_t$ is the number of newly marked individuals released at sampling occasion t. Although these closed form estimators provide an intuitive explanation to how CR methods work, in practice MLE estimators are preferred as they can allow for heterogeneity between firms both in a way that the CJS estimates are conditional on previous captures, and also that they allow for introducing covariates.

Appendix B : Tables

Table B.1: Example for capture history data

	1993	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**
Akzo	0	0	0	0	1	1	1	1	0	1	0	0
Alfa Acciai	0	0	0	0	0	0	0	1	0	0	0	0
Alken Maes	0	0	0	0	0	0	1	0	0	0	0	0
Allied Arthur Pierre	0	0	0	0	0	0	0	0	0	0	1	0
Alstom	0	0	0	0	0	0	0	0	0	0	0	1
ALZ NV	0	0	1	0	0	0	0	0	0	0	0	0
Amann	0	0	0	0	0	0	0	1	0	0	0	0
AP Moeller-Maersk	0	0	0	0	0	1	0	0	0	0	0	0
Aragonesas/Uralita	0	0	0	0	0	0	0	0	0	0	1	0
Areva	0	0	0	0	0	0	0	0	0	0	0	1
Arkema/Elf Aquitaine	0	0	0	0	0	0	0	0	0	0	1	0
Atochem	0	0	0	0	0	0	1	1	0	1	0	0

Table B.1 stores capture history data, indicating where an investigation successfully captured a cartelising firm (the same investigation may have captured involvement in more than one cartels). 1 implies capture and 0 denotes no capture.

Table B.2: AIC statistics

Model	**QAICc**	**ΔQAICc**	**QAICc weight**	**Model likelihood**
$\phi(./.)p(t)$	541.753	0.000	0.706	1.000
$\phi(././.)p(t)$	543.623	1.870	0.277	0.393
$\phi(../.)p(t)$	550.294	8.541	0.010	0.014
$\phi(./.)p(./t)$	552.730	10.971	0.003	0.004
$\phi(./.)p(5yrs)$	553.909	12.156	0.002	0.002
$\phi(./.)p(3yrs)$	559.481	17.728	0.000	0.000
$\phi(t/.)p(t)$	564.110	22.357	0.000	0.000
$\phi(./.)p(././t)$	566.149	24.396	0.000	0.000
$\phi(./.)p(../t)$	567.561	25.808	0.000	0.000
$\phi(.)p(t)$	572.576	25.808	0.000	0.000
$\phi(./.)p(./.)$	572.968	31.215	0.000	0.000
$\phi(./t)p(t)$	574.170	32.417	0.000	0.000
$\phi(t/.)p(./t)$	574.916	33.163	0.000	0.000
$\phi(././.)p(./.)$	574.990	33.240	0.000	0.000
$\phi(t/.)p(t/t)$	579.606	37.854	0.000	0.000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\phi(t/t)p(t)$	597.123	55.370	0.000	0.000

Table B.3: Parameter estimates for annual capture data

$\phi(./.)p(t)$	**Estimate**	**Std.err**	**95% CI**	
			Lower	**Upper**
1985	0.00000	0.00000	0.00000	0.00000
1986	0.00000	0.00000	0.00000	0.00021
1987	0.00000	0.00000	0.00000	0.00072
1988	0.00000	0.00000	0.00000	0.00000
1989	0.16196	0.11038	0.03777	0.48755
1990	0.00000	0.00000	0.00000	0.00000
1991	0.14836	0.08674	0.04339	0.40083
1992	0.43172	0.10623	0.24536	0.63966
1993	0.02226	0.02621	0.00215	0.19423
1994	0.40897	0.10369	0.22988	0.61598
1995	0.00000	0.00000	0.00000	0.00000
1996	0.00000	0.00000	0.00000	0.00000
1997	0.07589	0.05135	0.01918	0.25648
1998	0.21249	0.08225	0.09335	0.41421
1999	0.28764	0.09522	0.13971	0.50099
2000	0.04342	0.03653	0.00803	0.20286
2001	0.05834	0.04029	0.01450	0.20689
2002	0.20918	0.07695	0.09608	0.39697
2003	0.03182	0.02702	0.00586	0.15495
2004	0.01612	0.01926	0.00152	0.15036
2005	0.08547	0.04789	0.02735	0.23696

Table B.4: Robust design estimates for 3-year secondary sampling periods

Primary period	p	ψ^{AB}	ψ^{BA}	β	λ
1984-1990	0.052	-	-	0.149	0.158
	[0.020-0.128]	-	-	[0.138-0.160]	[0.100-0.240]
1991-1995	0.257	0.780	-	0.070	0.589
	[0.181-0.352]	[0.549-0.911]	-	[0.040-0.122]	[0.487-0.684]
1996-2000	0.270	0.684	0.222	0.224	0.408
	[0.105-0.538]	[0.302-0.915]	[0.025-0.763]	[0.209-0.239]	[0.150-0.730]
2001-2005	0.048	-	0.213	-	0.526
	[0.030-0.076]	-	[0.007-0.910]	-	[0.364-0.684]
2006-2009	0.020	0.480	-	0.073	0.234
	[0.003-0.140]	[0.008-0.990]	-	[0.014-0.301]	[0.106-0.438]

The parameters used in Tables B.5 and B.4 are as follows: p is the probability of detection at time t in secondary period i, ψ^{AB} and ψ^{BA} denote transition parameters (A - collude and B - compete) immediately before primary sampling periods (thus estimates for the first period do not exist). β denotes the probability that a firm in the collusion state in primary period t is a new arrival (within that primary period) to the study area for that state at secondary capture occasion j (for simplicity β is assumed to be constant over secondary periods) and λ denote the probability that an individual in the study area in the collusion state at secondary capture occasion j is still in that study area at capture occasion $j+1$ (again, for simplicity and the estimability of the model λ is assumed to be constant over secondary periods).

Table B.5: Robust design estimates for 3-year secondary sampling periods

Secondary period	p	ψ^{AB}	ψ^{BA}	β^{AB}	β^{BA}
1984-1990	0.038	-	-	0.155	0.158
	[0.015-0.093]	-	-	[0.147-164]	[0.100-0.240]
1991-1993	0.632	0.891	-	0.064	0.214
	[0.409-0.810]	[0.812-0.939]	-	[0.024-0.154]	[0.139-0.315]
1994-1996	0.206	-	-	0.261	-
	[0.106-0.363]	-	-	[0.162-0.391]	-
1997-1999	0.098	0.733	0.160	0.057	0.814
	[0.051-0.181]	[0.462-0.897]	[0.050-0.406]	[0.001-0.968]	[0.333-0.975]
2000-2002	0.086	0.571	0.039	0.317	-
	[0.041-0.170]	[0.265-0.823]	[0.001-0.627]	[0.222-0.429]	-
2003-2005	0.081	-	-	0.297	-
	[0.046-0.140]	-	-	[0.233-0.371]	-
2006-2009	0.009	-	-	0.069	0.229
	[0.004-0.023]	-	-	[0.011-0.320]	[0.105-0.430]