



A tip of the iceberg? The probability of catching cartels

Peter L Ormosi

CCP Working Paper 11-6

Abstract: Reliable estimates of crime detection rates could help design better sanctions and improve our understanding of the efficiency of law enforcement. For cartels, insufficient knowledge on the rate of discovery has hindered effective enforcement in the past. 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 and markets. It draws on capture-recapture methods, similar to those used to make inferences on various wildlife population characteristics in ecology. An application of this method provides evidence that less than one-fifth of all cartels are discovered.

21 November 2011

JEL Codes: C18, D43, K21, L41

Acknowledgements: I am grateful for the useful comments to Iwan Bos, Stephen Davies, Joseph Harrington, Morten Hviid, Bruce Lyons, Franco Mariuzzo, Matthew Olczak, Greg Shaffer, and the participants of ZEW 2011 June and ETH Zurich Seminar Series. A special thanks goes to Oindrila De and Andreas Stephan for kindly providing access to their data on EC cartels. The usual disclaimer applies. The support of the Economic and Social Research Council and the ESRC Centre for Competition Policy is also gratefully acknowledged.

Contact Details:

Norwich Business School and ESRC Centre for Competition Policy, University of East Anglia, NR4 7TJ, Norwich, United Kingdom, email:

p.ormosi@uea.ac.uk

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

Peter L. Ormosi[†]

Norwich Business School, University of East Anglia, UK

November 21, 2011

Abstract

Reliable estimates of crime detection rates could help design better sanctions and improve our understanding of the efficiency of law enforcement. For cartels, insufficient knowledge on the rate of discovery has hindered effective enforcement in the past. 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 and markets. It draws on capture-recapture methods, similar to those used to make inferences on various wildlife population characteristics in ecology. An application of this method provides evidence that less than one-fifth of all cartels are discovered.

JEL Classification codes: C18, D43, K21, L41

*I am grateful for the useful comments to Iwan Bos, Stephen Davies, Joseph Harrington, Morten Hviid, Bruce Lyons, Franco Mariuzzo, Matthew Olczak, Greg Shaffer, and the participants of ZEW 2011 June and ETH Zurich Seminar Series. A special thanks goes to Oindrila De and Andreas Stephan for kindly providing access to their data on EC cartels. The usual disclaimer applies. The support of the Economic and Social Research Council and the ESRC Centre for Competition Policy is also gratefully acknowledged.

[†]Norwich Business School and 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 loss - for example OECD member countries report estimates of harm of USD 55 billion world-wide for 16 cartels between 1996 and 2000¹ - and most of the estimations are based on a - potentially biased - selection of a small number of cases. The main obstacle to establishing the harm caused by illegal collusion is the lack of knowledge on what proportion of cartels are discovered. The focus of this paper is on providing a method for overcoming this obstacle. Knowing the probability of cartel detection would help in establishing the sufficiently deterrent level of fine and it could also help in the ex-post evaluation of anti-cartel policies. Due to the simplicity of the proposed method it 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.

The large proportion of undetected cartels hinders the design of an optimal anti-cartel policy. The number of detected cartels is a potentially biased indicator of the effectiveness of cartel enforcement. 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.² A change in the number of observed cases

¹OECD Reports, Fighting Hard Core Cartels: Harm, Effective Sanctions and Leniency Programmes, 2002, <http://www.oecd.org/dataoecd/41/44/1841891.pdf>

²For the purposes of this paper detection rate is used synonymously to the probability of cartel discovery.

can only be meaningfully interpreted if, at the same time, the change in detection rate is also known.³ Although previous cartel detection rate estimates do exist, these are one-off figures assuming stationarity and as such are not useful for deriving rates of change.

This paper offers a way to estimate cartel detection and survival probabilities that are time-dependent, thereby allowing the estimation of rates of change in the probability of cartel detection. The applied method, most commonly referred to as capture-recapture (CR) or mark and recapture 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 advocates the use of these advanced CR methods to obtain estimates for the detection rates of cartelising firms. Parallels between wildlife and economic agents will be drawn and it will be argued that the assumptions used in

³For example an increase in the number of cases from year t to year $t + 1$ paired with constant detection rate would mean reduced deterrence.

ecological studies can be similarly interpreted for economic agents. A key cornerstone of this method is that there are recaptures in the analysed samples, i.e. a large enough number of firms are repeatedly captured for involvement in cartels.⁴ 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.

Time-dependent estimates can also be used to derive the rates of change in detection probability, which could help in measuring the deterrent effect of anti-cartel policies. For their simplicity, 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. Examples could include corruption cases, drug offenses, tax evasion, or drink driving. Provided that there is a sufficient number of repeat offenders, CR methods could be used to derive time-dependent detection rate estimates, and potentially make inferences of the deterrence effect of public policies.

The paper commences with a review of the relevant literature on cartel detection. This is followed by an introduction to capture-recapture methods and a discussion of the assumptions required. Finally, an application is given on EC cartels detected between 1985 and 2005, and specific robustness issues and extensions are discussed.

⁴This is supported by empirical evidence. Connor (2010) for example reports a rapidly increasing rate of recidivist behaviour.

2 Other ways of estimating cartel detection rate

Only a limited number of previous works on cartels have managed to offer a solution for successfully accessing issues such as the probability of cartel detection and cartel deterrence.⁵ A common denominator of these works is that they set out to create and use a link between observed and unobserved factors to make inferences on the unobserved. Some of these papers grew out of the seminal article by Bryant and Eckard (1991), who use a simple statistical birth and death process to model cartel formation and breakdown, and then estimate model parameters using the duration of cartels as the observed factor. The model transition probabilities are given by the probability of cartel formation, the probability of cartel breakdown, and the probability of no cartels being formed or breaking down from period t to $t + 1$. A derived characteristic of birth and death processes is that the lifetime of cartels - assuming independence of observations - is exponentially distributed, and the central moments of this distribution depend on the cartel formation and breakdown probabilities of the underlying Markov process. Knowing the duration of detected cartels, these probabilities can be estimated. The intuition behind their paper is simple: if the distribution of cartel duration contains many short-lived cartels and few long-lived ones, then it is a sign of high detection rate and a small number of cartels alive, and vice versa. The authors suggest that in the period between 1961 and 1988 the annual

⁵Some of the earliest of these works use survey-based methods. These include Beckstein and Landis-Gabel (1982) and Feinberg (1985) with a detection rate estimate of less than 50% based on questionnaires among American and European antitrust lawyers respectively. Although survey estimates often provide a useful initial grasp of the analysed social phenomena, for the purposes of this paper more attention is given to those previous works that use robust, theory based, quantitative empirical methods.

probability of getting caught did not rise above 13-17% in any given year. Using the same method Golub, Detre, and Connor (2008) estimate US cartel detection rates for a later period and found similar figures. For the EU, Combe, Monnier, and Legal (2008) derive a cartel detection rate of 12.9-13.3% for the period between 1969-2007.

Miller (2009) models the world of cartels as a Markov chain, where three types of transition can link each state. For any industry, a probability is assigned to (a) non-colluding firms starting to collude, (b) colluding firms being discovered by the CA, and (c) colluding firms ceasing collusion.⁶ Using the properties of Markov processes, the expected number of colluding and competing industries are calculated both before and after introducing a leniency programme and a convergence path is established for the number of cases. This convergence path contains, as model parameters, the three transition probabilities of interest. Using the convergence path, the change in the number of detected cartels - as a result of changing model parameters - can be simulated. The simulated number of cases is then compared to the number of cases observed from enforcement records to find which simulated parameter change explains best what is observed in real life.⁷ Using data from US indictments and information reports between 1985 and 2005, Miller shows that the introduction of the 1993 leniency programme increased the rate of detection but did not affect deterrence.

⁶Although regulatory exposure is not the only reason why cartels may break down, Madhavan, Masson, and Lesser (1994) provides evidence that it may be the main one.

⁷The empirical estimation is fairly straightforward using a reduced form Poisson regression to examine (1) whether the number of cartel discoveries increases immediately after the introduction of leniency programmes, and (2) whether the number of discoveries decreases subsequently below the initial (pre-lenieny) level. To estimate (1) Miller examines the coefficient of a leniency dummy, whereas to analyse (2) the conditional mean of cartel discoveries is looked at.

Although a purely theoretical work, another paper could potentially be used for empirical estimates. Harrington and Chang (2009) (and an alteration of their model to allow for leniency by Chang and Harrington (2010)) suggest 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). 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. Hyytinen, Steen, and Toivanen (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 typically were not illegal.

A common limitation of the above works is that they rely on homogenous Markov chains, i.e. detection and deterrence rates are assumed to be constant. Secondly, the unit of analysis is industry, not firm, therefore the model does not allow for individual firms to join or leave the cartel.⁸ Thirdly, these methods do not account for firm or market based heterogeneity. Fourthly, individuals are assumed to be independent, but potential (and very likely) violations are not discussed. Finally, the above works typically only allow the analysis of simple comparative statics, such

⁸See Bos and Harrington (2010).

as the sign (and not the magnitude) of change in the parameters of interest as a result of a policy change, and do not offer the possibility of a dynamic analysis of the relevant parameters. The following discussion offers a method which overcomes some of these limitations, whilst at the same time relying on fewer and less restrictive assumptions.

3 The proposal

3.1 Capture-recapture methods

The paper proposes a capture-recapture (CR) method 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.⁹ 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 (where population can change for these reasons). This paper uses a variation of open population methods, the Cormack-Jolly-Seber model (CJS).¹⁰ CJS models are restricted in the sense that they rely solely on the recapture of previously marked

⁹For a detailed introduction to CR methods see: Amstrup, McDonald, and Manly (2005), Williams, Conroy, and Nichols (2002) or Burnham and Anderson (2002).

¹⁰Cormack (1964), and Lebreton, Burnham, Clobert, and Anderson (1992)

specimens, and use maximum likelihood methods to estimate survival and capture probabilities (but not the population size).¹¹ 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 cartels 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.¹²

3.2 A general CJS model

Let ϕ_t denote the probability of an individual surviving time $t = 1, 2, \dots$, 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 $t = 1, 2, \dots$ 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

¹¹Another 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).

¹²To understand the intuition behind CJS models, a simple explanation is given in Appendix A.

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 $t, t+1, \dots$. An individual that 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:

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

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.¹³ This is why the estimated parameters can only be interpreted for individuals that have been captured at some point.¹⁴ 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 cal-

¹³This can be thought of as the individual being inserted into the sample at this time.

¹⁴Due to the reversibility of CR models, the likelihood could be conditioned on any capture occasions.

culated 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, Pollock, and Norris (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):

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

Although this parametric form assumes that both capture and survival rates are time-dependent, it is possible to use a model where these are assumed to be constant over time. Goodness of fit tests are conducted below to find the best fitting parametric form.

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:¹⁵

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

¹⁵The observed capture history is therefore an observation of a multinomial distribution of all possible capture histories.

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.

3.3 Analysis of wildlife vs 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. Many aspects of analysing the undetected behaviour of economic agents are similar to the process of the trapping exercise conducted in wildlife. The issues that arise for economic agents can be found in wildlife and have therefore been addressed by rigorously designed biometric methods. Open population CR models for example allow individuals to be born and die during the period of analysis, and migrate over to another geographic area, where the researcher does not follow them.¹⁶ This is very similar to the situation one is expected to find for economic agents, i.e. firms can be wound up, or can decide to pull out of a market, and new firms can appear in the same market throughout the sampling process.

Another characteristic of wildlife is that individuals are not homogeneous. For example some animals are never caught because they live in an area where traps cannot be placed out. CJS methods are used in these situations to allow the distinction of two subpopulations, one that is never caught and one where individuals are capturable.¹⁷ The parallel can be seen for cartelising firms, where some firms are probably never caught, for example because their evasion skills are better than the

¹⁶Amstrup et al. (2005), Chapter 3

¹⁷Pollock, Nichols, Brownie, and Hines (1990), Chapter 4.

others’.

The capturability of animals may also vary within the sub-population of captured individuals. For example some animals are younger or larger than others, which has an impact on their survival and capture probabilities and is treated by accounting for this heterogeneity.¹⁸ Firms are similar in the sense that some may be more likely to be captured simply because of their size or the different duration of their illegal behaviour.

Another source of heterogeneity in nature is called trap-response in ecology: the capture probability of animals can increase (trap-happy) or decrease (trap-shy) as a result of previous capture.¹⁹ For firms, one could also expect that a discovery by the CA has an impact on the future capturability and survival of the same firm.

3.4 Interpretation of CR parameters for the analysis of 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 cartelising 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 cartelising firms that

¹⁸As discussed by Pledger, Pollock, and Norris (2003) for example.

¹⁹Crespin, Choquet, Lima, Merritt, and Pradel (2008).

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.²⁰ Therefore survival in this case means that the cartelising firm is still in the capturable subset (i.e. it still exist, it did not stop collusion forever, and it did not join the subpopulation of firms never captured).²¹

3.5 Assumptions

To more formally verify the applicability of CR methods to firm behaviour, this section examines whether the required assumptions are appropriate for the analysis of cartels.

3.5.1 Assumption A1 - Annual (discrete) sampling period

A1. *The CA engages in cartel enforcement (CR sampling) in discrete annual periods $t = 1, 2, \dots$. The population of cartelising firms does not change during sampling, but can change between sampling occasions.*

²⁰For example, for the capture history $CH_m = (1, 0, 1, 0)$ all we care about is that the cartel is still in the capturable subset in the third year, whether there was a cartel breakdown in the unobserved second year does not matter for this estimation. However, one of the extensions of this paper offers a method to include temporary collusion breakdown.

²¹Cartels - by their nature - must have started and may have even finished before capture. In this setting, survival only means survival of the possibility of being captured again. This could also be thought of as the survival of detectable cartel related evidence, which is generated when the cartel is formed and this evidence remains alive until capture (even if capture happens later than the breakdown of the cartel).

This assumption treats each year as one sampling occasion, and the parameter estimates are therefore annual capture and survival estimates. Classical CJS methods assume that samples are taken instantaneously.²² In practice this assumption is nearly always necessarily violated. Estimation bias can be avoided for example if it can be assured that within the sampling period (i.e. within each analysed year) there is no change in the analysed population.²³ The goodness of fit tests in Section 4.2.1 reveal that the resulting bias remains within tolerable limits. The robustness discussion also compares the model fit when using 6-month, 1, 2, and 3-year sampling periods.

Another way of testing the robustness of this assumption - which could be the subject of further research - would be by comparing the model used in this paper with continuous-time sampling estimates. In theory continuous models would be the best approximation of real life cartel enforcement. However, given the complexity of continuous models, their use would sacrifice the attractiveness of the simplicity of CJS methods.²⁴

3.5.2 Assumptions A2 and A3 - Homogeneity

A2. *The probability of any firm $m = 1, 2, \dots, n$ being captured by the CA at sampling occasion t is given by p_t (provided that it has been captured at least once and*

²²To illustrate the importance of this assumption imagine that cartel survival is analysed annually. The survival (i.e. to remain capturable in the future) of a cartelising firm to the next period is different for a firm that was captured in January from a firm that was captured in December.

²³This issue of long sampling times has been discussed by Williams, Conroy, and Nichols (2002).

²⁴Olsen (2006) for example proposed that CJS models may be used directly on continuous mark-recapture data, much in the same way as the Kaplan-Meier estimator is used for survival data, where semi-parametric hazard models are designed for the capture and mortality process.

that it has survived until t).

A3. *Any firm $m = 1, 2, \dots, n$ surviving sampling occasion t has equal probability ϕ_t of survival to $t + 1$.*

Besides homogeneity, Assumptions A2 and A3 imply two other things. One is that the unit of analysis is firms, and the other one is the time-dependence of the parameters, which relaxes the stationarity assumptions used in previous literature. A test for time-dependence, comparing models with time-dependent and constant parameters, will be conducted below.

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

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 cartelis-

ing firms having the same capture/survival probability (and not that marked and unmarked firms have equal capture and survival probabilities). This means that the modeler would not have to worry about the effect of factors such as increased regulatory vigilance aimed at industries that have been captured once, or customers becoming more alert in spotting cartels by firms that have been condemned before.

As far as temporary (or short-term) trap-response is concerned, there are many reasons why colluding firms could exhibit different behaviour following regulatory capture. Some of them probably adjust their behaviour 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 very low (i.e. many firms would disappear from the sample for either of these two reasons), and the "survived" firms would only include those that have no intention to stop repeating collusion, which group would be attributed with a high subsequent survival probability.

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 (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):

$$L = \prod_{m=1}^{2^K-1} \left\langle \sum_{d_m=l_m}^K \left\{ \left[\prod_{t=f_m}^{d_m-1} \left(\phi_1^{x_{mt}} \phi_2^{(1-x_{mt})} \right) \right] (1-\phi_1)^{x_{md_m}} (1-\phi_2)^{(1-x_{md_m})} \times \left(\prod_{t=f_m+1}^{d_m} p_t^{x_{mt}} (1-p_t)^{(1-x_{mt})} \right) \right\} \right\rangle \quad (3)$$

Heterogeneous firms and markets. Firm/market specific characteristics can also violate the homogeneity assumption. The most simple 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:

$$L = \prod_{m=1}^{2^K-1} \left\langle \sum_{c=1}^C \left[\sum_{d_m=l_m}^K \left\{ \left(\prod_{t=f_m}^{d_m-1} \phi_{ct} \right) (1-\phi_{cd_m}) \times \left(\prod_{t=f_m+1}^{d_m} p_{ct}^{x_{mt}} (1-p_{ct})^{(1-x_{mt})} \right) \right\} \right] \right\rangle \quad (4)$$

For example in ecology an obvious stratification could be based on the sex of the captured animals. For cartels it is not so easy to think about characteristics that so unambiguously subdivide the population of marked firms into a small number of sub-populations, but samples could be clustered along the following categories: industries, 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: $p_t, \phi_t = \exp(\alpha'_t \mathbf{X}_t) / [1 + \exp(\alpha'_t \mathbf{X}_t)]$, 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.²⁵

3.5.3 Assumption 4 - Independence

A4. *The capture and survival probabilities of individuals are independent of each other (independence is only needed for the marked subpopulation).*

Although the independence assumption is necessary for estimates in earlier works such as Bryant and Eckard (1991), they do not address the potential violation of this assumption, which may result in biased estimates. 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.²⁶ For cartels it is unlikely in many instances for two cartels to be independent from each other.²⁷ Moreover, the same regulatory action (e.g. a dawn raid) may discover more than one cartels. The goodness of fit tests in Section 4.2 look at the amount of overdispersion caused by the violation of this assumption, and a treatment is also

²⁵In a time-dependent model with l covariates affecting both survival and capture rates, there would be an additional $2K \times l$ coefficients to estimate.

²⁶See Anderson, Burnham, and White (1994) for an analysis of overdispersion caused by lack of independence.

²⁷A wildlife analogy would be species whose members exist in schools or flocks.

offered.

3.5.4 Further assumptions

A5. The whole geographical area of study is sampled with equal intensity. If new areas are added to the sampling area, they have randomly chosen characteristics of the initial study area.

The relevance of this assumption is specific to the empirical part of this paper. It accounts for the fact that economies may be expanding (for example in Europe, any analysis would need to account for the fact that countries may have joined the EU during a given time period).

A6. Marked individuals do not lose their marks.

Although this assumption is typically more relevant to ecological studies where animals are physically tagged, one issue may arise in relation to cartelising firms. Firms may 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.

4 Application to European cartels

4.1 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.²⁸

The cartels included in the sample are the ones 'captured' by the European Commission. For the purposes of data collection the case reports from the European Commission's cartel database are used. The first capture in the sample was in 1984, and the last one in 2005, in which time period 110 cases with 492 cartelising 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 685 offences, including 94 recidivist firms, and 194 recidivist offences.

The data is organised in a $n \times K$ matrix \mathbf{X} , (n is the number of individual cartelising 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 was colluding, otherwise $x_{mt} = 0$.²⁹ 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. Appendix B contains an example of how capture history data is organised for the ML estimation.

²⁸Plea bargaining was not an issue for EC cases in the time of this analysis (pre-2005) but will be for future cases.

²⁹Firms are identified if they have at least once been captured by the Commission.

4.2 Model fit

4.2.1 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, Burnham, Clobert, and Anderson (1992), and the deviance statistic, which is χ^2 distributed, is used in the GOF test as: $\hat{c} = \frac{\chi^2}{df}$. For the saturated model $\hat{c} = 1$, therefore any 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.³⁰

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 $\hat{c}_{bootstrap} = \frac{\text{model deviance}}{\text{mean of simulated deviances}}$.

\hat{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 3 (where capture can have an immediate effect on survival - trap-response). Table 1 shows the measures of overdispersion for these two general models. The lower \hat{c} -statistics imply that the model, which allows for an im-

³⁰Anderson, Burnham, and White (1994).

mediate 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 1: GOF of the most general models

	\hat{c}	$\hat{c}_{bootstrap1}$
Equation 2	2.20181	1.796297
Equation 3	1.39898	1.378984

Lebreton, Burnham, Clobert, and Anderson (1992) suggest that $\hat{c} < 3$ indicates a tolerable level of overdispersion. Although the statistics in Table 1 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).³¹

4.2.2 Model choice

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

³¹White (2002) shows that in most circumstances \hat{c} performs better than $\hat{c}_{bootstrap}$, therefore the likelihood terms provided below are adjusted by \hat{c} .

The notations used below follow a very simple logic: $\phi(t)$ and $p(t)$ denote time dependent survival and capture parameters, and $\phi(\cdot)$ and $p(\cdot)$ refer to models with constant survival and capture probabilities respectively. $\phi(t/t)$ denotes a model 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. Finally, the model denoted as $\phi(\cdot/t)p(t)$ 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.³² The examples given in Equations (2) and (3) are denoted as $\phi(t)p(t)$ and $\phi(\cdot/\cdot)p(t)$ respectively. 13 different parametrisations of the likelihood function in Equation 2 are estimated, and the most efficient one is chosen using Akaike's Information Criterion.³³ The test statistics are presented in Table 2, where AIC_c is the corrected AIC , ΔAIC is the difference in comparison to the model with the lowest AIC . Q implies that the likelihood terms were adjusted for overdispersion, and quasi-likelihood adjusted AIC is reported.

The $QAIC_c$ weight in Table 2 shows that given the dataset, model $\phi(\cdot/\cdot)p(t)$ is around 2.7 times more likely than the second most efficient model.³⁴ The model

³²The software Mark was used for the estimations to follow: <http://warnercnr.colostate.edu/~gwhite/mark/mark.htm>

³³The AIC is given by: $AIC = -2\ln(\mathcal{L}) + 2K$, where \mathcal{L} is the model likelihood, and K 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: $AIC_C = -2\log(\mathcal{L}) + 2K [n / (n - K - 1)]$.

³⁴Burnham 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

likelihood column captures the same information, the relative likelihood of model $\phi(. /. .)p(t)$ to model $\phi(. /.)p(t)$ is 0.3692. 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 rate of cartelising firms in the year immediately following capture is different from survival rate in subsequent years.

Table 2: AIC statistics

Model	QAICc	ΔQAICc	QAICc weight	Model likelihood
$\phi(. /.)p(t)$	556.5530	0.0000	0.72945	1.00000
$\phi(. /. /.)p(t)$	558.5458	1.9928	0.26932	0.3692
$\phi(t /.)p(t)$	569.5064	12.9534	0.00112	0.0015
$\phi(.)p(t)$	575.3504	18.7974	0.00006	0.0001
$\phi(. /.)p(t/t)$	576.5875	20.0345	0.00003	0.0000
$\phi(. /t)p(t)$	579.6344	23.0814	0.00001	0.0000
$\phi(. /.)p(. /t)$	580.1880	23.6350	0.00001	0.0000
$\phi(t)p(t)$	588.7011	32.1481	0.00000	0.0000
$\phi(t/t)p(t)$	589.8691	41.6377	0.00000	0.0000
$\phi(. /.)p(t/.)$	598.1907	41.6377	0.00000	0.0000
$\phi(. /.)p(. /.)$	604.3377	47.7847	0.00000	0.0000
$\phi(t/t)p(t/t)$	605.1795	48.6265	0.00000	0.0000
$\phi(.)p(.)$	619.6519	63.0989	0.00000	0.0000
$\phi(t)p(.)$	623.0551	63.0989	0.00000	0.0000

for difference between the two models, and $\Delta AIC_c > 7$ supports strong evidence that the two models are different.

4.3 Parameter estimates and discussion

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

4.3.1 The detection rate of cartelising firms

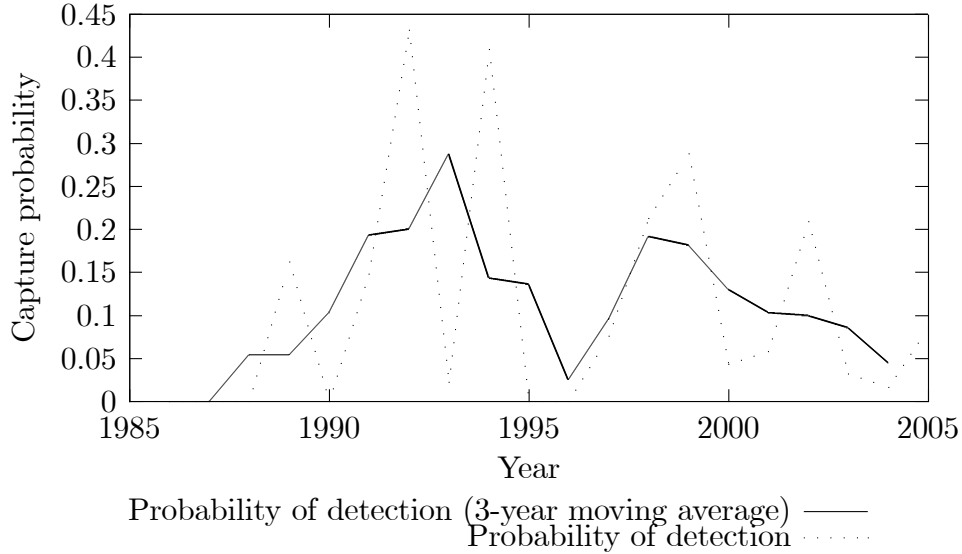
Figure 1 plots the maximum likelihood estimates for $p(t)$ using the $\phi(./.)p(t)$ model as shown in Equation (3).³⁵ The estimates are also plotted as 3-year (mid-point) moving averages.

Figure 1 shows that the detection rate of cartelising firms stays under 20% most of the time (with the exception of three years), frequently dropping under 10%. The moving average line on Figure 1 reveals two peaks in cartel detection rate. One in 1993, and the other one in 1998. One possible reason for the first increase in cartel detection probability could be an effect of the introduction of the US leniency programme (1993). The second increase may be related to the introduction of EC leniency programmes (1996) and the establishment of the Cartel Unit within the Commission (1998). The subsequent reform of leniency programmes in 2002 seemingly had no effect on the probability of detecting cartels.

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

³⁵The table of estimations can be found in Table B.2 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 (1985-2005)



way as in Bryant and Eckard (1991). The mean (over-time) estimate is similar to the constant detection rate estimated by Combe, Monnier, and Legal (2008). The comparative advantage of the CR-method is that the change of capture rate over time can now be estimated, which provides a very convenient tool for a more profound policy analysis. Moreover, as is suggested under possible extensions, this could also be used for making inferences on the deterrent effect of cartel enforcement.

4.3.2 The survival of cartelising firms

The estimated survival parameters (and the 95% CI) are: $\phi_1 = 0.29108$ [0.19431; 0.41432] for the year immediately following capture, and $\phi_2 = 0.88093$ [0.79291; 0.93463] otherwise. It is not clear whether the subpopulation of uncaptured col-

cluding 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.³⁶

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.³⁷ If colluding 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.³⁸ The welfare consequences of 'death' because of avoiding future detection 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

³⁶According 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.

³⁷As explained earlier, these death rates do not include the temporary breakdown of cartels.

³⁸'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.

or because it colludes too well.³⁹ Intuitively however it could be said that 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.

5 Further robustness and sensitivity discussion

5.1 Annual sampling occasions

Table 3 reveals that the choice of 1-year sampling periods provides a good fit relative to other choices (as explained above, out of the two measures \hat{c} typically offers a better measure). The parameter estimates are also presented for the best fitting model $\phi(./.)p(t)$, for the 6-month, 1-year, 2-year, 3-year sampling occasions, and are reported in Table B.3 in the Appendix.

Table 3: GOF for different sampling lengths

	1 year		2 years		3 years		6 months	
	\hat{c}	\hat{c}_{bs}	\hat{c}	\hat{c}_{bs}	\hat{c}	\hat{c}_{bs}	\hat{c}	\hat{c}_{bs}
$\phi(t)p(t)$	2.202	1.796	3.292	2.244	1.912	2.187	2.194	1.692
$\phi(t/t)p(t)$	1.399	1.379	1.980	1.447	1.217	1.782	1.627	1.322

The GOF tests in Table 3 and the parameter estimates in Table B.3 show that the

³⁹Connor (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.

1-year long sampling periods offer a comparatively good fit. 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 3-year model would mean losing some of the captures (3 captures in a 3-year model would only be recorded as 1 single capture). Following this logic, the 1-year model is superior to the 2, and 3-year models in this respect. 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. Also, in the 6-month model we have - at least - twice as many parameters, and therefore much smaller degrees of freedom, which may also cause a problem with small datasets.

5.2 Firm as observation

The paper uses firms as a unit of analysis to allow for the possibility of individual firms joining or leaving cartels. Although this offers more flexibility than using industries, it is still not a fully realistic depiction of how cartels work. Repeat offences by large diversified firms could also be a result of decisions brought at the level of individual divisions, therefore the researcher would first have to choose whether firms or divisions should be used as observation. This ambiguity is down to the fact that for CR methods to work, firms are used analogously to animals, even though firms are not a single undivided entity the same way as individual animals are. When dealing with diversified firms, therefore the right unit of observation may be the individual divisions of firms. Nevertheless, using divisions as observation would

create serious obstacles to empirical work. Most obviously, division-level data would be more difficult to acquire. Secondly, one would need reliable information on firms' corporate culture to be able to distinguish between cases where the behaviour of divisions is determined by a high-level executive, in which case division-level analysis is not the most appropriate one. Finally, one could still ask the question, are divisions a good enough observation, why not go further and focus on the managers behind the cartels?

If a firm-level analysis is conducted when division should have been used as the correct observation, then there is a positive bias in the estimated capture probabilities. This is easy to see as the probability of capturing any individual division is always smaller than (or equal to) the probability of capturing firms that embodies the given division.⁴⁰ 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.

5.3 Heterogeneity

One aspect of heterogeneity (trap-response) has been controlled for above and test results suggest a very small amount of overdispersion in the model used. 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 (4). One

⁴⁰It will be equal when there is only one colluding division in the firm.

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.

6 Possible extensions

6.1 Estimating the deterrent effect of enforcement

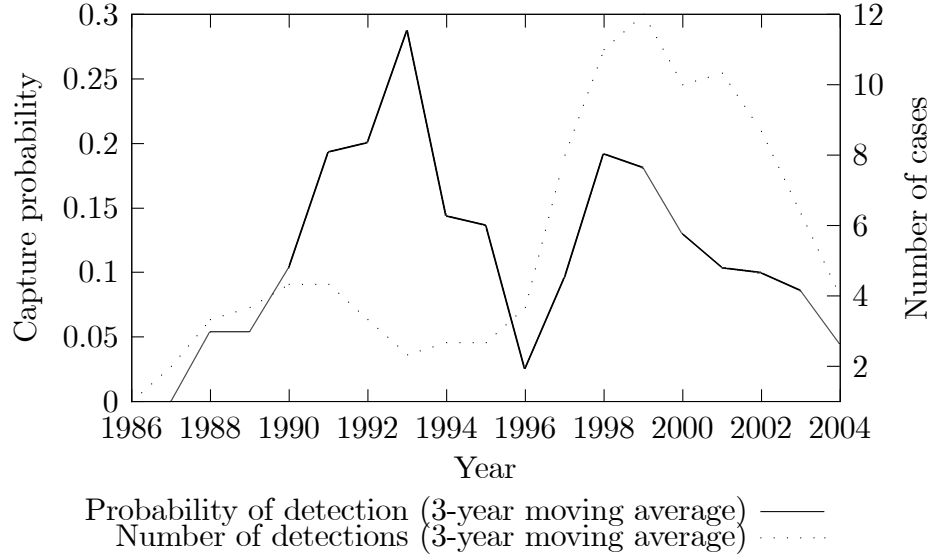
Several authors highlight the difficulty of gauging the effect of anti-cartel enforcement, which is caused primarily by the lack of information on deterred cartels.⁴¹ An evaluation of enforcement activities that focuses on the number of cases and not on deterrence may lead to enforcement agencies prosecuting fewer cases than is socially optimal in order to reduce deterrence and to raise the cartel rate.⁴² This extension proposes a method to make inferences on the deterrent effect of law enforcement. Figure 2 projects estimated detection rates over the number of cases. The purpose of this exercise is to compare the change in these two variables in the expectation that it reveals some information about the deterrent effect of cartel enforcement.

To understand the relationship between the number of cases and detection rate, consider Figure 3, which offers a classification of cartels to fully depict how cases

⁴¹See for example: Werden (2008) and Bergman (2008).

⁴²Harrington (2011).

Figure 2: Detection rate vs number of cases (1985-2005)



are distributed over the total population of privately profitable cartelising firms, including those that were deterred as a result of anti-cartel laws and enforcement. According to this classification, cases can be broken down into deterred and undeterred ones. Within the undeterred subset, there are cases that are discovered by the CA (the tip of the iceberg) and cases that remain hidden.

Figure 3: The universe of cartels

Deterred ω	Undeterred $(1 - \omega)$	
	Undetected $(1 - p)$	Detected p

Figure 3 denotes deterrence and detection rates as ω and p respectively. Assuming

that Figure 3 is a realistic representation of the universe of cartelising firms, denote the number of detected individuals at time t as $C(t)$, which is given by:

$$C(t) = [1 - \omega(t)]p(t)N \quad (5)$$

Where N is the total number of cartelising firms (deterred and undeterred).⁴³

As shown on Figure 2, in the pre-1996 period the number of detected cartels (C) increases slightly then remains roughly constant. The detection rate p on the other hand shows a steady increase. This is only possible if at the same time the deterrence rate (ω) increases. Similarly, the post leniency period (1997-) shows that C and p move together. Intuitively this may suggest that the deterrence rate remains approximately constant. Finally, the 1996 introduction of leniency programme was accompanied by a sudden increase in both C and p . We cannot decide without a formal analysis what this means for ω . For this purpose the differentials ΔC and Δp can be worked out from the estimates in Table B.2 in the Appendix. Using these differentials $\Delta\omega$ can be derived from Equation 5 as:

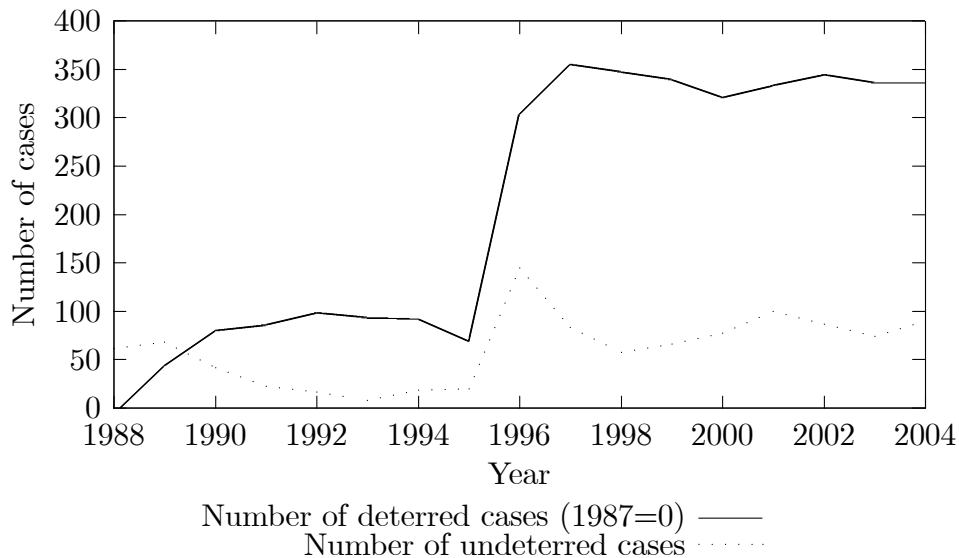
$$N\Delta\omega = \frac{C_t\Delta p - \Delta C p_t}{p_t^2}$$

Note that N is moved to the LHS as it is not estimated, and $\Delta\omega = \omega_t - \omega_{t-1}$. Figure 4 plots $N\omega_t$ (the number of deterred cases relative to $N\omega_{1987} = 0$), and $N(1 - \omega_t)$ (the number of undeterred cases). As p_t is an upper-bound estimate,

⁴³To be more precise N is the number of cartels that are in the 'caturable' subpopulation.

$N\Delta\omega$ and $N\omega_t$ are both lower bounds (assuming that the ratio of the capturable and uncapturable subpopulations does not change too much between two subsequent years).

Figure 4: The deterrent effect of anti-cartel enforcement (1988-2004)



The plotted curves show that the deterrent effect of EC cartel enforcement increased slightly between the end of the 80s and 1992, then stagnated for a few years before a large improvement around 1996 (the introduction of the European corporate leniency programme). From 1997 deterrence has been at a roughly constant level. It clearly stands out that the 1996 introduction of the leniency programme had a strong impact on cartel deterrence. In the context of Miller’s (2009) simulation an increased detection and deterrence rate should display a case number pattern that increases around the introduction of leniency programmes but eventually drops un-

der the pre-leniency level. This cannot be confirmed or rejected from the evidence provided above, as the analysed time period is not long enough to see what level the number of cases converges following the policy change (it is decreasing, but whether it drops to a low enough level remains to be seen).⁴⁴

6.2 Allowing for cartel breakdown and re-forming

Levenstein and Suslow (2006), Levenstein and Suslow (2009) and Harrington and Chang (2009) emphasise the importance of thinking about cartels as a multistate process, where cartels necessarily break down and often re-form following some time. Harrington and Chang (2009) make the assumption that the probability of detection is the same if firms are not currently colluding but did collude in the past. Miller (2009) also allows for industries switching between competition and collusion.

To allow cartels to break down and then re-form again, assume that cartelising firms can be in one of two stages at all times, (A) collusion, (B) competition. The capture history $(A, 0, 0, B)$, recorded for the years 2001-2004, would in this case mean that the firm was captured whilst actively being involved in a cartel in 2001. Between 2002 and 2003 the firm was not captured, so we do not know whether it was colluding or competing. When the same firm is captured in 2004, it is in a state of competition. The fact that the firm is still captured means that the firm was involved in a cartel sometime during 2002-2003, but it left the cartel and was in the competition state (B) at the time of the capture. This would require the capture of

⁴⁴The same technique could be used to assess the impact of other policy changes on cartel formation, and complement research such as Krepps (1997) on the impact of the US National Industrial Recovery Act, or Levenstein and Suslow (2007) on the U.S. Export Trading Company Act.

cartels that have already broken down before detection. The EC dataset provides a number of examples for this.

It is very easy to see that periods without observation rapidly increase the number of estimable parameters. Consider the above capture history but with only one period of uncertainty, $CH : (A, 0, B)$. The probability of observing this capture history (following the diagram) would be:

$$L(A, 0, B) = \phi_1^A [\psi_1^{AB} (1 - p_2^B) \phi_2^B (1 - \psi_2^{AB}) + (1 - \psi_1^{AB}) (1 - p_2^A) \phi_2^A \psi_2^{AB}] p_3^B$$

Where ψ_t^{sr} is a movement parameter representing the probability of movement from state s to r , $s, r = \{A, B\}$, and the survival and capture parameters are also state-specific. It is also assumed that $\psi_t^{AB} + \psi_u^{AA} = 1$, if $t = u$. The movement probabilities are net, i.e. if a firm moves among states between observations, only the initial and final states are of interest. Similarly to the model presented above, this model is also conditioned on the first capture (i.e. firms only enter the sample after being captured) therefore only information after the first capture can be obtained from the model.

Now introduce another time period where the cartelising firm was not captured, with a capture history of $CH : (A, 0, 0, B)$. The likelihood of observing this pattern

becomes much more complicated:⁴⁵

$$L(A, 0, 0, B) = \phi_1^A \left\{ \begin{array}{l} \psi_1^{AB}(1 - p_2^B)\phi_2^B \left[\begin{array}{l} \psi_2^{BA}(1 - p_3^A)\phi_3^A\psi_3^{AB} + \\ +(1 - \psi_2^{BA})(1 - p_3^B)\phi_3^B(1 - \psi_3^{AB}) \end{array} \right] + \\ +(1 - \psi_1^{AB})(1 - p_2^A)\phi_2^A \left[\begin{array}{l} \psi_2^{AB}(1 - p_3^B)\phi_3^B(1 - \psi_3^{BA}) + \\ +(1 - \psi_2^{AB})(1 - p_3^A)\phi_3^A\psi_3^{AB} \end{array} \right] \end{array} \right\} p_4^B$$

The fact that there are two periods where the authority makes no observation results in 4 possible paths.⁴⁶ Although the estimation of these parameters is possible, the large number of parameters would require a much larger dataset.⁴⁷ If, however, sufficient data is available, these types of models can be estimated using the MARK software package used above.⁴⁸

7 Concluding remarks

Measuring the impact of law enforcement is a difficult task especially in areas with a high proportion of undetected behaviour, which is thought to be the case for cartels. Building on the similarity between the wildlife trapping of animals and the discovery

⁴⁵This is fundamentally a CH depiction of a hidden Markov process.

⁴⁶As a sidenote, the capture history $(A, 0, 0, B)$ is only possible, if the cartel was operational either in time 2 or 3 (i.e. the second capture happens because the cartel has been operational since the first capture). This means that the path: $\phi_1^A\psi_1^{AB}(1 - p_2^B)\phi_2^B(1 - \psi_2^{BA})(1 - p_3^B)\phi_3^B(1 - \psi_3^{AB})p_4^B$ is not possible, as this would mean no collusion at all after the first capture, i.e. capture in period 4 would not have been possible.

⁴⁷Counting with k sample times and S states there are a total of $(k - 1)(S^2 + 2S)$ parameters to be estimated.

⁴⁸A possible way to simplify this, would be to use a forward-backward algorithm (frequently applied in computer sciences) together with hidden Markov chains. This, however, is left to future work.

of illegal cartels, the paper advocates for using capture-recapture models, - a method widely used for population studies in ecology - to study cartel detection, survival, and deterrence 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*, therefore it does not require an underlying theoretical model that typically comes with numerous assumptions on how cartels are formed and broken down. (2) *Minimal data requirements*: in its simplest form CR methods only need data on capture (encounter) times. (3) *Time-dependent parameters*: It allows time-dependent parameter estimates, e.g. an estimate of detection rate over time, which improves the possibilities of analysing the effect of policy changes. The time-dependent estimates can also be used to derive the rates of change of detection probability, which could be used to make inferences on the deterrent effect of anti-cartel policies. (4) *Heterogeneity*: the method offers the possibility to control for firm and market specific characteristics, and therefore identify specific circumstances that facilitate or hinder cartel detection and survival. (5) *Analysis is firm-based*, which allows the possibility of individual firms quitting or rejoining the cartel and does not assume that the whole industry is either collectively in a cartel or in competition. (6) *No need to identify all states of life*: CR methods do not rely on transitions between states in the same way as cartel formation-breakdown models do, therefore there is no need to identify all states of life, only to distinguish between unlawful and lawful conduct. A CR analysis estimates the parameters of a population of choice, provided that this population is clearly discernible from other populations. In the case of cartels

this means that the researcher is exploring the parameters of the population that the CA intends to capture (i.e. illegal cartels). No assumption is needed on what other possible states of life may exist. It is the policymaker that draws the line, beyond which a conduct is illegal, and CR estimates describe the characteristics of individuals that fall beyond that line.

The application to EU cartels is used to illustrate CR methods at work. Estimates show that less than a fifth of cartels have been detected in the EU between 1985 and 2005. Cartel survival - interpreted as future capturability - is around 30% in the year following capture (i.e. in this crucial period around 70% cartelising 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 it becomes more wary of the EC's investigations), but for those firms that remain capturable later on, apparent survival increases to almost 90%. Finally, deterrence seems to have been largely boosted by the introduction of the 1996 leniency programme.

Given its simplicity, the proposed method could develop to be an important tool for cartel related policy analysis. Moreover, the relevance of this paper goes beyond the scope of analysing cartel enforcement. CR methods could be more generally applied to other areas of law enforcement that are characterised by high proportions of undetected illegal behaviour. Examples could include corruption cases, drug offenses, tax evasion, or drink driving. Provided that there is a sufficient number of repeat offenders, CR methods could be used to derive time-dependent detection rate estimates, and potentially make inferences of the deterrence effect of public policies.

References

- AMSTRUP, S., T. McDONALD, AND B. MANLY (2005): *Handbook of Capture-Recapture Analysis*. Princeton University Press.
- ANDERSON, D. R., K. P. BURNHAM, AND G. C. WHITE (1994): “AIC Model Selection in Overdispersed Capture-Recapture Data,” *Ecology*, 75, 1780–1793.
- BECKSTEIN, A., AND H. LANDIS-GABEL (1982): “Antitrust Compliance: Results of a Survey of Legal Opinion,” *Antitrust Law Journal*, 52, 459–468.
- BERGMAN, M. A. (2008): “Quis Custodiet Ipsos Custodes? or Measuring and Evaluating the Effectiveness of Competition Enforcement,” *De Economist*, 156(4), 387–409.
- BOS, I., AND J. E. HARRINGTON (2010): “Endogenous Cartel Formation with Heterogeneous Firms,” *RAND Journal of Economics*, 41(1), 92–117.
- BRYANT, P., AND W. ECKARD (1991): “Price Fixing: The Probability of Getting Caught,” *Review of Economics and Statistics*, 73, 531–540.
- BURNHAM, K. P., AND D. R. ANDERSON (2002): *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, 2nd edn.
- CHANG, M.-H., AND J. E. HARRINGTON (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., C. MONNIER, AND R. LEGAL (2008): “Cartels: The Probability of Detection in the European Union,” *BEER paper*, No. 12.
- CONNOR, J. M. (2010): “Recidivism Revealed: Private International Cartels 1990-2009,” *Competition Policy International*, 6(2).
- COOCH, E., AND G. WHITE (2010): *Program Mark: A Gentle Introduction*. Publisher, 9th edn.
- CORMACK, R. M. (1964): “Estimates of Survival from the Sighting of Marked Animals,” *Biometrika*, 51, 429–438.
- CRISPIN, L., R. CHOQUET, M. LIMA, J. MERRITT, AND R. PRADEL (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.
- FEINBERG, R. (1985): “The Enforcement and Effects of European Antitrust Policy: A Survey of Legal Opinion,” *Journal of Common Market Studies*, 23, 373–384.
- GOLUB, A., J. DETRE, AND J. M. CONNOR (2008): “The Profitability of Price Fixing: Have Stronger Antitrust Sanctions Deterred?,” *available at: <http://ssrn.com/abstract=1188515>*.
- HARRINGTON, J. E. (2011): “When is an Antitrust Authority Not Aggressive Enough in Fighting Cartels?,” *International Journal of Economic Theory*, 7(1), 39–50.

- HARRINGTON, J. E., AND M.-H. CHANG (2009): “Modelling the Birth and Death of Cartels with an Application to Evaluating Antitrust Policy,” *Journal of the European Economic Association*, 7(6), 1400–1435.
- HURVICH, C. M., AND C.-L. TSAI (1989): “Regression and Time Series Model Selection in Small Samples,” *Biometrika*, 76, 297–307.
- HYYTINEN, A., F. STEEN, AND O. TOIVANEN (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.
- KREPPS, M. B. (1997): “Another Look at the Impact of the National Industrial Recovery Act on Cartel Formation and Maintenance Costs,” *Review of Economics and Statistics*, 79(1), 151–154.
- LEBRETON, J.-D., K. P. BURNHAM, J. CLOBERT, AND D. R. ANDERSON (1992): “Modelling Survival and Testing Biological Hypotheses Using Marked Animals: A Unified Approach with Case Studies,” *Ecological Monographs*, 62, 67–118.
- LEVENSTEIN, M. C., AND V. Y. SUSLOW (2006): “What Determines Cartel Success?,” *Journal of Economic Literature*, 44(1), 43–95.
- (2007): “The Economic Impact of the U.S. Export Trading Company Act,” *Antitrust Law Journal*, 27(2), 343–386.
- (2009): “Breaking Up is Hard to Do: Determinants of Cartel Duration,” Available at SSRN: <http://ssrn.com/abstract=1471393>.

- MADHAVAN, A. N., R. T. MASSON, AND W. H. LESSER (1994): “Cooperation for Monopolization? An Empirical Analysis of Cartelization,” *Review of Economics and Statistics*, 1(161-176), 76.
- MILLER, N. H. (2009): “Strategic Leniency and Cartel Enforcement,” *American Economic Review*, 99(3), 750–768.
- OLSEN, I. C. (2006): “Assessing the Effects of Lengthy Sampling Periods on Mark-Recapture Methodology,” *Doktoravhandling for graden doktor ingeniør*.
- PLEDGER, S., K. H. POLLOCK, AND J. L. NORRIS (2003): “Open Capture-Recapture Models with Heterogeneity: I. Cormack-Jolly-Seber Model,” *Biometrics*, 59(4), 786–794.
- POLLOCK, K. H., J. D. NICHOLS, C. BROWNIE, AND J. E. HINES (1990): “Statistical Inference for Capture-Recapture Experiments,” *Wildlife Society Monographs*, 107.
- SEBER, G. (1965): “A Note on Multiple-Recapture Census,” *Biometrika*, 52, 249–259.
- (1982): *The Estimation of Animal Abundance and Related Parameters*. Macmillan, New York, 2nd edn.
- WERDEN, G. J. (2008): “Assessing the Effects of Antitrust Enforcement in the United States,” *De Economist*, 156(4), 433–451.
- WHITE, G. C. (2002): “The Use of Auxiliary Variables in Capture-Recapture Modeling: An Overview,” *Journal of Applied Statistics*, 29(1-4), 103–106.

WILLIAMS, B. K., M. J. CONROY, AND J. D. NICHOLS (2002): *Analysis and Management of Animal Populations*. Academic Press.

Appendix A : Open population CR models - an intuitive explanation

Although both Amstrup, McDonald, and Manly (2005) and Williams, Conroy, and Nichols (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: 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.3: Parameter estimates with different sampling lengths

	6months		1year		2years		3years	
	Estimate	Std.err	Estimate	Std.err	Estimate	Std.err	Estimate	Std.err
ϕ_1	0.338	(0.059)	0.291	(0.056)	0.270	(0.045)	0.375	(0.097)
ϕ_2	0.945	(0.015)	0.881	(0.035)	0.831	(0.059)	0.660	(0.079)
1985	0.000	(0.000)	0.000	(0.000)				
	0.000	(0.000)						
1986	0.000	(0.000)	0.000	(0.000)				
	0.000	(0.000)						
1987	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)		
	0.000	(0.000)						
1988	0.000	(0.000)	0.000	(0.000)				
	0.000	(0.000)						
1989	0.000	(0.000)	0.162	(0.110)	0.182	(0.107)	0.000	(0.000)
	0.116	(0.068)						
1990	0.000	(0.000)	0.000	(0.000)				
	0.000	(0.000)						
1991	0.000	(0.000)	0.148	(0.087)	0.148	(0.074)		
	0.099	(0.049)						
1992	0.185	(0.061)	0.432	(0.106)			0.104	(0.053)
	0.206	(0.064)						
1993	0.000	(0.000)	0.022	(0.026)	0.447	(0.091)		
	0.021	(0.021)						
1994	0.152	(0.056)	0.409	(0.104)				
	0.222	(0.068)						
1995	0.000	(0.000)	0.000	(0.000)	0.405	(0.087)	0.282	(0.087)
	0.000	(0.000)						
1996	0.000	(0.000)	0.000	(0.000)				
	0.000	(0.000)						
1997	0.022	(0.022)	0.076	(0.051)	0.089	(0.044)		
	0.064	(0.037)						
1998	0.020	(0.02)	0.212	(0.082)			0.206	(0.074)
	0.155	(0.055)						
1999	0.075	(0.038)	0.288	(0.095)	0.341	(0.087)		
	0.167	(0.057)						
2000	0.000	(0.000)	0.043	(0.037)				
	0.035	(0.025)						
2001	0.049	(0.029)	0.058	(0.04)	0.058	(0.034)	0.217	(0.076)
	0.000	(0.000)						
2002	0.073	(0.034)	0.209	(0.077)				
	0.099	(0.039)						
2003	0.013	(0.014)	0.032	(0.027)	0.188	(0.061)		
	0.014	(0.014)						
2004	0.014	(0.014)	0.016	(0.019)			0.169	(0.062)
	0.000	(0.000)						
2005	0.044	(0.026)	0.085	(0.048)	0.089	(0.039)	0.066	(0.034)
	0.030	(0.022)						