Known unknowns: How much financial misconduct is detected and deterred?

Ashton, John; Burnett, Tim ; Diaz Rainey, Ivan ; Ormosi, Peter

Journal of International Financial Markets, Institutions and Money

DOI:
10.1016/j.intfin.2021.101389

Published: 01/09/2021

Peer reviewed version

Known unknowns: How much financial misconduct is detected and deterred?

John Ashton
Bangor University

Tim Burnett
Aston University

Ivan Diaz-Rainey
University of Otago

Peter Ormosi
University of East Anglia

October 6, 2020

Abstract

Have financial businesses changed their behaviour in the aftermath of global financial crisis? We address this question by introducing a new and more parsimonious method to quantify the level of financial misconduct and apply this to financial offences between 2004 and 2016. This exercise allows us to investigate whether Capture-Recapture methods can be deployed to handle problems of partial observability and how they compare to previous methods set out to achieve the same goal. In our two stage approach, first, we estimate the rate at which offending businesses are detected, then we look at how the number of detected offenders changed after 2010, and use these two layers of information to make inferences on the deterrent effect of financial regulation. Our results offer evidence that a drop in the number of detected offences post-global financial crisis was driven largely by improved deterrence.

Keywords: misconduct behaviour, misconduct risk, regulatory punishments, partial observability, Capture-Recapture, deterrence

We would like to thank Yang Wang and participants from seminars at Cardiff University (2017), University of East Anglia (2018), University of Liverpool (2018) and SKEMA Business School (2019), and a presentation at EARIE (2017) for their helpful comments. Faults remain our own.

1 Introduction

Since the global financial crisis, there has been greater awareness of the risks posed by the misconduct of financial institutions and their employees. At the same time, there has been heightened regulatory oversight of the behaviour of financial firms, accompanied by an increase in the severity of the punishments imposed. These developments have spawned a rich post-crisis literature examining organisational cultures of offending (e.g. Burdon and Sorour (2018), Graham et al. (2017), Parsons et al. (2018)), the ethical basis of financial markets (Sobolev (2019)), methods of regulatory reform (Palermo et al. (2017), Leaver and Reader (2019), and Roulet (2019)) and associated corporate governance developments (e.g. Koch-Bayram and Wernicke (2018), Shi et al. (2017); and Zorn et al. (2017)). At the heart of this burgeoning discourse, adverse regulatory outcomes and the incidence of financial offending have been used to justify, analyse and comprehend wrongdoing by financial firms.

Despite the centrality of such regulatory reporting underlying this developing literature, it remains unclear how much financial offending is undertaken yet not reported. We address this concern and contribute to this literature through advancing a novel method to quantifying the proportion of financial offending which is detected and not detected and if these rates of regulatory detection have proved successful in deterring future financial offending in the UK.[1]. The purpose of this exercise is to address the question of whether financial regulation has improved or otherwise since the global financial crisis.

A key issue in understanding the extent and scope of crime is 'partial observability'; the idea that we only ever observe the number of wrongdoers who are detected. Such headline figures tell us little about the number of miscreants who evaded capture, or whether our actions are dissuading individuals from engaging in misconduct. In the context of quantifying financial misconduct, this results in distorted estimates of regula-

[1] 'For example, in the UK, fines and remediation totalled \$38.7 billion (\$56 billion) between 2011 and 2014, accounting for 60% of bank's profits (The Economist 2016).1 A similar process is witnessed in the USA where financial institutions paid around \$139 billion in fines between 2012-14 (Zingales, 2015)

tory efficacy and uncertainty over whether the increased severity of recent punishment has served to provide an effective deterrent to breaching regulations.

This uncertainty is troublesome, as the frequency and scale of regulatory sanctions has also promoted unease in financial industries, reduced valuations of the incumbent firms' sector and reinforced public skepticism as to whether this sector can ever be reformed (Group of 30, 2015).[2] What is clear, however, is that regulation, compliance and enforcement activities come at considerable cost for both firms and regulators, with significant and consequential repercussions for corporations and their management of financial misconduct (Marcel and Cowen, 2014). From the perspective of firms, the levying of large fines may be viewed as unfair, placing an inordinate cost on the shareholders of financial firms rather than those persons and corporate entities responsible for offending (Goodhart, 2017). Moreover, financial regulation can have unintended consequences such as relocating offending from one nation or sector to another (Zeume, 2017) or can result in victims of the crime being punished by the market when they are identified in the 'naming and shaming' of perpetrators (de Batz, 2020). Meanwhile, from a regulatory perspective, a large number of countries have reactively invested heavily in changing their regulatory architecture in recent years; not least the UK where the Financial Conduct Authority (FCA) was established after the break-up of the Financial Services Authority in 2013 (hereafter the FSA).

For the above reasons, it is undoubtedly important to enhance methods which can quantify the effectiveness of regulatory detection and deterrence in order to best minimise misconduct by firms and individuals. In light of the almost total absence of empirical analysis addressing these questions in the financial sector, this paper provides an innovative approach to deliver the first broad evidence that the post-crisis overhauling of UK financial regulation, which started in 2010, has improved the rates of detection and deterrence of financial misconduct.

[2] There is a growing body of literature identifying the often negative (Delis et al., 2016; Danisewicz et al., 2018) and positive (Pasiouras, 2016) outcomes from regulatory enforcement actions.

Our main contribution is methodological; we use a Capture-Recapture (CR) method to deal with the aforementioned partial observability concerns. This method, frequently used in life sciences such as ecological studies, allows estimation of unobserved population parameters from taking repeated samples for this population. Applying CR to financial services provides a new approach for estimating detection and, crucially, inferring subsequent effects on deterrence—which would otherwise be unobserved. This, alone, constitutes a valuable and novel addition to the body of research concerning the accountability of financial regulators and the efficacy of regulatory arrangements. Naturally, we do not claim that our method provides an infallible solution. But we believe that it helps us develop our understanding of a problem, which otherwise would stay largely unobserved. Moreover, in terms of practical implementation, our method is more parsimonious than most others proposed before,[3] although still suffering from one main limitation of previous models, i.e. it only provides upper-bound estimates of the probability of detecting offending businesses.

We offer evidence that whilst detected breaches of UK financial regulation fell after 2010, detection rates have increased, driven mainly by the improved ability to detect mis-selling and fraudulent behaviour (although still less than 1 in 4 offences are detected, and this number is less for reporting offences). Using our proposed framework, these two findings together imply that the corresponding level of deterrence rose in this period. The causes of these changes in deterrence are uncertain yet could include highly-publicised changes in regulatory structures, enhanced media coverage of financial misdemeanors, the effectiveness of punishments, cultural change in the banking industry, as well as the enhanced detection quantified in this work.

[3] Table A1 provides a summary of alternative approaches and their data requirements.

2 Literature review

The scope of financial regulation defies precise definition (Allen and Carletti, 2010) and optimal regulatory outcomes are hard to gauge (McCraw, 1975). This uncertainty has resulted in financial laws and regulation being viewed to have both a positive (e.g. La Porta et al., 2006) and negative (e.g. Stigler, 1964) influence on the operation of financial markets. Similarly, increased enforcement of financial regulation has been interpreted as evidence both of more active and successful regulators (Stigler, 1970) and failure for 'allowing' regulatory transgression (Becker and Stigler, 1974). Clearly, this conflagration of criticism faced by regulators, as either undertaking too little or too much regulatory action simultaneously requires further investigation. To engage with these issues, this discussion considers the importance of addressing the partial observability problem in the context of financial and economic markets, and previous studies which have addressed the issue.

2.1 Quantifying partial observability in financial wrongdoing

As noted earlier, an impediment to measuring the success of regulatory policy in detecting and deterring aberrant behaviour is the inability to observe those cases where a misconduct exists yet is not detected. While we are aware how many firms and individuals have been caught for breaching regulations, due to the illicit nature of financial misconduct, it is unclear how many firms' transgress regulations and are not caught. This is problematic because the number of cases detected does not lead to an unambiguous assessment of regulatory performance. In an example where the number of cases detected increases, this could imply that the regulator has become better at detecting misconduct, but it could also be a sign of weakening deterrence. This partial observability presents a major challenge to the assessment of regulation; non-detection of misconduct is likely to lead to underestimation of the true level of misconduct, overestimate the effectiveness of regulation and base assessment on a biased sample.

Sample selection issues such as partial observability arise when analysis is limited to a non-random subsample of interest. Observations of firms caught for regulatory failings are selected through a process that is not independent of the outcome of interest (i.e. whether the firm has breached a certain regulation and is affected by a diversity of non-random influences). This non-random selection can arise from explicit and incidental sources, such as data availability (i.e. the sample maybe truncated or censored) or arises when other, unobserved endogenous variables determine the selection process. Reflecting the general nature of this statistical challenge, a variety of methods have arisen in different disciplines to address partial observability in the context of corporate offending.

Some of the methods emerged in the accounting and finance literature (we label these control detection methods), building on the work of Poirier (1980) and Feinstein (1990). Wang et al. (2010) and Wang (2011) examined the incidence of corporate fraud and how the attributes of captured firms can be used to estimate the characteristics of firms likely to undertake similar transgressions. These logistic regression models consider the latent processes underlying fraud commissioning and detection distinctly to estimate the characteristics of the population of potential offenders. Using this approach, Wang et al. (2010) report that firms are more likely to commit fraud when business conditions are good, yet less so when investor confidence becomes very high. Wang (2011) further broadened the range of factors linked to corporate fraud. This type of logistic regression model has subsequently been employed to assess the influence of social links between directors on fraud (Kuang and Lee, 2017) and accounting mis-statements (Zakolyukina, 2018), and has been developed to address partial observability directly (Lancaster and Imbens, 1996).

Quantifying partial observability using these logistic regression techniques also has drawbacks. These approaches require far more data than just the frequency of offending, demanding data as to the characteristics of firms concerned. Furthermore, the models forwarded by Wang et al. (2010) and Lancaster and Imbens (1996) have been reassessed and shown to be sensitive to the model assumptions (see Hahn et al. (2016) and Phillips and Elith (2013) respectively).

Studies of accounting fraud and mis-statements have also employed a range of methods to quantify undetected offending. Financial accounts and reports, the subject of many accounting frauds, have been used in the estimation of levels of detected and non-detected offending. Descriptive statistical methods (Dechow 2011), machine learning (Cecchini et al., 2010) and deviations from Benfords law (Amiram et al., 2015) have all also been used to predict unseen offending. These approaches whilst promising, are data intensive requiring data on the subject of offending, in addition to the occurrence and frequency of offending; Amiram et al. (2018) provides a review of these techniques.

The total level of fraud in a market has also been estimated using natural experiments. Dyck et al. (2013) looked at the failure of Arthur Andersen (AA) in one such experiment. In the early 1990s, and following the collapse of AA, a large number of firms suddenly required a new auditor. These firms were assumed to be closely examined by their new auditors, enabling estimates of fraud throughout corporate America to be made. From this sample of closely scrutinised firms, it was estimated that 14.5% of large US publically limited firms engaged in accounting fraud.

More recently, to estimate the prevalence of illegal activities in cryptocurrencies, Foley et al. (2019) offer two different approaches: first they reconstruct the network of transactions between market participants using blockchain data, and second, they study the characteristics of observed illegal activities in order to distinguish between legal and illegal users (much in the same vein as the above control detection methods). From a very different perspective, using a case study based approach, O'Donovan et al. (2019) study in great detail the client network of an offshore service provider, Mossack Fonseca, from which they infer conservative estimates of the prevalence of offshore secrets among businesses.

Lastly, Capture-Recapture (CR) methods have been used to address partial observability in a number of settings.[4] These techniques accommodate situations where populations change over time, when heterogeneity within the sample exists, and if time dependence influences recapture. In its most simple form, CR models estimate a population through examining repeated random samples taken from the population of interest. In this process, samples are marked and replaced, with common observations recorded. The proportion of recaptured individuals is then used to infer population parameters such as population size, capture and survival rate.

Though originally developed for use in ecological settings to overcome uncertainty around animal populations, similarities with the intrinsic uncertainty concerning illicit behaviour mean these CR approaches have been applied to the analysis of the frequency of economic crimes and similar forbidden conduct where the true scale of activity is obscured. For example, applying these techniques Ormosi (2014) estimated that 13-17% of European cartels were caught in any given year by competition law regulators between 1985 and 2009. Other crimes such as prostitution (Rossmo and Routledge, 1990), marijuana cultivation (Bouchard, 2007), car theft (Collins and Wilson, 1990) and criminal desistance (Bushway et al., 2003) have also been examined with these methods.

In summary, addressing partial observability is an emergent subject and, as such, applying new techniques is important to address measurement concerns. We propose that CR methods deserve to be investigated to establish how well they can tackle the statistical concerns raised above (such as partial observability and sample selection), as they focus on the estimation of population characteristics from incomplete data. In this respect, previous work has shown it to be effective in the study of 'white-collar' crime (Ormosi, 2014). In Table A1 in the Appendix we summarise this overview of previous methods in order to make a direct comparison to our preferred choice of method. The

[4]For an introduction to CR methods see Amstrup et al. (2005), Williams et al. (2002) or Burnham and Anderson (2002).

table shows that alternative approaches and methods, with their different assumptions and requirements are either a concern owing to misapplication, inappropriate due to the particularities of the parameters they estimate, or difficult to implement owing to onerous data requirements. Finally, CR methods operate with fewer underlying assumptions and are therefore more parsimonious than these alternative approaches.

3 A simple theoretical framework

To address the problem of partial observability we formulate a simple model of detection and deterrence. Denote the population size of all registered financial sector firms by N, the probability of deterring a regulated business from committing an infringement by ω, and the probability of detecting an infringement by ρ. The number of cases detected (n) is then given by $n = (1 - \omega)\rho N$ (i.e. a product of the total number of firms, and compound probabilities of the proportion of these firms not deterred from misconduct, and the probability of detecting these firms' aberrant behaviour). From this, the probability of deterrence is defined as:

$$\omega = 1 - \frac{1}{\rho}\frac{n}{N} \tag{1}$$

We denote the proportion of firms under financial regulation that engage in regulated misconduct by $\eta = n/N$. From Equation (1), it is straightforward to conclude that deterrence increases if $\Delta\eta/\eta > \Delta\rho/\rho$, where, in the analysis of Section 5, $\Delta\eta$ and $\Delta\rho$ denote changes in, respectively, the average values of η and ρ between the period up to and including 2010 and the average for the period 2011 onward. The choice of Dec 2010 as cut-off point is primarily motivated by the fact that the post-crisis overhaul of UK financial regulation started in 2010. Moreover, we also find that the end of 2010 denotes a structural break point in the dataset (see Section 4.1). Therefore, we can establish the following proposition.

Proposition 1 *A sufficient set of conditions to establish we have observed increased deterrence after 2010 is:* $\Delta\eta < 0$ *and* $\Delta\rho \geq 0$*, or* $\Delta\eta \leq 0$ *and* $\Delta\rho > 0$*.*

Deterrence increases if the pre- and post-average proportion of firms engaging in misconduct declines (or remains static) after 2010, coupled with stagnant or increasing average probability of detecting an offending firm over the same interval. Accordingly, to test this proposition, our empirical strategy consists of two main elements: First, we estimate the impact of our structural break (Dec 2010) (see Figure 3) on detection probability ($\Delta\rho$), then we estimated how the relative number of detected cases (to elicit $\Delta\eta$) changed after 2010. Finally, from our estimates of $\Delta\eta$ and $\Delta\rho$ we can infer how regulatory deterrence has changed in the UK since 2010.

4 Data and methods

In this section we outline the sources of the data, its format and the processes employed to code and transform it into firm-level data usable for the study. We also introduce the descriptive and inferential techniques employed.

4.1 Data sources and variables

The study employs data from a number of sources. The primary data sources are the 'Final Notices' issued by the UK financial regulators, the FSA (in operation between 2001-13) and the FCA (operating since 2013).

A sample of 1,869 UK Final Notices were collected, varying in document length from one to ninety pages and issued to firms and individuals between 2002-2015. The Notices all include the date of the offence, the duration of offending, the date of the regulatory intervention (i.e. date of the 'Final Notice' from which we create yearly and quarterly measures of offending), firm characteristics, punishments and the nature of the offence.

This hand-collected data was supplemented and manually cross-checked, using Financial Regulator Annual Reports and the Financial Services Register. Furthermore, Supervisory, Warning and Decision notices and press releases issued by the FSA and FCA, as well as appeals to the Financial Services and Markets Tribunal (359 documents in total) were also consulted to augment and confirm Final Notice details. These Final Notices, similar to regulatory reporting internationally, include multiple and different forms of offending which vary from small to substantial levels. Regulatory reporting at the individual contract, transaction or customer level is not available publicly in the UK. To alleviate the aggregation bias emerging in all such regulatory reporting we consider all data at the firm level.

The data was initially collected and coded at the level of individual offences according to classifications previously applied to Final Notices within annual reports issued by the FSA (FCA) and to comply with existing forms of coding used within the Financial Services Register. The data was categorised according to the different types of offences recognised by regulators. The coding exercise included a unique identification number allocated to all regulated firms and individuals in the Financial Services Register. We use this to ensure that there are no cases of double counting by firms and individuals changing their names over the sample period.

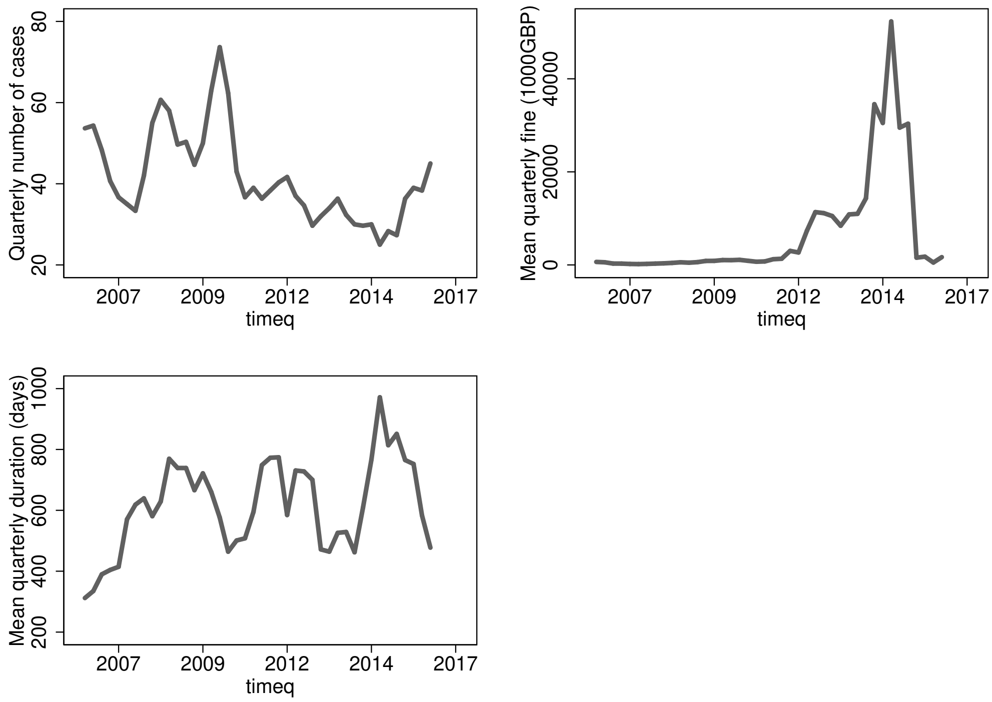
To transform the Final Notices into a firm-level data set, a number of assumptions have been made, resulting in the exclusion of some observations. In 22 cases, a Final Notice refers to a rejected application to extend a regulatory function, and in 32 Final Notices, the judgement concerned a form of market abuse such as insider dealing and/or an accounting or listing reporting irregularity involving a non-financial firm. Furthermore, in 135 Final Notices a person or firm has provided financial services whilst not being regulated. These cases fall outside our frame of reference (focusing on regulated financial firms or employees, and breaches of financial regulation) and are, therefore, excluded from the analysis.

The remaining observations relate to 1,389 firms, including situations where firms, or their employees, have been issued multiple Final Notices in the study period. For 68 Final Notices, multiple firms were involved; in these cases, each firm involved was considered distinctly. Finally, the data was annualised, such that we considered whether a firm or its employee(s) had offended in a given year (multiple offences within a single year were only considered once). Overall, 1,295 firms only offended in one year and nearly 100 committed offences in two or more different years.

Lastly, data was collected from regulators' annual reports and accounts and other sources. This wider data collection, and specifically data drawn from the Financial Services Register, allowed the determination of the population of regulated firms operating in the UK during the sample period. Further, it allowed for the creation of control variables on regulatory resources, thereby allowing our analysis to differentiate between the effects of regulatory resources versus wider macro-economic concerns. We summarise the data and variables used in four tables in the Appendix. The variables employed as co-variates in our regressions are described in Table B1. Table B2 outlines the descriptive variables for the enforcement cases considered at an offence-level. Table B3 provides descriptive statistics for firm level data over time and descriptive statistics of the co-variates are reported in Table B4. Of particular interest is the significant rise in regulatory resources, such as employees and operating costs of regulators between 2002 and 2016 (Table B4).

Figure 1 shows that the quarterly number of cases dropped after 2010 (the vertical line shows Q1 2011). The average fines levied on firms and individuals displays an upward trend (the fall after 2015 is due to the censoring point in our data). The quarterly average duration of offences appears to move around a steady trend.

Figure 1: Number of cases, duration, fines (3-year moving averages)



Regarding the type of observed financial misconduct and the punishments applied, Table 1 shows that reporting and compliance offences are the most frequently observed (55%). This incorporates many actions: from non-payment of regulatory fees, to failures to submit transactions data. Mis-selling of financial services (the sale of a financial service, which is not needed by a customer) and fraud (many of which are associated with corresponding criminal proceedings) together make up around a third of cases, whilst other case types, such as money laundering, feature in much smaller numbers. Turning to punishments, non-financial punishments such as prohibition of individuals from working in the financial sector, or cancellation of regulatory permissions to trade as a financial services firm are used more frequently.

Figure 2 presents the ratio of 'captured' offending firms (those which were caught)

Table 1: Types of offences and punishments

Type of Offence*	%	**Punishments***	%
Reporting/compliance	55.7	Public censure	3.5
Complaints handling	3.7	Prohibition	21.7
Market abuse	8.2	Fine	27.4
Fraud and theft	17.3	Cancelled regulatory permissions	51.6
Mis-selling	13.8	Disgorgement	1.6
Money laundering	1.3	Other punishment	0.1
Other offence	1.4		
Client funds	0		

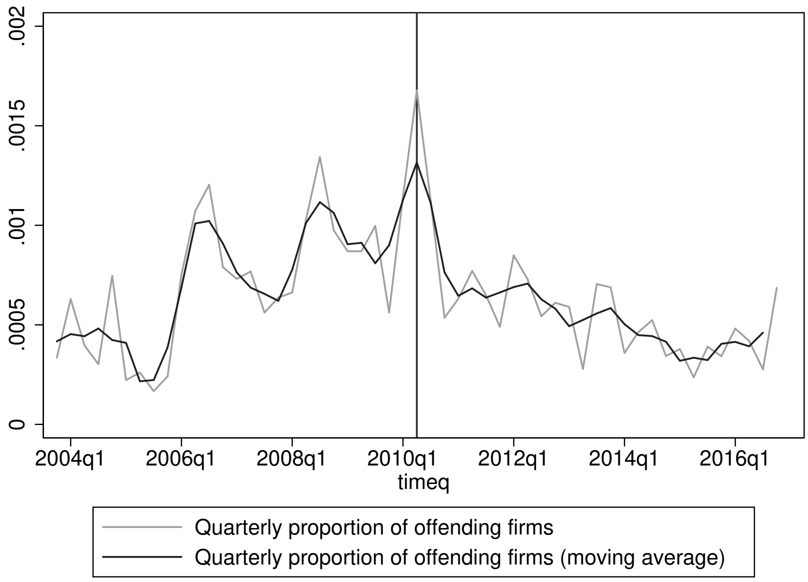
*percentages do not add up to 100% as more than one type of offence or punishments may be relevant to any one case.

to the total number of registered financial firms, which is what we denoted as η in the framework presented in Section 4.2.[5] To smooth the two curves and focus on the longer run trends rather than short-term variation, we also report 3-quarter moving averages.

Figure 2 shows how $\eta = n/N$, the proportion of the total number of registered firms, found guilty of some form of misconduct, changes over the sample period: increasing until an apparent break point in 2010, at which point it declines – in line with the overall number of offences (the vertical line denotes Q2 2010).

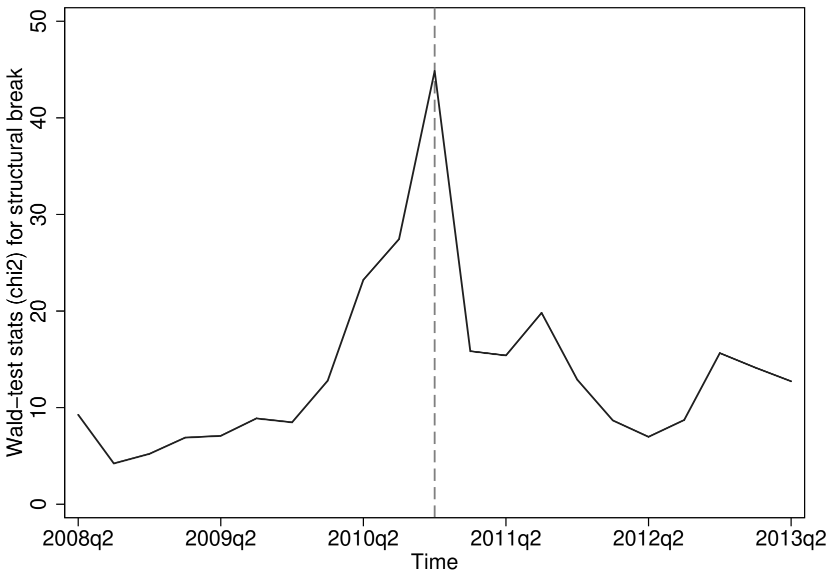
[5]Based on firm level data on observed offending and reoffending, the total number of regulated firms operating in UK financial services industries are reported in Table A3 in the Appendix

Figure 2: Proportion of offending regulated firms



To formally confirm the existence of a structural break in 2010, we run a set of Wald tests of whether the coefficients in a time-series regression vary over the periods defined by possible break dates. Figure 3 shows the test statistics for these break dates, which shows a peak at Q4 2010, implying the highest probability of a structural break at this point in time.

Figure 3: Wald test statistics on a set of potential break dates



In the context of partial observability, and considering the information from Figure 2 in isolation, one could jump – potentially mistakenly – to the conclusion that enforcement has become less effective in the UK in the post-2010 period. We will show below that this would be an erroneous conclusion as both Figure 1 and 2 mask key information. Given $n = (1 - \omega)\rho N$, a drop in the number of detected offences could be a sign of a decline in detection rates, but it could also be due to improved regulatory environment with improved deterrence and fewer offences to detect. Whereas the former would be an undesired change, the latter is clearly a positive development.

The need to unpick these contorted interpretations and accurately assess regulatory performance motivates the use of the CR framework to distinguish between these possible explanations and allow identification of these different effects.

4.2 Method

To estimate one of the components in Proposition 1, the change in the rate of detection (ρ), we turn to Capture-Recapture (CR) methods. Ormosi (2014) offers a detailed explanation of the terminology, however, given the novelty of the method in the analysis of business behaviour we provide an intuitive and a moderately technical explanation below.

CR methods are based on taking repeated samples of the analysed population. With every new sample, one looks at the proportion of recaptured individuals (those which have also been captured in previous samples) in order to make inferences on population parameters (such as survival and detection rates). In their simplest forms, CR methods would assume that the population does not change between samplings, or that the only change is through death and birth (closed population methods). To account for a more realistic scenario (e.g. continuously changing population, heterogeneity across individuals, time-dependence) a number of robust open population CR methods have been developed for estimating dynamically changing population characteristics.

To give a simple example, imagine that someone takes repeated samples from a population. With every sample they record an identifier of the individuals that they sampled and then put them back in the population to be available for subsequent samplings. Individuals can 'die' between samplings (or survive to be recaptured in future sampling), or might survive but evade future capture - in both latter cases they are never seen again. The idea is to design a likelihood function that describes, for each sampling period, some probability of detection and survival. For this likelihood function the survival and detection parameters with the highest likelihood of generating the observed data are estimated.

Using formal notation, the CR likelihood function describes the probability of observing an individual at time t (detection probability), and the probability of the individual subsequently surviving to period $t+1$ (survival probability).

Applying this intuition to financial misconduct, ‘to capture’ refers to the detection of financial misconduct, therefore we denote by ρ_{tm} the probability of detection of a financial misconduct of firm m at time t. The estimation of detection rates in an open population CR setting are conditional on previous capture, i.e. it only provides information on those firms that are caught at least once, which might be different from those that are never caught. Because detection rate is conditioned on previous detection, it can also be thought of as a rate of recapture. For this reason, our detection rate estimates can only be interpreted as an upper-bound of the ‘true’ detection rate. It is an upper bound because, by definition, the detection rate of those offenders that are never captured must be smaller than the detection probability of those offenders that are caught at least once. 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.

The survival rate (ϕ_{tm}) in this application is an apparent survival estimate. It is apparent because, if a captured individual (a detected offender) is not captured again in future time periods it is not known whether it has ’died’ because it does not exist anymore, because it refrains from future financial misconduct, or because it joins the sub-population of those offending firms that are never re-captured (for example because the firm developed techniques to evade regulatory detection). For the analysis of financial offenders this means that an offender ‘survives’ if it still exists, and can potentially commit an offence again. This could also be thought of as the ’survival’ of detectable evidence related to the offence, which is generated when the offence is committed and this evidence remains alive until discovery.

The construction of our likelihood functions follows a very simple logic, explained through the following general example (for a more detailed explanation see Ormosi, 2014). Take a time period bookended by t and $t + 3$, where sampling takes place at t,

$t+1$, $t+2$, and $t+3$. An individual (say a regulated financial company), denoted by m that was captured (found guilty) at t and $t+2$, but not seen at $t+1$, and $t+3$, will have a capture history:

$$CH_m = (1, 0, 1, 0) \tag{2}$$

The probability of observing this pattern m is given by:

$$\Pr\{CH_m | release\ at\ t\} = \phi_t(1-\rho_{(t+1)})\phi_{(t+1)}\rho_{(t+2)}[(1-\phi_{(t+2)})+\phi_{(t+2)})+\phi_{(t+2)}(1-\rho_{(t+3)})] \tag{3}$$

This function displays an important feature of CR models. The observation of each individual (or, in this case, firm) is conditional on being captured at time t.[6] At period t, we capture the offending firm (i.e. there is a recorded offence) and this firm is 'released' back into the population. Does it survive to period $t+1$? Yes, we know that because although the individual (the firm) is not seen at $t+1$, it survived, as it is later seen at $t+2$. For this reason we record some probability of survival at time t, denoted as ϕ_t in Equation (3). Moving on to period $t+1$, we know that there was no detection (so we record the probability of no detection, $(1-\rho_{(t+1)})$, and we also know that the individual survived to $t+2$ because it is captured at that stage – we record this as a probability of survival at time $t+1$, denoted as $\phi_{(t+1)}$. The rest of Equation (3) follows a similar logic. The expression in the squared brackets denotes the scenarios associated with not seeing the given individual after $t+2$ (i.e. there is no information on whether the individual survived after $t+2$ or not), and accounts for both the possibility that it has not survived, or that it did but we didn't detect it in $t+3$.

In the present case, to record data for a CR analysis, we need to log the capture

[6] This is why the estimated parameters can only be interpreted for individuals (or, in the present case, firms) that have been captured at some point.

histories of every firm for every time period similarly to Equation (2). The capture histories for all firms are then organised into an $i \times K$ matrix $\mathbf{X}$ (i is the number of offending firms detected over the time period studied, K is the number of years – or sampling periods – in our sample, $m \in i$, and $t \in K$), where $x_{mt} = 1$ if firm m was captured at sampling occasion t and $x_{mt} = 0$ otherwise.[7]

Let ϕ_{tm} denote the probability of an offending firm m surviving time $t = 1, 2, ...$, which is the conditional apparent survival from year t to year $t+1$, given that the same firm is 'alive' at the beginning of year t. Denote the probability of firm m being captured at sampling occasion $t = 1, 2, ...$ by ρ_{tm}.

If we denote the time of the first capture of firm m as t_m, the last capture as l_m and the departure ('death' or migration) from the sample as $d_m (> l_m)$ we can generalise the probability of observing any capture history (shown in a simple form in Equation (4)). In this general form we can sum up for all possible departure (i.e. disappearing from the population) times d_m, which is necessary as d_m is typically not observed. Note that this is a general parametric form, which assumes that both capture and survival rates are time-dependent. As we are mainly interested in the effect of the 2010 break on the rate of detection (capture) and survival, later we will estimate a model where ϕ_t and ρ_t can assume only two values (pre, and post-2010):

$$\Pr(CH_m \mid f_m) = \sum_{d_m = l_m}^{K} \left\{ \left(\prod_{t=f_m}^{d_m - 1} \phi_t \right) (1 - \phi_{d_m}) \times \left(\prod_{t=f_m+1}^{d_m} p_t^{x_{mt}} (1 - p_t)^{(1 - x_{mt})} \right) \right\} \quad (4)$$

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:

[7] In Table B5 in the Appendix we provide an sample section of our capture history matrix.

$$L = \prod_{m=1}^{2^K-1} \Pr(CH_m \mid f_m) \tag{5}$$

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

Of course financial misconduct is fundamentally different from the typical applications of CR models, which warrants a more detailed discussion of whether the assumptions required for unbiased CR estimates are tenable for our research purposes. As this is a rather technical discussion, we included it in Section A.1 in the Appendix.

4.3 Model choice

As implied by the above discussion, CR models can have many (fully time dependent parameters) or relatively few estimated parameters (time constant parameters), and the choice of the relevant model is down to two things: the assumption of the researcher (e.g. is there any reason to think that parameters are stationary) and the goodness of fit of the chosen model.

To determine which model specification to use, intuition would suggest that, as we are interested in the change after Jan 2011, it would make sense to look at a simple model where detection and survival rates can assume two values for the two time intervals: before and after Jan 2011. In addition, we would be interested in the effect that capture has on our parameters of interest in the time periods after capture (called 'trap dependence', with reference to animals which become wary of traps following capture). For each of the two intervals therefore we should have two parameters estimated, one only measuring detection and survival rates immediately after capture, and another one for all other years. For example, if a firm is detected as offender in 2011, we would have an estimate of detection and survival probabilities within 1 year of the detection, and another estimate for 2012 onward. We show that this intuition is closely reflected by the ranking of models in terms of their goodness of fit.

Using a number of model fitting tests (explained in Section A.2 in the Appendix), it appears that the models where we only estimate before and after values perform better than the other (4 out of the 5 best performing models were such). Based on goodness of fit, the best performing model is the one that we intuitively thought would be most credible: where we estimate parameters before-and-after Jan 2011, and we allow for trap-response (i.e. the parameter immediately following detection is different from the subsequent parameters). For the discussion that follows, we focus on this model.

5 Results

5.1 Detection and survival rates

First, we looked at the change in detection rate, $\Delta\rho$. As explained earlier, detection rate estimates can only be interpreted as an upper bound estimate, and the true detection rate is possibly smaller than our estimates.

The estimates for the change before and after Jan 2011 are shown in Table 4. The table has three main rows. In the first we report estimates for the whole sample, including all offences. These estimates can be thought of as average detection rates across all types of offences. The second main row shows estimates for reporting offences only, and the second row contains average detection rates for all offences except reporting offences (fraud and theft, mis-selling, complaints handling, market abuse, and money laundering).

Table 2 shows that when averaging over all offences, the probability of recapture (detection) in the immediate aftermath of a previous detection did not increase after 2010. However, the probability of recapture after 1 year following a previous capture has increased significantly (from 10% to 25%). As these are upper bound estimates

Table 2: Recapture rates before and after Jan 2011

		Before Dec 2010		After Dec 2010	
	n	Recapture within 1 year	Recapture after 1 year	Recapture within 1 year	Recapture after 1 year
All offences	1591	0.046	0.099	0.274	0.253
[95% CI]		[0.016; 0.126]	[0.062; 0.154]	[0.069; 0.660]	[0.159;0.376]
Reporting offences	971	0.055	0.155	0.27	0.067
[95% CI]		[0.007; 0.324]	[0.026; 0.556]	[0.026; 0.839]	[0.009; 0.360]
All other offence	662	0.143	0.105	0.216	0.273
[95% CI]		[0.074; 0.259]	[0.066; 0.164]	[0.076; 0.482]	[0.166; 0.414]

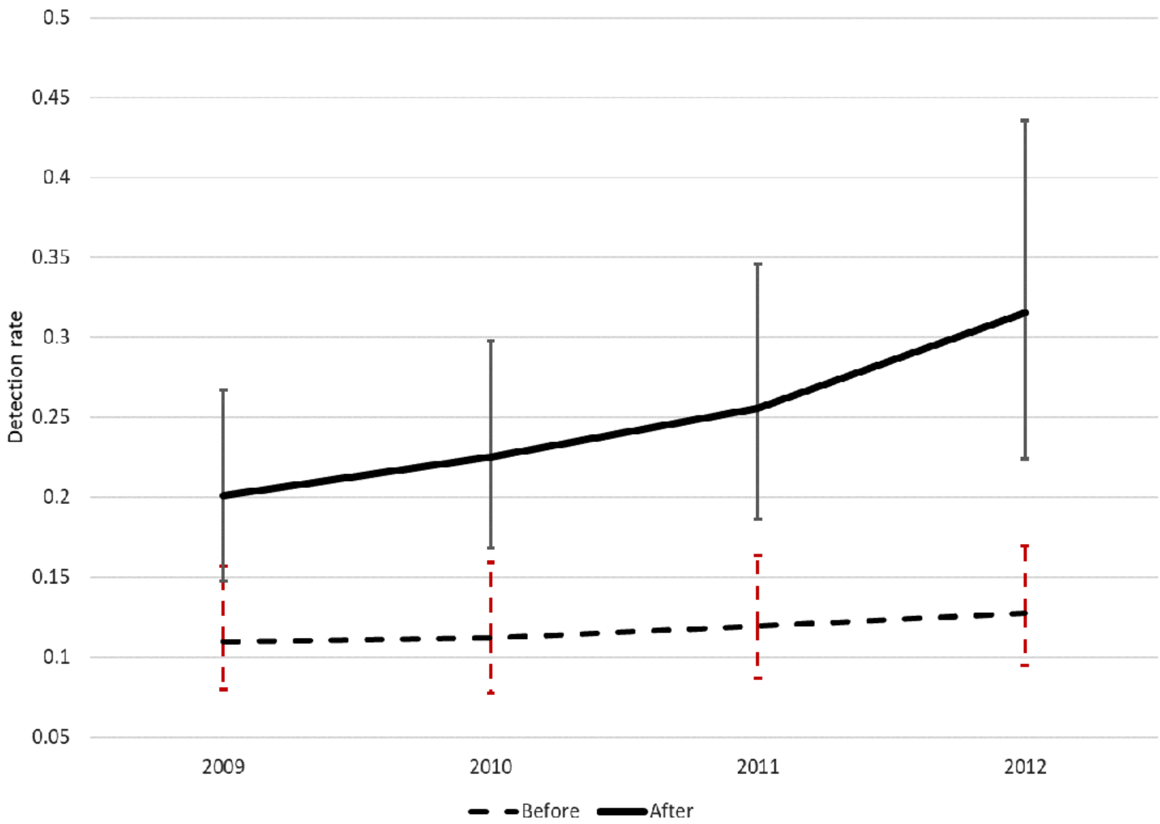
(as explained earlier), this means that before 2011 the upper bound of the probability of detecting a financial offence was 10%, whereas the upper bound of the probability of detecting an offence after 2011 was 25%. If one assumes that the bias from the unobserved firms did not change after 2010, this is evidence that detection rates have increased.

When looking at reporting/compliance related offences only, we find no evidence of changing detection rates. However when looking at all other offences, excluding reporting offences (this subset consists dominantly of mis-selling and fraud related offences) then we find a significant increase in detection rates. This result implies that the observed increase in detection rates is driven by improved detection rates of mis-selling and/or fraudulent behaviour.

As a sensitivity check, we re-estimated the main model for all offences, but assuming a structural break at different time points (2009, 2010 (used above), 2011, 2012). Figure 4 shows the long-term recapture rate estimates for each of these assumed structural breaks. For each assumed structural break one estimate shows the before, and another denotes the after-break estimates.

Figure 4 indicates that the difference in recapture rates between before and the after the break gradually opens up for years following 2009. In 2009, the difference between the estimates is not yet significant, it becomes significant (at 95% level) in 2010, and the difference grows in 2011 and 2012. This provides strong support to our story that the re-design of the UK financial regulatory landscape, which started in 2010, gradually affected the behaviour of UK financial businesses.

Figure 4: Comparing detection rate estimates using various years as structural break (with 95% CI)



Although not central to our main story, in Table 5 we report survival rate estimates as well. Survival includes a number of things (the firm still exists, and that it is still capable of committing an offence) as explained in Section 5.2. The results below include two interesting findings: in general, the chance of survival (an offender remaining in operation following a detection) is very low in the year of the detection. However, businesses that survive the critical first year after detection, have a very good survival probability. This is in line with intuition. This main finding remains the same when before and after are compared.

Table 3: Estimated survival rates, pre-, and post-2010

	Recapture within 1 year	Recapture after 1 year
Before 2011	0.155	0.986
[95% CI]	[0.118; 0.201]	[0.237; 0.999]
After 2010	0.121	0.704
[95% CI]	[0.069; 0.205]	[0.595; 0.793]

5.2 The number of offences

Next, we formally test whether the drop in the number of detected offending firms in the UK is significant by regressing the proportion of firms guilty of misconduct ($\eta = n/N$) on a number of independent variables. For this we regress the quarterly number of cases on a before-after dummy variable and a number of covariates. Because some of these variables vary significantly in their magnitude, we use standardised values for all but the dummy variables; hence the coefficients should be interpreted as the standard deviation change in the dependent variable associated with a 1 standard deviation change in the independent variable. To remove the effect of size, the stock index, the net operational costs, the employee number, and the employee costs were standardised by dividing through by total assets. Table 6 displays the results of four different model specifications. The first column shows the estimates where the dependent variable is the proportion of detected offending firms (η) and is estimated using a number of time-dependent covariates as previously specified. The second column is the same as the first column but without covariates. Columns 3 and 4 estimate the same models but now using the number of detected offenders as dependent variable.

The first row of Table 4 shows the before-after estimator ($\Delta\eta$), which is significant and negative for all model specifications. This is unsurprising, given the visually apparent drop in the number of detected offenders after 2010, as observed in Figure 2. This is evidence that our second sufficient conditions to establish an increased deterrence rate (Proposition 1) holds as $\Delta\eta < 0$ (i.e. the change in detected firms has declined since 2010). Notwithstanding the lack of significance associated with the visibility of fines (which we expected to negatively impact on errant behaviour), we refrain from further interpretation of the effect of the co-variates to maintain focus on the effect of the post-crisis effect indicated by the 2010 structural break.

Table 4: Regression results on the proportion and number of offending firms (η)

	(1)	(2)	(3)	(4)
	Proportion	Proportion	Number	Number
Before-after dummy	-2.167***	-0.838***	-1.994***	-0.642**
	(0.492)	(0.24)	(0.526)	(0.27)
Number of cases with fines (1 year lag)	0.0641		0.124	
	(0.137)		(0.14)	
Stock index	-1.780**		-1.688**	
	(0.831)		(0.829)	
Total assets	0.18		0.0876	
	(0.143)		(0.16)	
Net operational costs	-2.127		-3.513*	
	(1.665)		(1.833)	
Employees	1.785		2.531**	
	(1.122)		(1.149)	
Employee costs	2.753*		3.426**	
	(1.594)		(1.694)	
Year	0.226**		0.279***	
	(0.0914)		(0.0879)	
Observations	49	49	49	49

Standard errors in parentheses
=* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

In Proposition 1 we formulated a sufficient pair of alternative conditions needed to establish that financial regulations were more deterring of misconduct after 2010. One of these conditions required that the proportion of detected offenders' decreased ($\Delta\eta < 0$) and the rate of detection did not decrease ($\Delta\rho > 0$). Evidence supporting both of these conditions in the UK is provided, where detection rates have remained constant and the number of detected cases dropped significantly. We believe this is strong evidence that the UK regulatory environment improved after 2010 as the rate of deterrence has risen.

6 Robustness checks

We present two cases where we diverged from our original assumption. First, we look at estimating a model without trap-dependence, and second, we estimate our main model using a sample that only contains firms as offenders.

6.1 No trap dependence

Table 5 shows the detection rates where we assumed that there was no trap dependence. These can be thought of as before-after averages. These results are qualitatively the same as the results presented earlier. Detection rates – as an average for the whole sample – increased significantly. This was driven by the increase in offences other than reporting/compliance, more specifically, the increase in detection mis-selling offenders is where detection rates improved and it remained unchanged in other offences. Both of these robustness checks deliver results that point in the same direction as our main results.

Table 5: Detection rates without trap dependence

	Before Dec 2010	After Dec 2010
All offences	**0.036**	**0.075**
[95% CI]	[0.027; 0.048]	[0.051; 0.110]
Reporting offences	0.009	0.011
[95% CI]	[0.003; 0.028]	[0.003; 0.043]
All other offence:	**0.046**	**0.142**
[95% CI]	[0.033; 0.065]	[0.092; 0.213]
Mis-selling	**0.013**	**0.154**
[95% CI]	[0.007; 0.023]	[0.062; 0.331]
Fraud	0.071	0.094
[95% CI]	[0.030; 0.162]	[0.026; 0.285]

6.2 No individual offenders

Below we present the results where individual offenders were removed from the sample and the sample only contains firms as offenders. Table 6 shows the detection (recapture) rates for business offenders only. The results are qualitatively unchanged from those reported in Table 2.

Table 6: Recapture rates before and after Jan 2011 - firms only

		Before Dec 2010		After Dec 2010	
	n	Recapture within 1 year	Recapture after 1 year	Recapture within 1 year	Recapture after 1 year
All offences	901	0.187	0.112	0.192	0.225
[95% CI]		[0.123; 0.273]	[0.079; 0.154]	[0.089; 0.366]	[0.155;0.315]
Reporting offences	591	N/A	0.136	N/A	0.132
[95% CI]		N/A	[0.011; 0.689]	N/A	[0.019; 0.545]
All other offence	310	0.027	0.076	0.122	0.282
[95% CI]		[0.004; 0.177]	[0.036; 0.151]	[0.024; 0.446]	[0.152; 0.465]

Finally in Table 7 we show the estimates for the change in the proportion of detected offences ($\Delta\eta$) when only considering business offenders in our sample. Again, the results are of the same sign (and somewhat different magnitude) as our headline results.

Table 7: The proportion and number of offending firms (η) -–firms only

	(1)	(2)	(3)	(4)
	Proportion	Proportion	Number	Number
Before-after dummy	-1.725***	-0.769***	-1.383**	-0.467*
	(0.566)	(0.254)	(0.678)	(0.275)
Number of cases with fines (1 year lag)	-0.0448		0.0271	
	(0.181)		(0.182)	
Stock index	-0.614		-0.436	
	(0.886)		(0.859)	
Total assets	0.249		0.108	
	(0.156)		(0.201)	
Net operational costs	-1.345		-3.177	
	(2.351)		(2.500)	
Employees	1.218		1.966	
	(1.106)		(1.208)	
Employee costs	1.224		2.353	
	(2.290)		(2.367)	
Year	0.187		0.239**	
	(0.115)		(0.115)	
Observations	49	49	49	49

Standard errors in parentheses
=* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

To conclude this section, we summarise the findings of the above results in Table 8. Here we present the headline results for using various subsamples. The biggest such subsample is reporting offences (see Table 1) and there was enough data to allow us to estimate the above models for this subset (and the inverse of this subset, i.e. offences other than reporting). The table shows some variation in detection rates and in how the number of cases changes but all subsamples point to the same evidence of increasing deterrence.

Table 8: Summary results for various subsamples

	$\Delta\rho$	$\Delta\eta$	implied change in deterrence
All cases	+	-	+
Reporting only	0	-	+
Other than reporting:	+	0	+
Mis-selling	+	-	+
Fraud	0	-	+

7 Conclusions

Since the financial crisis there has been much reflection as to the effectiveness of financial regulation. The UK financial regulator in particular was candid as to its failings surrounding this crisis and areas where improvement could be effected (Ferran, 2011). Despite the importance of critically assessing regulatory performance, too much analysis has focused on deconstructing causes of past crisis events and often politically reactive regulatory developments. This study puts forward and applies a new method for assessing the efficacy of financial regulation, through assessing regulatory detection and deterrence rates to aid this assessment of misconduct regulation.

Our results indicate that while the number of detected cases did drop significantly, this was not a sign of weakening enforcement, but rather strengthening deterrence after 2010. The results were particularly driven by detection of fraud and misselling, rather than compliance offences. Beyond their policy relevance these findings also contribute to the long-standing discussion on the efficacy of regulation and optimal levels of regulation and punishment.

Our results raise a host of further questions as what might be driving this process. This study considers a period of time which witnessed increased punishments, changing regulatory structures, cultural change in the industry, and enhanced reputational damages due to increased media focus on misbehaving financial businesses; one, several, or all of these could have been influential. A deeper understanding of these candidate explanations is beyond the scope of this study yet remains an important and pressing area for future investigation, and would provide valuable insights for the growing literature identifying managerial (Koch-Bayram and Wernicke, 2018; Zorn et al., 2017) and cultural explanations (Parsons et al., 2018) of financial misconduct.

In order understand the effectiveness of regulatory action, it is vitally important to move beyond repetition of existing methods and to develop new techniques to refine estimations and disentangle alternative causality influences on financial misconduct. This study forwards a technique, which is less data demanding and emerges from a developed statistical tradition with ecology and biology. Moreover, when compared to previous attempts, our paper offers a more parsimonious approach to address partial observability issues—with clear implications for its practical implementation. We hope this contribution can act as a trigger for further work both examining levels of financial offending and other white collar crimes, and also provide support for the growing business and management literature seeking to comprehend and constrain such wrongdoing.

References

Allen, F. and Carletti, E. (2010). An overview of the crisis: Causes, consequences, and solutions. *International Review of Finance*, 10(1):1–26.

Amiram, D., Bozanic, Z., Cox, J. D., Dupont, Q., Karpoff, J. M., and Sloan, R. (2018). Financial reporting fraud and other forms of misconduct: a multidisciplinary review of the literature. *Review of Accounting Studies*, 23(2):732–783.

Amiram, D., Bozanic, Z., and Rouen, E. (2015). Financial statement errors: Evidence from the distributional properties of financial statement numbers. *Review of Accounting Studies*, 20(4):1540–1593.

Amstrup, S. C., McDonald, T. L., and Manly, B. F. (2005). *Handbook of capture-recapture analysis*. Princeton University Press.

Anderson, D., Burnham, K., and White, G. (1994). Aic model selection in overdispersed capture-recapture data. *Ecology*, 75(6):1780–1793.

Becker, G. S. and Stigler, G. J. (1974). Law enforcement, malfeasance, and compensation of enforcers. *The Journal of Legal Studies*, 3(1):1–18.

Bouchard, M. (2007). A capture–recapture model to estimate the size of criminal populations and the risks of detection in a marijuana cultivation industry. *Journal of Quantitative Criminology*, 23(3):221–241.

Bryant, P. G. and Eckard, E. W. (1991). Price fixing: the probability of getting caught. *The Review of Economics and Statistics*, 73(3):531–536.

Burdon, W. M. and Sorour, M. K. (2018). Institutional theory and evolution of 'a legitimate'compliance culture: The case of the uk financial service sector. *Journal of Business Ethics*, pages 1–34.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York, 2 edition.

Bushway, S. D., Paternoster, R., and Brame, R. (2003). Examining the prevalence of criminal desistance. *Criminology*, 41(2):423–448.

Cecchini, M., Aytug, H., Koehler, G. J., and Pathak, P. (2010). Detecting management fraud in public companies. *Management Science*, 56(7):1146–1160.

Collins, M. F. and Wilson, R. M. (1990). Automobile theft: Estimating the size of the criminal population. *Journal of Quantitative Criminology*, 6(4):395–409.

Danisewicz, P., McGowan, D., Onali, E., and Schaeck, K. (2018). The real effects of banking supervision: Evidence from enforcement actions. *Journal of Financial Intermediation*, 35:86–101.

de Batz, L. (2020). Financial crime spillovers. does one gain to be avenged? *Journal of Economic Behavior & Organization*, 173(C):196–215.

Dechow, P. M., Ge, W., Larson, C. R., and Sloan, R. G. (2011). Predicting material accounting misstatements. *Contemporary accounting research*, 28(1):17–82.

Delis, M. D., Staikouras, P. K., and Tsoumas, C. (2016). Formal enforcement actions and bank behavior. *Management Science*, 63(4):959–987.

Dyck, I., Morse, A., and Zingales, L. (2013). How pervasive is corporate fraud? *Rotman School of Management Working Paper 2222608*.

Feinstein, J. S. (1990). Detection controlled estimation. *The Journal of Law and Economics*, 33(1):233–276.

Ferran, E. (2011). The break-up of the financial services authority. *Oxford Journal of Legal Studies*, 31(3):455–480.

Foley, S., Karlsen, J. R., and Putniņš, T. J. (2019). Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies? *The Review of Financial Studies*, 32(5):1798–1853.

Goodhart, C. (2017). Has regulatory reform been misdirected? *Journal of Financial Regulation and Compliance*, 25(3):236–240.

Graham, J. R., Harvey, C. R., Popadak, J., and Rajgopal, S. (2017). Corporate culture: Evidence from the field. *National Bureau of Economic Research Working paper 23255.*

Group of 30 (2015). *Banking Conduct and Culture. A call for Sustained and Comprehensive Reform.* G30, Washington D.C., USA.

Hahn, P. R., Murray, J. S., and Manolopoulou, I. (2016). A bayesian partial identification approach to inferring the prevalence of accounting misconduct. *Journal of the American Statistical Association*, 111(513):14–26.

Koch-Bayram, I. F. and Wernicke, G. (2018). Drilled to obey? ex-military ceos and financial misconduct. *Strategic Management Journal*, 39(11):2943–2964.

Kuang, Y. F. and Lee, G. (2017). Corporate fraud and external social connectedness of independent directors. *Journal of Corporate Finance*, 45:401–427.

La Porta, R., Lopez-de Silanes, F., and Shleifer, A. (2006). What works in securities laws? *The Journal of Finance*, 61(1):1–32.

Lancaster, T. and Imbens, G. (1996). Case-control studies with contaminated controls. *Journal of Econometrics*, 71(1-2):145–160.

Leaver, M. P. and Reader, T. W. (2019). Safety culture in financial trading: an analysis of trading misconduct investigations. *Journal of Business Ethics*, 154(2):461–481.

Lebreton, J.-D., Burnham, K. P., Clobert, J., and Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs*, 62(1):67–118.

Marcel, J. J. and Cowen, A. P. (2014). Cleaning house or jumping ship? understanding board upheaval following financial fraud. *Strategic Management Journal*, 35(6):926–937.

McCraw, T. K. (1975). Regulation in america. *Business History Review*, 49(2):159–183.

Miller, N. H. (2009). Strategic leniency and cartel enforcement. *American Economic Review*, 99(3):750–68.

O'Donovan, J., Wagner, H. F., and Zeume, S. (2019). The value of offshore secrets – evidence from the panama papers. *The Review of Financial Studies*, Forthcoming.

Olsen, B., Munster, V. J., Wallensten, A., Waldenström, J., Osterhaus, A. D., and Fouchier, R. A. (2006). Global patterns of influenza a virus in wild birds. *Science*, 312(5772):384–388.

Ormosi, P. L. (2014). A tip of the iceberg? the probability of catching cartels. *Journal of Applied Econometrics*, 29(4):549–566.

Palermo, T., Power, M., and Ashby, S. (2017). Navigating institutional complexity: The production of risk culture in the financial sector. *Journal of Management Studies*, 54(2):154–181.

Parsons, C. A., Sulaeman, J., and Titman, S. (2018). The geography of financial misconduct. *The Journal of Finance*, 73(5):2087–2137.

Pasiouras, F. (2016). Financial consumer protection and the cost of financial intermediation: Evidence from advanced and developing economies. *Management Science*, 64(2):902–924.

Phillips, S. J. and Elith, J. (2013). On estimating probability of presence from use–availability or presence–background data. *Ecology*, 94(6):1409–1419.

Poirier, D. J. (1980). Partial observability in bivariate probit models. *Journal of Econometrics*, 12(2):209–217.

Pollock, K. H., Nichols, J. D., Brownie, C., and Hines, J. E. (1990). Statistical inference for capture-recapture experiments. *Wildlife Monographs*, pages 3–97.

Rossmo, D. K. and Routledge, R. (1990). Estimating the size of criminal populations. *Journal of Quantitative Criminology*, 6(3):293–314.

Roulet, T. J. (2019). Sins for some, virtues for others: Media coverage of investment banks' misconduct and adherence to professional norms during the financial crisis. *Human Relations*, 72(9):1436–1463.

Shi, W., Connelly, B. L., and Hoskisson, R. E. (2017). External corporate governance and financial fraud: Cognitive evaluation theory insights on agency theory prescriptions. *Strategic Management Journal*, 38(6):1268–1286.

Sobolev, D. (2019). Insider information: The ethicality of the high frequency trading industry. *British Journal of Management*.

Stigler, G. J. (1964). Public regulation of the securities markets. *The Journal of Business*, 37(2):117–142.

Stigler, G. J. (1970). The optimum enforcement of laws. *Journal of Political Economy*, 78(3):526–536.

Wang, T. Y. (2011). Corporate securities fraud: Insights from a new empirical framework. *The Journal of Law, Economics, & Organization*, 29(3):535–568.

Wang, T. Y., Winton, A., and Yu, X. (2010). Corporate fraud and business conditions: Evidence from ipos. *The Journal of Finance*, 65(6):2255–2292.

Williams, B. K., Nichols, J. D., and Conroy, M. J. (2002). *Analysis and Management of Animal Populations*. Academic Press.

Zakolyukina, A. A. (2018). How common are intentional gaap violations? estimates from a dynamic model. *Journal of Accounting Research*, 56(1):5–44.

Zeume, S. (2017). Bribes and firm value. *The Review of Financial Studies*, 30(5):1457–1489.

Zorn, M. L., Shropshire, C., Martin, J. A., Combs, J. G., and Ketchen Jr, D. J. (2017). Home alone: The effects of lone-insider boards on ceo pay, financial misconduct, and firm performance. *Strategic Management Journal*, 38(13):2623–2646.

Appendix

A Methodological appendix

A.1 Assumption for CR methods

Below we provide an overview of the assumptions required for unbiased CR estimates, and their suitability for the analysis of business behaviour.

A1. Discrete sampling occasions: *The financial regulator engages in market monitoring (CR sampling) in discrete annual periods $t = 1, 2, ...$, and the population of financial offenders does not change during sampling occasions, but can change between sampling occasions.*

This assumption treats each year as one sampling occasion and the parameter estimates are therefore annual capture and survival estimates. The use of Cormack-Jolly-Seber (CJS) CR methods assumes that samples are taken instantaneously.[8] In practice however this assumption is nearly always necessarily violated and we have to use discrete sampling. In order for this violation not to cause bias, Assumption A1 is needed, which requires that within the sampling period (i.e. within each analysed year) there is no change in the analysed population. To illustrate the importance of this assumption, imagine that financial offending survival is analysed. This assumption means that the survival (i.e. to remain capturable in the future) of an offending firm to the next period is the same for a firm that was captured in January as a firm that was captured in December.[9]

A2a. Homogeneity: *The probability of any firm $m = 1, 2, ..., n$ being captured by the financial regulator at sampling occasion t is given by ρ_t (provided that it had been captured at least once and that it had survived until t).*

[8] Lebreton et al. (1992)

[9] This issue of long sampling times has been discussed by Williams et al. (2002). Olsen et al. (2006) uses simulation data to show the bias caused by lengthy sampling periods. Ormosi (2014) showed that this was not significantly biasing the results when using annual sampling of cartelising businesses.

Table A1: Partial observability methods compared

Paper	Method	Unit of analysis	Data source	Scope of data required	Relaxes independence assumption	Relaxes stationarity assumption	Allows heterogenous firms	Allows movement between states (offend - not offend)
Miller (2009)	Markov chain	Industry groups	public	narrow	No	No	No	No
Bryant and Eckard (1991)	Duration of offending	Firm	public	narrow	No	No	No	No
Tan et al. (2015)	Heckman selection models	Firm	public	extensive	No	No	Yes	No
Poirier (1980), Feinstein (1990), Wang (2011)	Detection controlled estimation/ Bivariate probit model	Firm	public	extensive	No	No	Yes	No
Dyck et al. (2013)	Experimental methods	Firm	non public	extensive	No	No	No	No
Dechow et al. (2011)	Descriptive	Industry groups	public	extensive	No	No	No	No
Amiram et al. (2015)	Benfords Law	Industry groups	public	extensive	No	Yes	Yes	No
Cecchini et al. (2010)	Machine Learning	Firm	public	extensive	No	Yes	No	Yes
Foley et al. (2019)	Network approaches	Industry groups	non public	extensive	Yes	Yes	Yes	Yes
O'Donovan et al. (2019)	Event study	Industry groups	non public	extensive	No	No	Yes	No
Ormosi (2014)	Capture - recapture	Firm	public	narrow	No	Yes	Yes	Yes

A2b. *Any firm* $m = 1, 2, ..., n$ *surviving sampling occasion* t *has equal probability* ϕ_{tm} *of survival to* $t + 1$*.*

As the proposed model only provides estimates for captured firms, the homogeneity assumption is reduced to all marked offenders having the same capture/survival probability (and not that marked and unmarked offenders have equal capture and survival probabilities). Assumptions A2 and A3 also imply time-dependence of the parameters, which relaxes on the stationarity assumptions used in previous literature that looked at the partial observability problem. A test for time-dependence will be conducted before the empirical estimation, where models with time-dependent and constant parameters will be compared.

In practice, the homogeneity assumption is rarely satisfied (temporary and/or permanent heterogeneity). The simplest way of relaxing this assumption would be to acknowledge heterogeneity, and interpret the estimated parameters as a UK aggregate for all marked offenders. However, an appealing feature of modern open population CR methods is that we can go beyond this and control for differences between the individual offender. Two main sources of heterogeneity are 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 treated as permanent (marked offenders showing different capture/survival rates to the ones never captured), or temporary (within the marked sub-population, parameters directly following capture are systematically different). Pollock et al. (1990) pointed out that when using the Cormack Jolly Seber model, survival and capture parameters are based on marked individuals and are therefore not affected by permanent trap-response. In our model we test temporary trap-response by estimating a model that allows 1-year trap dependence. Depending on whether the model is a time-dependent or a constant one, there are numerous possible model specifications. For example, the likelihood function of a model with constant and temporary (1 year) trap-dependent survival rate is given below. Here the survival rate is constant across time periods but for each individual there is a difference between the year directly following capture (ϕ_{tm}) and all subsequent years (ϕ) (note that in this case we only estimate two survival parameters) and time-dependent capture probability is:

$$L = \prod_{m=1}^{2^K-1} \sum_{d_m=l_m}^{K} \left\{ \phi_{fm} l_m \phi(K-d)(1-\phi) \times \left(\prod_{t=f_m+1}^{d_m} p_t^{x_{mt}} (1-p_t)^{(1-x_{mt})} \right) \right\} \tag{6}$$

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 some characteristics. Given our sample size, we cannot apply stratified models (for dimensionality issues as we would need different estimates for each stratum) and therefore we need to rely on our assumption of homogeneity.

A4. Independence: *There is independence between the individual offenders with respect to capture and survival (independence is only needed for the marked subpopulation).*

The violation of independence may produce an overestimate of variances, and may produce biased estimates, however there is little evidence to support the latter (Anderson et al., 1994). There is a potential source of bias given that offenders involved in the same offence are not independent from each other (for example if the same regulatory action discovers more than one offender), but such co-offending or co-discovery is rare in our sample.

A more important potential violation is that businesses are aware of previous captures and might adjust their behaviour in response. This would mean that our subsequent annual samplings would contain a continuously evolving set of offenders. As we estimate annual rates of detection and survival, this change should be picked up by the changing level of estimates. If however a behavioural change happens within samplings (within a calendar year), then we have a violation that are currently unable to deal with. The reason we are relaxed about this is because our main focus is not on the precise magnitude of detection and survival rates, rather the testing of whether there was a structural change in the rate of detection. Therefore our estimates should be reliable as long as the violation of the independence assumption is also time-dependent.

A5. Study area: *The whole geographical area of study is sampled with equal intensity. If new areas were 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 financial regulations are continuously changing and therefore it is possible that some behaviour only became illegal halfway through our study period. Table A2 below shows that it should not have affected our sample as the proportion of various offending types are roughly constant across the two study periods of our interest (before and after Jan 2011).

Table A2: Proportion of case types, before and after Jan 2011

	Reporting	Compliance	Market Abuse	Fraud	Mis-selling	Money Laundering	Other
Before 2011	0.562	0.033	0.061	0.207	0.123	0.007	0.007
After 2010	0.57	0.028	0.086	0.138	0.147	0.014	0.016

A6. Marked individuals do not lose their marks.

Although this assumption is typically more relevant to ecological studies where animals are physically marked, one issue may arise in relation to financial offenders. Firms may change their name during the period of analysis (e.g. as a result of mergers). This was accounted for when data was collected and cleaned for the empirical analysis, where all offending firms were cross-referenced against each other across a number of parameters including address, ownership, and employees (based upon the publicly available register of financial firms). As all regulated firms and individuals have a unique identifier code attributed by the Financial Services Register, these firm-ownership changes are accurately tracked.

A.2 Choosing the right model

To find the best fitting model, we estimated 11 different models (these are different parametrisations of Equation 2). We choose the most efficient one using Akaike's Information Criterion.[10] The test statistics are presented in Table 3, where AICc is the corrected AIC, Delta AIC is the difference in comparison to the model with the lowest AIC.

For the model names we use the following rules: ϕ and ρ denote survival and detection probabilities respectively. The notation (.) implies a model where the given parameter is assumed to be time-constant; (t) indicates time-dependent parameters. (.)(.) denotes that our estimated parameter is estimated for both the period before and after January 2011, and that it is constant across all years within our before and after intervals. (./.) refers to a model where the given parameter is time-constant within its interval, but its values are allowed to differ between the year of detection and any other subsequent year (trap-dependence). For example, $\phi(./.)(./.)$ refers to a model where we assume that firms' survival rate in the year of the detection is different from all subsequent years, and we estimated this for before and after 2011.[11]

[10]The AIC is given by: $AIC = -2\ln(L) + 2K$, where L is the model likelihood, and K is the number of parameters estimated. An unbiased, corrected version of AIC was given by Hurvich and Tsai (1989): $AICc = -2\ln(L) + 2K(n/(n - K - 1))$.

[11]These widely accepted notations are also used by the software Mark, used for our estimation.

Table A3: AIC statistics and model ranking

Model	AICc	Delta AICc	AICc Weights	Num. Par
$\phi(./.)(./.)\ \rho(./.)(./.)$	1491.124	0	0.52611	8
$\phi(./.)(./.)\ \rho((.)(.))$	1491.561	0.4362	0.42302	6
$\phi(./.)(./.)\ \rho(./.)$	1496.051	4.9267	0.0448	6
$\phi(./.)\ \rho(./.)$	1500.385	9.2608	0.00513	4
$\phi(./.)\ \rho(./.)(./.)$	1503.776	12.6518	0.00094	6
$\phi(t)\ \rho(t)$	1583.325	92.2008	0	22
$\phi(.)(.)\ \rho(./.)(./.)$	1588.558	97.4332	0	6
$\phi(.)(.)\ \rho(.)(.)$	1593.878	102.7531	0	4
$\phi(t)\ \rho(.)$	1597.917	106.7924	0	11
$\phi(.)\ \rho(t)$	1600.056	108.9314	0	15
$\phi(.)\ \rho(.)$	1601.596	110.4718	0	2

B Additional tables

Table B1: Variables considered

Variable name	Description	Units	Data source
Capture	Whether a firm has received a Final Notice in the sample period.	A dummy variable recorded in each year of the sample	FSA/FCA Final Notices.
Market	The financial market in which the firm primarily operates. This definition focuses on the firms regulated activities	One of eight classifications including banking, consumer credit, insurance, investments, payments, stockbroker/ asset management/corporate finance /hedge funds and not know	FSA/FCA Final Notices.
Type of offence	The classification of the offence. These are not mutually exclusive.	One of six classifications including market abuse, fraud and theft, mis-selling, reporting and compliance, money laundering and complaints handling.	FSA/FCA Final Notices.
Offence duration	The 'relevant time period' of the offending as defined within the FSA/FCA Final Notice. Alternatively, the time between the first period of offending and the end of the offending.	Days	FSA/FCA Final Notices.
Punishment	The outcome of the Final Notice. Multiple outcomes are commonly reported.	One of six punishments including public censure, fines individual prohibition, variation of regulatory permissions, disgorgement of profits and other measures.	FSA/FCA Final Notices.
Regulated firm numbers	The number of regulated financial firms	Number of financial firms	FSA/FCA. Annual Reports and Accounts
Regulator Net Assets	The assets net of liabilities to provide a perspective on regulators resources	(£m or equivalent)	FSA/FCA. Annual Reports and Accounts
Regulator net operational costs	Net operational costs	(£m or equivalent)	FSA/FCA. Annual Reports and Accounts
Regulator employees	The regulatory workforce size	Number of employees	FSA/FCA. Annual Reports and Accounts
Regulator Employee costs	The costs of the regulatory workforce.	Total employment costs of the regulator.	FSA/FCA. Annual Reports and Accounts
Main Stock Market Index Change %	Change in the appropriate stock market.	Change in the FTSE 100	Market websites.

Table B2: Descriptive statistics of conduct offences in the UK

Final Notice/ order year issued	Cases	Average duration of offence (days)	Average fine £
2002	15	747.8	913,000
2003	48	635	573,750
2004	89	639.33	804,516
2005	48	505.39	1,045,366
2006	203	335	483,044
2007	144	514.26	242,159
2008	218	612.65	431,731
2009	182	738.3	816,587
2010	251	546.55	1,050,650
2011	145	676.44	1,168,463
2012	161	692.83	5,431,582
2013	138	529.36	11,672,332
2014	113	566.83	33,489,179
2015	114	813.44	21,559,410
2016	182	562.41	12,77,635

Table B3: Firm level offending and reoffending

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Offending firms															
First offence	14	37	57	36	186	128	168	125	183	96	114	96	78	61	151
Second Offence	0	1	2	4	5	2	4	8	19	14	14	18	22	21	13
Total offences	14	38	59	40	191	130	172	133	202	110	128	114	100	82	164
Total regulated firms															
Number	41,791	42,901	53,830	53,172	53,375	54,346	55,182	56,000	57,058	58,918	60,991	66,870	84,596	n/a	n/a
%	0.03	0.09	0.11	0.08	0.36	0.24	0.31	0.24	0.35	0.19	0.21	0.17	0.12	n/a	n/a

Table B4: Covariates Descriptive statistics

UK	Main Stock Market Index Change (%)	Regulator Net Assets	Regulator net operational costs	Regulator employees	Regulator Employee costs
2003	11.66	3.30	208.26	2288.00	198.30
2004	6.74	17.80	224.70	2312.00	119.30
2005	15.92	22.60	246.30	2356.00	158.30
2006	9.49	21.60	272.20	2610.00	196.50
2007	2.31	13.70	263.70	2659.00	199.90
2008	-30.90	6.80	304.70	2535.00	197.80
2009	18.66	-19.50	346.50	2730.00	208.60
2010	-1.59	17.10	384.30	3150.00	269.10
2011	5.88	36.80	450.80	3337.00	293.10
2012	3.47	58.00	474.70	3502.00	314.00
2013	11.97	40.00	528.20	3596.00	326.90
2014	-2.26	22.70	434.50	2511.00	278.80
2015	-4.67	-17.10	452.70	3155.00	337.00
2016	17.22	-23.10	479.00	3285.00	307.80

Table B5: Example of a section of our capture history matrix

Firm	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Sedley Richard Laurence V.	0	0	0	1	0	0	0	0	0	0
Greenhow & Company	0	0	0	0	0	0	0	0	0	0
MF Global UK Limited	0	0	0	0	1	0	0	0	0	0
Abbey National plc	0	0	0	0	0	1	0	1	0	0
Nationwide Building Society	1	0	0	0	0	0	0	0	0	0
Northern Rock Plc	0	0	0	1	0	0	0	0	0	0
Yorkshire Building Society	0	0	0	0	0	0	0	1	0	0
Bradford & Bingley plc	0	0	0	0	0	0	1	0	0	0
Royal Liver Assurance Limited	0	0	0	0	0	1	0	0	0	0
Guardian Linked Life A. Ltd	0	0	0	0	0	0	0	0	0	0