

How much insider trading happens in stock markets?*

Vinay Patel^a and Tālis J. Putniņš^{a,b}

^a *University of Technology Sydney*

^b *Stockholm School of Economics in Riga*

December 30, 2021

Abstract

We estimate that the actual prevalence of illegal insider trading is at least four times greater than the number of prosecutions. Using structural estimation methods that account for incomplete and non-random detection and hand-collected data of all US prosecuted insider trading cases, we estimate that insider trading occurs in one in five mergers and acquisitions and in one in 20 earnings announcements. Key drivers of the propensity for insider trading include the value of the information, and the number of people in possession of the information, and the stock's liquidity. Detection and prosecution are more likely when there are abnormal trading patterns and more regulatory resourcing.

Keywords: insider trading, prosecution, detection-controlled estimation, M&A, earnings

JEL classification: G14

* Email: Vinay.Patel@uts.edu.au and Talis.Putnins@uts.edu.au. We thank Joseph Barbara, Phillip Drummond, David Easley, Sean Foley, Kingsley Fong, Neal Galpin, Petko Kalev, Maureen O'Hara, Emiliano Pagnotta, Elvira Sojli, David Walsh, Terry Walter, and participants at the Financial Research Network Conference, and seminar participants at Monash University, University of Technology Sydney, and La Trobe University for helpful comments.

1. Introduction

“I have a major present for you... information.”

— Marketing officer at Akamai Technologies talking to a hedge fund trader (Danielle Chiesi).¹

Despite the significant resources devoted to combatting illegal insider trading, it remains pervasive, although no one knows quite how pervasive.² A lower bound on its prevalence is given by prosecution cases—the US Securities and Exchange Commission (SEC) prosecutes approximately 50 insider trading cases per year.³ Although most would agree that prosecution cases reflect only a fraction of all illegal insider trading, opinions are vastly varied about the total amount of illegal insider trading and how this rate varies through time. While some regulators downplay the prevalence of undetected insider trading as being negligible, at the other end of the spectrum, some observers argue that insider trading is “*rampant*”, has been occurring for a long period of time, and involves many individuals, from hedge funds to corporate insiders.⁴ It is therefore unclear how much illegal insider trading occurs and what fraction is detected and prosecuted by regulators. In this paper we aim to answer these questions.

A further issue we seek to address is when, and in which stocks, is insider trading more likely? And what makes a given instance of insider trading more likely to be detected and prosecuted? While such questions are important in understanding insider trading, its impact on markets, and the effectiveness of regulators, they are difficult to answer because insider trading prosecutions are a *non-random* sample of insider trading. If the detection and prosecution of insider trading is anything like that of market manipulation, the observed prosecutions might be only the tip of the iceberg and can lead to substantial biases in inference due to the non-random sample.⁵

We tackle these issues using hand collected data on all US prosecuted cases of insider trading during the past 21 years and *detection-controlled estimation* (DCE) models. These structural estimation techniques explicitly address the fact that prosecutions are a non-random sample of all insider trading. DCE models have been applied in other settings involving incomplete detection, such as breaches of occupational health and safety (Feinstein, 1989), tax evasion (Feinstein, 1991), corporate fraud (Wang, Winton, and Yu, 2010),

¹ See <https://www.ft.com/content/d2f32724-7bfe-11e0-9b16-00144feabdc0>.

² For example, the US Securities and Exchange Commission anticipates using resources of \$543 million and 1,504 staff in the Enforcement Division during 2017 (SEC, 2017). These numbers do not count resources devoted to other divisions or to State Prosecutors.

³ See <https://www.sec.gov/news/newsroom/images/enfstats.pdf>.

⁴ Preet Bharara (US Attorney for the Southern District), as quoted in *Frontline*, “Preet Bharara: Insider trading is ‘rampant’ on Wall Street” (January 7, 2014). As an example of the former, John Shad (SEC Chairman) suggests that securities fraud accounts for a fraction of 1% of all securities transactions, see *The New York Times*, “No new insider laws are needed, Shad says” (June 19, 1986).

⁵ For example, Comerton-Forde and Putniņš (2014) find that only one in three hundred cases of closing price manipulation in US and Canadian stock markets is detected and prosecuted.

market manipulation (Comerton-Forde and Putniņš, 2014), and most recently illegal activity in cryptocurrencies (Foley, Karlson, and Putniņš, 2019). DCE involves *jointly* estimating a model of the violation process (in this case insider trading) and the detection process, thereby obtaining unbiased estimates of the drivers and outcomes of each process. This approach provides estimates of the amount of insider trading, the fraction that is detected, the time-series patterns in insider trading and detection, the characteristics of stocks and information that make them more susceptible to insider trading, and the characteristics that make instances of insider trading more likely to be detected/prosecuted.

We focus on insider trading in US stocks around mergers and acquisitions (M&As) and quarterly earnings announcements between 1996 and 2016 because they are the most frequent of the major price sensitive announcements made by companies. They are also the most frequent information types in prosecution cases and therefore provide the richest information for estimating insider trading rates and characteristics. For example, approximately 70% of US SEC and Department of Justice (DOJ) insider trading prosecutions relate to M&As and earnings (50% and 20%, respectively). Earnings announcements are regular and scheduled, whereas M&As are irregular and unscheduled, allowing us to contrast insider trading in scheduled and unscheduled news.

Our first main finding is that insider trading occurs in an estimated one in five M&A events and one in 20 quarterly earnings announcements. These estimates imply that insider trading is more prevalent in M&A events than earnings announcements, even after accounting for differences in detection rates. These estimates also imply that there is at least four times more actual insider trading than there are prosecution cases. We estimate that the probability of detection/prosecution of insider trading in both M&A and earnings announcements is approximately 15%. Therefore, what we see in prosecutions is just the tip of the iceberg.

The notion that only a minority of actual insider trading violations (less than 20%) are detected and prosecuted is consistent with theories of rational crime such as the literature following the Becker (1968) framework. Given the substantial penalties for convicted insider trading violations including financial, reputational, and potential jail time, and the smaller potential profits from insider trading, the detection/prosecution probability would have to be low otherwise no rational agent would engage in insider trading. Of course, behavioral biases such as overconfidence, cognitive dissonance, and incorrect beliefs about detection probabilities may also play a role in explaining the levels of insider trading and nothing in our results precludes such effects.

Our second main finding is that the detection likelihood increases throughout the two decades in our sample period. This trend is strongest for earnings announcements, with the detection probability reaching as high as 30% in 2016. The trend coincides with increases in regulatory resources through time and the introduction of the SEC Whistleblower Program in 2010. For both announcement types, the probability of

insider trading increases from 1996 to 2007 but falls following the introduction of the Whistleblower Program in 2010. Our analysis therefore suggests that this program has been successful in both increasing the detection rate and decreasing the prevalence of insider trading consistent with the theoretical predictions of Kacperczyk and Pagnotta (2020).⁶

We also use the model to identify the characteristics that affect the probability of insider trading. For both unscheduled and scheduled announcements, we find that the probability of insider trading is higher for more liquid stocks, consistent with the notion that liquidity helps an insider trade strategically and hide their informational advantage (e.g., Kyle, 1985; Admati and Pfleiderer, 1988; Collin-Dufresne and Fos, 2015, 2016), ultimately leading to larger trading profits. We find that illegal trading is more likely when the value of inside information is higher, for example, when there is a larger return on the M&A/earnings announcement date. This result is consistent with rational crime theories where the propensity for misconduct increases with the potential profits and similarly in Kyle (1985) informed traders tend to trade more when there is a larger difference between observed stock prices and fundamental value. Additionally, we find that the likelihood of insider trading increases when the opportunities for information leakage are larger, for example, from an increased number of M&A financial advisors, similar to theories of competing informed traders (e.g., Holden and Subrahmanyam 1992; Back et al., 2000).

We find that the probability of detection and prosecution increases with the amount of regulatory resources devoted to enforcing insider trading, as captured by regulatory budget and future prosecution rate. Regulators are also more likely to detect cases of illegal trading when there are abnormal pre-announcement spikes in returns and volumes prior to M&A and earnings announcements, consistent with the optimal enforcement model of DeMarzo et al. (1998) where abnormal trading volume is used to identify insider trades.

The results are robust to alternate model identification assumptions such as different sets of instrumental variables. We also use two independent validation tests. The first test uses the number of informal inquiries to validate our estimates for the probability of insider trading. The second test compares the pre-announcement abnormal returns and trading volume of announcements with a high estimated likelihood of insider trading against those with a low estimated probability. We find differences in trading characteristics consistent with the expected impact of illegal insider trading, providing support for the empirical probabilities of insider trading.

Our paper contributes to several areas of the literature. First is the literature on the benefits and costs of insider trading and how insider trading impacts markets. While insider trading can improve the

⁶ Based on the Becker (1968) and Kyle (1985) frameworks, Kacperczyk and Pagnotta (2020) develop a model that predicts increases in the probability of detection/prosecution (e.g., insider trading detected via the SEC Whistleblower Program) results in more cautious and reduced insider trading.

informational efficiency of prices (e.g., Carlton and Fischel, 1983; Fernandes and Ferreira, 2009), insider trading is illegal in most countries and punishable by civil/criminal sanctions.⁷ In addition to creating significant regulatory and surveillance costs, insider trading also harms investor confidence and the perceived fairness of markets, which can reduce investor participation, thereby decreasing market liquidity and increasing trading costs (Leland, 1992). Insider trading can also have adverse effects on a firm's cost of capital (Bhattacharya and Daouk, 2002). Determining the appropriate regulatory response to these costs and benefits of insider trading requires an understanding of how effectively insider trading can be detected and prosecuted and how it responds to surveillance. We provide the first such estimates. By estimating the prevalence and determinants of insider trading, this paper increases our understanding of where regulation/surveillance effort can have the largest impact (e.g., in specific industries or certain types of price sensitive information).

Second, a debate has emerged in the literature as to whether standard liquidity and information asymmetry measures suggested by market microstructure models can capture or “detect” the presence of informed trading (e.g., Collin-Dufresne and Fos, 2015). Some of the recent empirical contributions to this debate rely on prosecuted instances of insider trading. For example, in contrast to microstructure theory, Kacperczyk and Pagnotta (2019) find a largely negative relation between a battery of standard illiquidity measures (e.g., quoted spreads, price impact, and so forth) and insider trading in stock and options markets. By conditioning on short-lived information, Ahem (2020) reaches the opposite conclusion, documenting a positive relation between illiquidity measures and insider trading. However, after controlling for involvement by Financial Industry Regulatory Authority (FINRA) as a proxy for whether the case was detected via abnormal trading, only order imbalance can detect the presence of informed trading. Further, unlike order-flow measures, Akey et al. (2021) find that when insiders cannot strategically time their trades, spread and volume measures can detect insider trading.

These conflicting results illustrate the significant sample selection issues in prosecuted insider trading cases. Using prosecuted cases in such tests is problematic because the prosecuted cases reflect the characteristics of the insider trading *and* the detection/prosecution processes. For example, if detection or prosecution of insider trading is more likely when there is a large price run-up before the announcement, prosecuted insider trading cases will appear to have large price impacts even if insider trading actually has little or no price impact. Our findings contribute to this literature by quantifying the extent to which prosecutions underestimate the actual rates of insider trading and characterizing how prosecuted cases differ from other cases that are not prosecuted.

⁷ For example, penalties in the US include a maximum individual fine of \$5,000,000 and maximum prison sentence of 20 years (see https://www.sec.gov/Archives/edgar/data/25743/000138713113000737/ex14_02.htm).

Third, several studies examine the characteristics of insider trading in stock markets using SEC/DOJ prosecutions. Meulbroek (1992) and Meulbroek and Hart (1997) find that detected insider trades tend to occur on days with larger stock returns and detected cases have larger M&A bid premia. Ahern (2017) examines information sharing through insiders' social networks. Kacperczyk and Pagnotta (2019, 2020) examine trading strategies of detected insiders and how they respond to shocks in legal risk. Detected insiders trade faster in the face of competition from other traders, are more inclined to split their trades in smaller stocks, and they trade less aggressively when there is more risk of being caught. Our paper contributes to this literature by shedding light on the determinants of insider trading and its prosecution, the underlying rates of violation and detection, and explicitly addressing the issue of sample selection bias.

The prevalence of insider trading has been examined in options markets using indirect proxies. For example, Augustin et al. (2019) find abnormal options trading volume prior to 25% of their sample of M&As suggesting that informed trading in options markets is pervasive. This figure could overstate the amount of insider trading due to rumors and merger anticipation or understate it because not all insider trading involves options. Using a theoretical framework, Augustin et al. (2016) characterize how informed investors trade in the options market ahead of corporate news when they receive private, but noisy, information about the timing of the announcement and its impact on stock prices. They use SEC prosecutions of insider trading in options to validate their model's predictions. In contrast, our paper does not limit the sample to stocks with listed options (a minority of the stock universe). Our analysis also differs in that we use prosecution cases and structural estimation of insider trading and detection jointly rather than rely on price and volume proxies to estimate the prevalence of insider trading.⁸

Our results have implications for regulators and market participants. By identifying the determinants of insider trading and how it responds to increased regulatory effort, regulators can use our findings to tailor surveillance and enforcement effort to where it will have greatest impact and thereby improve detection and deterrence. For example, by identifying when and where insider trading is most likely, our findings can assist regulators in detecting more insider trading cases for a given amount of resources or reduce their costs in detecting the same number of cases. This in turn can have a deterrence effect on potential insider traders that weigh up the profits from insider trading against the expected costs of being detected and prosecuted.

This paper proceeds as follows. Section 2 describes the data sources including our hand-collected dataset of illegal insider trading. Section 3 details the econometric challenge and our solution: detection-controlled estimation (DCE). Section 4 reports the results. Section 5 concludes.

⁸ Many studies use price and volume measures to indirectly capture insider trading in options markets (e.g., Amin and Lee, 1997; Cao et al., 2005; Chan et al., 2015; Augustin et al., 2016; Hao, 2016; Augustin et al., 2019). A limitation of indirect proxies is that they are susceptible to false positives and negatives (e.g., insider trading that does not cause a large price change, or large price changes that occur due to reasons other than insider trading).

2. Data

We hand collect all prosecuted insider trading cases in the US from US SEC civil litigation cases and DOJ criminal litigation cases filed between January 1, 2000 and December 31, 2017. We obtain the records relating to each case from the SEC's litigation repository and additional information including court records from the US PACER service and Administrative Office of the US Courts.⁹ From the records of each case we manually extract a number of variables, including: the details of the corporate announcement, information that was the subject of insider trading, date, time, stock, type of announcement (M&A, earnings, etc.), nature of information (good news, bad news), and so on. We also extract as much information as possible about the actual illegal insider trades, including dates, times, volumes, direction, prices, and so on, as well as information about the accused person (name, age, occupation, etc.).

The first and last corporate announcement among these filings determines the beginning and end of our sample period. Thus, our sample of 453 insider trading cases spans the 21-year period 1996 to 2016. We restrict our sample to the two main types of information used in illegal insider trading cases: 365 M&A and 88 earnings announcements as these two types account for approximately 70% of the prosecuted insider trading cases, and therefore provide the richest information for estimating insider trading rates and characteristics.

We supplement the data about the illegal insider trading cases with a variety of information about penalties and regulatory enforcement effort. From the prosecution cases, we record the penalty incurred and outcome of the litigation. We also hand collect from the SEC Statement of Budgetary Resources and Annual Reports information regarding budgetary resources, spending authority, program/operational costs, number of staff, and number of informal inquiries/civil cases.

We obtain daily price and trading activity information for US common stocks listed on NYSE, AMEX and NASDAQ (i.e., share code = 10, 11 and exchange code = 1, 2, 3) from *CRSP*. Furthermore, we obtain information about all M&As involving US common stocks during our sample period from *SDC Platinum*, and all quarterly earnings announcements from *Compustat*. We merge all these data sources by stock-day. This step results in a dataset of M&A and earnings announcements for US common stocks between 1996–2016 with information about each announcement, the trading activity around each announcement, and importantly, information about which announcements involved illegal insider trading that was subsequently prosecuted.

⁹ SEC litigation releases are obtained from: www.sec.gov/litigation/litreleases.shtml.

The challenge is going from the observed prosecuted insider trading cases to inference about the entire population of illegal insider trading—prosecuted and *not* prosecuted. We describe our approach to dealing with this challenge in the next section.

3. Detection-controlled estimation model

3.1 The econometric challenge and its solution

A key challenge in analyzing illegal insider trading is that the directly observable cases, those that are prosecuted, represent only a non-random subset of all illegal insider trading. If insider trading prosecution rates are anything like those for market manipulation, then the prosecution cases are only the tip of the iceberg. While it is clear how the incomplete detection/prosecution makes it difficult to infer the prevalence of insider trading from prosecution cases, another important problem is that the characteristics of the prosecuted cases might not be representative of all illegal insider trading because prosecution is non-random.

The issue of incomplete and non-random detection of an underlying phenomenon is not unique to insider trading. It occurs for other crimes such as tax evasion, fraud and market manipulation, and detection settings such as nuclear power plant violations, mammogram screening for cancer, and so on. Although the challenges posed by incomplete detection might appear as just another case of sample selection bias, the problem is in fact more complicated than many other selection bias settings such as analyzing labor market outcomes where the bias comes from non-random participation in the labor force, or analyzing survey data where non-random non-response causes a bias. In such settings, the cause of the bias, the non-respondents or non-participants can be observed and therefore standard tools such as Heckman selection models can be applied to correct the bias. In the first stage of such a model, the response or participation decision is modeled. Then in the second stage, the model of the underlying phenomenon is adjusted to account for the non-randomness of the sample using estimates from the first stage.

In settings of incomplete detection, where the cause of the bias (the undetected cases) cannot be directly observed, Heckman models cannot be applied. However, there is an analogous class of models known as detection-controlled estimation (DCE) that are suitable for incomplete detection settings and serve the same purpose as Heckman selection models. Like Heckman models, DCE also involves estimating two models to overcome the selection bias—one model of the underlying phenomenon (in our case, illegal insider trading) and one model of the selection process (in our case the detection/prosecution of illegal insider trading). Just like a Heckman model, explicitly modeling the selection process in DCE allows the underlying phenomenon and its characteristics to be modeled free of selection bias. Unlike Heckman models, the two stages in the model cannot be estimated sequentially but instead have to be estimated

jointly. DCE models are therefore usually estimated with maximum likelihood estimation rather than the stage-by-stage OLS. Finally, while both Heckman models and DCE models can be estimated without instrumental variables, in which case the models rely on functional form and distributional assumptions, both models tend to be more robust if they include instruments that help separately identify the two stages in the model.

DCE models were developed by Feinstein (1990) and have since been successfully applied to a number of incomplete detection settings such as breaches of occupational health and safety (Feinstein, 1989), tax evasion (Feinstein, 1991), corporate fraud (Wang et al., 2010), market manipulation (Comerton-Forde and Putniņš, 2014), and most recently illegal activity in crypto-currencies (Foley et al., 2019). A further advantage of DCE models is that they not only correct for the selection bias caused by incomplete detection, but they can also be used to estimate the prevalence of the underlying phenomenon of interest. In our case, that means obtaining estimates of how much illegal insider trading occurs and what fraction is detected and prosecuted by regulators.

3.2. The model

We use a DCE model that consists of two equations—one that models the probability of insider trading for a particular company announcement and the second modeling the probability that conditional on illegal insider trading having occurred, it is detected and prosecuted.

< Figure 1 here >

Figure 1 illustrates the two stages of the DCE model which are estimated simultaneously. The unit of observation is an M&A or earnings announcement indexed by i . We estimate separate DCE models for these two types of announcements as they might differ in their propensity for insider trading and other characteristics. The first-stage equation models whether illegal insider trading occurred before announcement i . This model has an unobservable binary outcome (L_{1i}), which is driven by a continuous latent function (Y_{1i}) of a vector of characteristics (x_{1i}) that affect the probability of insider trading:

$$Y_{1i} = \beta_1 x_{1i} + \epsilon_{1i} \quad (1)$$

$$L_{1i} = \begin{cases} 1 & (\text{insider trading}) \\ 0 & (\text{no insider trading}) \end{cases} \text{ if } \begin{cases} Y_{1i} > 0 \\ Y_{1i} \leq 0 \end{cases} \quad (2)$$

The second stage models whether conditional on insider trading having occurred in the first stage, the insider trading is detected and prosecuted. This process also has a binary outcome (L_{2i}) driven by a continuous latent function (Y_{2i}) of a vector of characteristics (x_{2i}) that affect the probability that an instance of insider trading is detected/prosecuted:

$$Y_{2i} = \beta_2 x_{2i} + \epsilon_{2i} \quad (3)$$

$$L_{2i} = \begin{cases} 1 & (\text{detected/prosecuted}) \\ 0 & (\text{not detected/prosecuted}) \end{cases} \text{ if } \begin{cases} Y_{2i} > 0 \\ Y_{2i} \leq 0 \end{cases} \quad (4)$$

While the two processes are not separately observed their joint outcomes are observed. Set A consists of insider trading that has been detected and prosecuted ($L_{1i} = L_{2i} = 1$), and set A^c consists of announcements for which there was no insider trading ($L_{1i} = 0$) or the insider trading was not detected/prosecuted ($L_{2i} = 0$).

We estimate the model parameters (β_1 and β_2) using maximum likelihood estimation. These parameter estimates characterize the drivers of insider trading and what makes insider trading more/less likely to be detected/prosecuted. Appendix A contains the equations and likelihood function for the model. The estimated model parameters also allow us to calculate the probability of insider trading and the conditional probability of detection (see probability matrix in Table A1 of Appendix A).

We estimate the probability of insider trading and its detection (and associated confidence intervals) using a bootstrap method involving 1,000 samples. For each bootstrap sample, we estimate the DCE model using a random sample of announcements that is reduced in size for computational feasibility. Each random sample of announcements consists of a 10%/90% split between all announcements in which insider trading is detected, and a sample of announcements in which insider trading is not detected or did not occur, respectively. Using the DCE model parameters (β_1 and β_2), we estimate the probability of insider trading and its detection for each corporate announcement. We report the mean probability of insider trading (a measure of its expected prevalence) and the mean conditional probability of detection across the 1,000 bootstrap samples. Using a bootstrap of 1,000 samples, we obtain standard errors, from which we estimate confidence intervals.

3.3. Instrumental and control variables

While the structural DCE model can be identified based on functional and distributional assumptions, we improve the identification by using instrumental variables. The instrumental variables fall into two categories: those that affect the likelihood of insider trading but not detection, and those that determine the probability of detection/prosecution conditional on insider trading, but not insider trading. Identification without relying on distributional assumptions requires a minimum of one instrumental variable from either category. We use more than one instrument, resulting in an over-identified model, which allows us to test the validity of individual instruments and examine the robustness of the estimates to changing the exclusion restrictions.

Appendix B provides a list of the variables included in our DCE models and their definitions. To allow comparisons between the probability and characteristics of insider trading and its detection between M&A and earnings announcements, we include the same instrumental variables in our baseline DCE model.

Below we provide an overview of the variables used in our DCE models with a focus on the instrumental variables that provide identification.

The basic idea that underpins one group of our instruments is that detection/prosecution comes *after* insider trading and therefore information that becomes available after the trading can affect the probability of detection but cannot, by construction, affect the decision to conduct insider trading. This general approach to identifying instruments is also exploited by other papers using DCE models (e.g., Wang et al., 2010; Comerton-Forde and Putniņš, 2014; Foley et al., 2019).

3.3.1. Drivers of the decision to engage in illegal insider trading

Starting with the variables that are primarily associated with the decision to engage in illegal insider trading, we include three measures of stock liquidity: turnover (*Turnover*), the Amihud (2002) illiquidity (*Illiquidity*) measure, and bid-ask spread (*Bid-ask spread*). Theoretical models developed by Admati and Pfleiderer (1988) and Collin-Dufresne and Fos (2016) show that informed traders strategically time their trades during times of higher stock liquidity, thereby hiding the price impact of their trades. Similar reasoning might drive cross-sectional variation in the propensity for illegal insider trading. Insiders in more liquid stocks might be more likely to engage in illegal insider trading due to an increased ability to hide their trades and the ability to trade larger volumes and thereby earn larger dollar profits. Note that these are ex-ante measures of liquidity rather than ex-post measures of the actual price impacts and changes in liquidity in the days before a corporate announcement. While the former (ex-ante liquidity) might affect an insider's decision to engage in illegal insider trading, the latter (ex-post measures) might draw the attention of regulators and therefore affect detection. We deal with the latter case in the set of variables that drive the probability of detection.

Next, as drivers of the decision to engage in insider trading we include variables that relate to the value of the inside information. Following the intuition in Kyle (1985), we conjecture that insiders are more likely to engage in insider trading when they have more valuable information, that is, when the difference between observed stock price and fundamental value is large. The first measure is the bid premium (*Bid premium*), which is relevant in M&A cases. It is an ex-ante proxy for the likely magnitude of the target's return in response to the M&A bid announcement. The second measure used for earnings announcements is the magnitude of the earnings surprise (*Earnings surprise*) because the larger the surprise, the more valuable is ex-ante information about the earnings. We also use the abnormal stock return on the M&A or earnings announcement date (*Market reaction [0]*) as a general measure of the value of inside information, which alleviates data constraints: missing information about the bid premium in some cases and missing earnings surprises because not all stocks are followed by analysts.

We also expect that insider information will be more valuable the higher the information asymmetry about the company. To account for this possibility, we include a dummy variable for whether the stock is included in the S&P 500 index (*Index stock*). Stocks not included in the S&P 500 index are likely to have higher information asymmetry due to lower coverage and institutional following, meaning insiders have a greater information advantage thereby making insider information more valuable ex-ante.¹⁰

The next set of variables pick up the likelihood of information leakage by measuring the number of potential insiders with access to the information. Theoretical models show that a higher number of insiders increases the amount and aggressiveness of insider trading (e.g., Holden and Subrahmanyam, 1992; Back et al., 2000; Acharya and Johnson, 2010). For M&A events, this set of variables includes the number of legal and financial advisors on a deal (*Number of advisors*), and the number of bidders for the same target firm (*Number of bidders*). The more advisers, the more complex a deal and the more people are likely to know about it before the official announcement. The more competing bidders, the more insiders are aware of an acquisition target that is due to become “in play”.¹¹

The final variable that attempts to explain variation in the likelihood of insider trading is whether a stock has liquid listed options (*Optionable*). Options markets provide an alternative venue for insiders to trade. Using a sequential trading model, Easley et al. (1998) show that the amount of informed trading in stock and options markets depends on the liquidity and leverage of options markets. Empirically, Patel et al. (2020) show that a meaningful fraction of price discovery (or informed trading) occurs in both stock and options markets. An insider may be more inclined to trade if she can split her trades across both stock and options markets, thereby hiding their trades and increasing her potential profits by obtaining some exposure to the leverage inherent in options.

3.3.2. Drivers of the detection/prosecution of insider trading

Next, we consider variables that affect the probability that a given instance of illegal insider trading is detected and brought to prosecution. Similar to prior studies of misconduct such as fraud, we consider all variables that could affect either detection or the decision to prosecute a suspected case without trying to disentangle detection from prosecution. However, we conduct two validation tests (described in Section

¹⁰ Stock-level information asymmetry may also be captured by other instrumental variables that we consider in our DCE model (e.g., *Market capitalization*, *Turnover*, *Illiquidity*, *Bid-ask spread*, *Optionable*). For example, stocks that are smaller, or less liquid, or that are traded less, or that do not have liquid options contracts are typically not included in the S&P 500 index, and thus have higher levels of information asymmetry.

¹¹ We omit other variables measuring information leakage including the size (*Bidder toehold*) and changes of the bidder toehold stake (*Changes in bidder toehold*) from our DCE models due to missing observations in the *SDC Platinum* database. Larger toehold stakes or larger increases in toehold stakes before the announcement are likely to proxy for longer and more planned acquisition attempts.

4.5) using the number of informal inquiries and trading data preceding announcements to shed some light on insider trading that has been detected, but not prosecuted.¹²

Regulators operate market surveillance systems that alert them to the possibility of illegal insider trading when abnormal price and volume patterns are observed around corporate announcements. For example, FINRA operates a real-time surveillance system known as SONAR, which detects abnormal price and trading activity and matches such measures with filings in the SEC's EDGAR database. Importantly, the patterns of prices and volume immediately preceding a corporate announcement cannot be controlled by an insider, nor are they easy to anticipate ahead of time. The complete patterns of prices and volumes up to the announcement time can only be observed at or after the announcement. Yet, the decision to engage in insider trading must be made in advance of the announcement, that is, before the complete patterns of prices and volumes can be observed. Therefore, we include measures of abnormal returns (*Abnormal returns*) and volumes (*Abnormal \$volume*) measured up to the announcement date as variables that can affect the likelihood of detection/prosecution by drawing regulatory attention but are unlikely to affect the decision to engage in insider trading, as this decision is taken earlier (e.g., DeMarzo et al., 1998).

The next of these variables is the number of prosecutions in the future, that is, in the following 12-month period (*Future prosecutions*). Given that detection/prosecution intensity varies through time, future prosecutions provide additional information about the detection/prosecution intensity at a given point in time that is not observable ex-ante at the time an insider makes a trading decision. It is therefore only included in the detection/prosecution equation of the DCE model.

Another important variable is the regulatory budget (e.g., Del Guercio et al., 2017). We measure regulatory budget using the spending authority of the SEC/DOJ available to combat/monitor/detect insider trading per listed stock. The levels or changes in regulatory budget are not readily available and therefore less likely to be taken into consideration by insider traders.

3.3.3. Variables that drive both insider trading and detection/prosecution

Turning to the variables that are expected to affect both insider trading and detection, we include a range of basic stock characteristics, including market capitalization (*Market capitalization*), the listing exchange (1 – NYSE, 2 – AMEX, and 3 – NASDAQ), and industry (1 – Raw materials, 2 – Construction/manufacturing, and 3 – Services).

In both DCE equations we include a dummy variable following the introduction of the SEC Whistleblower Reward Program (*SEC Whistleblower*) in 2010. This program provides monetary rewards

¹² Informal inquiries represent suspected cases of insider trading that are not brought to formal prosecution. As we use the number of informal inquiries as a validation test of the DCE model estimates, we do not include informal inquiries as an instrumental variable in the DCE models. This allows for a cleaner and more powerful validation test.

to individuals who report insider trading violations and increases the protection to individuals from retaliation for whistleblowing.¹³ Based on the theoretical predictions of Kacperczyk and Pagnotta (2020), regulatory programs like the SEC Whistleblower Reward Program can therefore affect the probability of detection and prosecution by giving regulators an additional source of information about violations. By the same token, it can affect decisions to engage in insider trading by impacting the perceived probability of being detected and prosecuted.

We also include in both DCE equations a dummy variable for whether the M&A bid was preceded by rumors (*Rumored deal*). This variable is constructed using a flag within the *SDC Platinum* database. Rumored deals are more likely to have illegal insider trading either because the insider trading triggers price movements that lead to rumors or the rumors are based on information leakage that was also exploited for illegal trading. Rumored deals could also increase the detection/prosecution likelihood by drawing the attention of the regulator.

4. Results

4.1. Characteristics of announcements with prosecuted insider trading

We start with a simple comparison of the announcements in the prosecuted insider trading sample and all other announcements without addressing the bias from non-random detection and prosecution. Table 1 reports means of characteristics for announcements with (*Detected*) and without (*Not-Detected*) prosecuted insider trading as well as the differences in means. All variable definitions are in Appendix B. The differences in means reflect the drivers of insider trading *and* detection, which we disentangle in the next section using the DCE model.

For both announcement types, detected insider trading is more prevalent in stocks with more liquidity (i.e., higher *Turnover*, lower *Illiquidity*, and lower *Bid-ask spread*) consistent with an increased ability for insiders to conceal their trading in more liquid stocks. Detected insider trading is also more prevalent in stocks that are followed by a wider pool of stakeholders (*Index stock*), and stocks that have actively-traded options (*Optionable*).

Announcements that have prosecuted insider trading tend to have significantly larger abnormal stock returns on the announcement date (*Market reaction [0]*) which is consistent with the likelihood of insider trading being higher when the value of inside information is higher. For example, *Market reaction [0]* is 27.8% (4.8%) for M&A (earnings) announcements with detected insider trading compared to 15.9% (2.9%)

¹³ From inception until 2016, the Whistleblower Program has resulted in more than 14,000 tips, of which approximately 1,000 tips relate to insider trading (<https://www.sec.gov/files/owb-annual-report-2016.pdf>).

for *Not-Detected* announcements. Similarly, *M&A Bid premium* is relatively larger for the *Detected* sample. However, *Earnings surprise(s)* is indistinguishable between the *Detected* and *Not-Detected* samples.

Detected M&A deals also tend to have a larger number of legal and financial advisors (*Number of advisors*) consistent with the notion that more advisors increase the opportunities for inside information to be leaked to a larger pool of individuals. However, the *Number of bidders*, which could also increase the likelihood of information leakage, is higher in the *Not-Detected* sample of M&As.

Announcements with detected insider trading tend to have higher pre-announcement *Abnormal returns* and *Abnormal \$volume*, *Future prosecutions*, and *Regulatory budget*, consistent with increased probability of detection.

Detected insider trading is more prevalent in larger firms (*Market capitalization*), M&A *Rumored deal(s)* and *Leveraged buy-out deal(s)*. These relations could reflect either a higher likelihood of insider trading, a higher detection rate, or both. Furthermore, there tends to be more detected insider trading in NYSE listed stocks when it comes to M&As (*Listing Exchange 1 – NYSE*) and NASDAQ listed stocks when it comes to earnings (*Listing exchange 3 – NASDAQ*), as well as in the manufacturing/construction industry (*Industry 2 – Construction/Manufacturing*).

< Table 1 here >

4.2. *The prevalence of insider trading and its detection*

We estimate the baseline DCE model specified in the previous section using maximum likelihood and a 1,000-iteration bootstrap to generate standard errors and confidence intervals. For each bootstrap sample, we estimate the DCE model and from the DCE model parameters we estimate the probability of insider trading and its detection for each announcement. Table 2 reports the mean estimated prevalence of insider trading (including cases that are not detected or not prosecuted by regulators) and the detection/prosecution rate using the entire 21-year sample period.

The mean probability (estimated prevalence) of insider trading ahead of M&A events (V) is 19.76%, that is, insider trading is estimated to occur in approximately one in five M&A events. The bootstrapped 95% confidence interval is relatively small (ranging between 16.17% and 22.84%), indicating a high degree of confidence that illegal trading occurs in a considerable number of M&A events.¹⁴ For scheduled quarterly earnings announcements, we estimate a considerably lower underlying rate of insider trading: $V = 5.07\%$ with a confidence interval of [3.12%, 8.90%]. The rate implies that approximately one in twenty earnings

¹⁴ The confidence intervals that we report are conservative because the bootstrapped standard errors overestimate the standard errors in the full sample because of the down-sampling that we do to make the bootstrap computationally feasible.

announcements are subject to illegal insider trading, which is about one-quarter of the prevalence of insider trading for M&As.

The results show that the estimated detection rate (mean conditional probability of detection, D) for insider trading ahead of M&As is $D = 13.75\%$, with 95% confidence interval [12.02%, 15.78%]. The detection rate for insider trading ahead of earnings announcements is similar: $D = 14.26\%$, with 95% confidence interval [11.10%, 16.63%]. Our findings are consistent with rational crime theories which imply that rational individuals would not engage in illegal insider trading if the detection rate is too high (e.g., Becker, 1968).

Therefore, estimates from the baseline model suggest that insider trading is much more prevalent than reflected in prosecution cases (only a minority of insider trading is detected and brought to prosecution), is more prevalent around M&A events than quarterly earnings announcements, and conditional on insider trading occurring it is detected at a similar rate in both M&A events and earnings announcements.

< Table 2 here >

We compare the DCE model estimates of V and D to the actual frequency of prosecuted insider trading. The product $V \times D$ (the probability of insider trading times the conditional probability of detection/prosecution) is the model's estimate of the frequency of prosecuted insider trading. If the model's rates are correctly calibrated, the estimated $V \times D$ should be close to the actual frequency of insider trading prosecution as a percentage of all M&A events or earnings announcements.

For M&A events, the actual rate of insider trading prosecutions by US regulators is 5.44% (365/6,712), while the estimated product $V \times D$ from the DCE model for M&A events is approximately 3–4% indicating the model produces estimates of the amount of prosecuted insider trading that are close to the actual rates. Similarly, for earnings announcements, the actual rate of the prosecutions and the estimated rate of prosecutions are close, both being less than 1%. The overall violation and prosecution rates of the DCE model therefore appear well calibrated.

Our estimates for the fraction of M&A events subject to insider trading (19.76%) is approximately four times larger than the fraction of M&A events in which insiders have been detected and prosecuted by regulatory bodies (5.44%). For earnings, we estimate that the fraction of announcements involving illegal trading (5.07%) is at least five times larger than detected. Thus, the DCE model estimates suggest that the prosecuted cases are only the tip of the iceberg. Robustness tests (detailed later) show that the overall rates are fairly robust to changes in model specification.

4.3. Illegal insider trading and its detection through time

How does the prevalence of insider trading and the rate of prosecution vary through time? For M&A events, Figure 2 Panel A plots the time-series of the estimated underlying insider trading prevalence V , and Panel B plots the estimated detection rate D . For comparison, Panel B also plots the fraction of M&A events each year that have prosecuted insider trading (*Detected%*).¹⁵ To understand the drivers of these time-series trends in insider trading and its detection, we also plot the time-series of key instrumental variables used in the baseline DCE model (Figure 3).

< Figure 2 here >

Figure 2 shows that the prevalence of insider trading, V , increases from approximately 10% to 25% between 1996 and 2007. This increase coincides with a large increase in stock liquidity, consistent with the notion that liquidity helps insider traders conceal their trading and earn greater profits. For example, Figure 3 Panel A shows that between 1996 and 2007, *Turnover* roughly doubles in magnitude, and *Illiquidity* (*Bid-ask spread(s)*) declines by a factor of five (ten). The rate of insider trading falls to 15.50% as liquidity decreases during the global financial crisis. For example, *Bid-ask spread(s)* are approximately five times larger for sample stocks in 2009 when compared to 2007. V subsequently recovers to pre-crisis levels (>20%). The higher probability of insider trading following the global financial crisis is consistent with the Intralinks Annual M&A Leaks Report (SS&C Intralinks and CASS Business School) that shows the percentage of US M&As that involve information leaks increased post-crisis, reaching a high in 2015.¹⁶ The time-series of V is consistent with the notion that the introduction of the SEC Whistleblower Program in 2010 had a negative effect on V , which declines from 26.63% to 21.12% between 2010 and 2016 (we test this conjecture more formally with multivariate tests when we examine the determinants of insider trading).

< Figure 3 here >

The estimated probability of detection and prosecution of insider trading ahead of M&A events, D , also tends to increase through time (Figure 2 Panel B). From 1996 to 2007, the probability of detection increases from 5.21% to 30.30%, coinciding with increased *Regulatory budget*. For example, Figure 3 Panel

¹⁵ The annual estimates of V and D are obtained by first estimating the full sample DCE model to ensure there are sufficient observations to estimate the model coefficients, then using the estimated coefficients and actual values of the variables to calculate V and D probabilities each year.

¹⁶ For further details see: <https://www.intralinks.com/blog/2017/01/ma-deal-leaks-increase>. M&A deal leakage is defined as occasions when the cumulative abnormal return of the target firm in the pre-announcement period is significantly different to its expected return.

D illustrates that the SEC regulatory budget per US stock increases from \$30,000 to \$150,000 between 1996 and 2007. The detection probability falls during the global financial crisis (2008–2009) to approximately 10%.

The actual frequency of prosecutions (*Detected%* in Table 2 Panel B) roughly mirrors the trends in the estimated conditional detection probability. This result, indicating that, at least in part, the annual variation in the numbers of prosecuted insider trading cases is driven by variation in detection rates and not only variation in the underlying prevalence of insider trading.

< Figure 4 here >

For earnings announcements, the time-series trends in insider trading and detection are illustrated in Figure 4. Similar to the trends in insider trading ahead of M&A events, the probability of insider trading around earnings announcements tends to increase through time along with improvements in stock liquidity, as between 1996 and 2009, V increases from 1.33% to 11.10%. Earnings-related insider trading falls substantially from over 10% to below 5% following the introduction of the SEC Whistleblower Program in 2010 and remains low in the recent years. These results suggest that the SEC Whistleblower Program may have had a relatively larger deterrence effect on insider trading ahead of earnings than it did for insider trading associated with M&As.

The estimated probability of detection and prosecution of earnings-related insider trading increases throughout our sample, with the increases being larger than those for takeover-related insider trading. For example, during first part of our sample (1996 to 2002), both the estimated D and observed *Detected%* are below 3%. Subsequently, D increases to approximately 15% between 2003 and 2009, together with an 80% increase in *Regulatory budget* during the same period. D increases further to approximately 35% between 2010 and 2016 following the introduction of the SEC Whistleblower Program.

4.4. Drivers of insider trading and its detection

We investigate the drivers of insider trading and its detection using the coefficients of the estimated DCE models. Unlike simple comparisons of prosecuted insider trading versus other announcements, the DCE model coefficients overcome the sample selection bias by separately identifying the violation and detection/prosecution processes, thereby disentangling the drivers of insider trading from the drivers of detection. Table 3 reports the estimated model coefficients for M&A events, their marginal effects (in parentheses), and statistical significance (indicated by asterisks).¹⁷ Model 1 is the baseline DCE model (we

¹⁷ To improve the interpretation and comparison of the DCE model coefficient estimates and marginal effects, we: (i) log transforms right skewed variables (*Market capitalization* and *Regulatory budget*), (ii) square root transform

include the same instrumental variables in our DCE models when examining M&A and earnings announcements), Model 2 is a robustness test which omits the *Optionable* variable from the baseline model, and Model 3 is a robustness test which includes additional M&A deal specific variables in the baseline DCE model (see Section 4.5 for more details regarding robustness tests).

The likelihood of insider trading is higher in stocks with more liquidity, irrespective of whether stock liquidity is measured using *Turnover*, *Illiquidity* or *Bid-ask spread*. Our findings are consistent the theoretical predictions from with Admati and Pfleiderer (1988) and Collin-Dufresne and Fos (2016), where insiders strategically trade in stocks where they can conceal their information and maximize their trading profits. Anecdotal evidence in support of this notion is found in a call with Anil Kumar (McKinsey partner) and Raj Rajaratnam (Galleon Group hedge fund manager): “*buy wherever you want to right because this is liquid enough, you can buy it in the rest of your system as well.*”¹⁸

Inside information is more valuable when the difference between the observed stock price and fundamental value is larger (e.g., Kyle, 1985). The coefficient of *Market reaction [0]* implies that insiders are more likely to engage in illegal trading when the value of their inside information is higher. In addition, we find that the prevalence of insider trading is higher in smaller target firms (*Market Capitalization*). Smaller target firms are likely to have higher information asymmetry, therefore inside information is more valuable as insiders will have a greater information advantage over other investors.

In Model 3, we include M&A deal specific instrumental variables measuring the likelihood of information leakage. We find a positive relationship between the probability of insider trading and the number of financial legal advisors (*Number of advisors*) who have knowledge of the M&A deal prior to the announcement date. Our findings are consistent with theory (e.g., Holden and Subrahmanyam, 1992; Back et al., 2000), a higher number of individuals with access to non-public material information (e.g., *Number of advisors*), increases the opportunities for information leakage and thus the likelihood of insider trading. In contrast, we report a negative relationship between insider trading likelihood and the level of information leakage measured using the number of bidders for a target firm (*Number of bidders*) and deals which are preceded by a rumor (*Rumored deal*). A possible explanation for these findings is that potential insider traders might view M&A deals that have multiple bidders or which are preceded by rumors as deals that might draw the attention of regulatory bodies because such information is disclosed in company announcements or in the media, thereby discouraging illegal trading. In contrast, the number of legal and financial advisors is less observable and so has a smaller deterrence effect. In addition, our DCE model indicates that the likelihood of insider trading is higher for *Leveraged buy-out deal(s)*. Such findings are

Abnormal \$volume (where required negative values are multiplied by negative one before and after applying the transformation), (iii) standardize all instrumental variables to have a mean equal to zero and standard deviation of one, and (iv) winsorize all instrumental variables at the 1st and 99th percentiles.

¹⁸ See <https://www.ft.com/content/e0eaddca-8b94-49d5-b3f5-128981cd0bf7>.

consistent with Acharya and Johnson (2010), leveraged deals is another proxy for information leakage, as additional individuals (e.g., lending bank) have knowledge of the M&A prior to the announcement date.¹⁹

Through our DCE model, we also report that insider trading is more likely to occur in firms listed on the NYSE and NASDAQ stock exchanges (*Listing exchange 1* and *3*, respectively), relative to stocks listed on the AMEX exchange. Consistent with the time-series trends discussed in the previous section, the DCE estimates more formally indicate that the *SEC Whistleblower* program is an effective deterrent of insider trading in more recent times.

< Table 3 here >

The marginal effects in parentheses indicate the relative importance and magnitude of the drivers, which have been standardized (mean zero and standard deviation of one) for easier comparison. For example, the marginal effect of *Turnover* is 0.116, which indicates that a one-standard deviation increase in turnover in the year preceding an M&A event increases the probability of insider trading by 11.6% of the mean underlying rate of insider trading. Most of the instrumental variables are statistically significant and have an economically meaningful effect on the likelihood of insider trading.

The estimated drivers of detection suggest that detection/prosecution probabilities increase when there are larger abnormal returns and volume before the announcement (*Abnormal returns [-1]* and *Abnormal volume [-1]*), consistent with the design of optimal enforcement models (e.g., DeMarzo et al., 1998) and the use of real-time surveillance systems like SONAR which detect suspicious trades. Prosecution likelihood is also higher when regulators have a larger normalized budget (*Regulatory budget*) and higher intensity of enforcement as proxied by the future rate of prosecutions (*Future prosecutions*). All the instrumental variables associated with detection are statistically significant and have economically meaningful effects on the probability of detection and prosecution $D()$.

The control variables indicate that regulators are more likely to detect and prosecute insider trading in larger stocks (*Market Capitalization*) and in AMEX-listed stocks (*Listing Exchange 2*). At the same time, the probability of insider trading is lower in larger stocks and AMEX-listed stocks.

< Table 4 here >

Table 4 reports the DCE model estimates for earnings announcements. Similar to M&As, and consistent with strategic insider trading (e.g., Admati and Pfleiderer, 1988), there is a higher probability of

¹⁹ Acharya and Johnson (2010) find that insider trading is more likely prior to buyout deals when there is an increased number of financing parties who have knowledge of the deal in advance of the announcement date.

insider trading prior to earnings announcements in more liquid stocks (based on *Turnover* and *Bid-ask spread*). Insider trading is also more likely when the value of earnings information is larger as measured by the market reaction upon announcement (*Market reaction [0]*). However, for both M&A events and earnings announcements, the decision to engage in illegal insider trading is not significantly associated with the ability to further conceal information, or the opportunity to increase the leverage of their trades using options (*Optionable*). *Optionable* indicates the total yearly options volume is in the top tercile.

The control variables are statistically significant and indicate that the probability of insider trading is more likely in larger stocks (*Market capitalization*), NYSE and NASDAQ listed stocks (*Listing exchange 1* and *3*, respectively), and in the raw materials, construction/manufacturing, and services industries (*Industry 1, 2* and *3*, respectively).

Detection and prosecution of insider trading ahead of earnings announcements is more likely when there is abnormal volume around the announcement (*Abnormal \$volume [-1]*) and when regulatory budgets are larger (*Regulatory budget*). Regulators appear to focus their detection and prosecution efforts in larger stocks (*Market capitalization*) and stocks operating in the construction/manufacturing industry (*Industry 2*).

Consistent with the time-series trends noted earlier, the SEC Whistleblower Program in 2010 has a large effect on the likelihood of detection and prosecution of insider trading ahead of earnings announcements. The marginal effects indicate that depending on the model, the program is estimated to increase the detection probability by a factor of 2.7 to 3.0, that is approximately tripling the detection probability. At the same time, the program is also estimated to have a substantial deterrence effect, reducing the likelihood of illegal trading. Our findings are consistent with insiders internalizing legal risk such as in the Kyle (1985) meets Becker (1968) framework proposed by Kacperczyk and Pagnotta (2020).

4.5. Robustness tests

We conduct several robustness tests of the DCE model estimates. In our first set of tests, we exploit the fact that the DCE model is over-identified to examine the robustness of the results to the exclusion restrictions. We relax the exclusion restrictions for individual instrumental variables one at a time, re-estimating the DCE model nine times. For example, in the first variation we estimate the baseline model without the *Turnover* variable, in the second we exclude the *Illiquidity* variable, and so forth.

< Table 5 here >

Table 5 reports the results of the robustness tests. Panel A reports the baseline DCE model estimates for the sake of comparison. Panel B reports the mean *V* and *D* (and associated 95% confidence intervals)

for the nine tests of the exclusion restrictions. The results suggest that the baseline DCE model is not overly sensitive to the individual exclusion restrictions. Similar to the baseline results, approximately one in five (20) M&As (earnings) announcements are estimated to have insider trading, and insider trading is detected in approximately 15% of M&A and earnings announcements.

Table 5 Panel C similarly shows that the estimates are not overly different to the baseline estimates when we exclude the *Optionable* variable and re-run the tests of the nine exclusion restrictions. We also draw similar conclusions about the main drivers of insider trading and its detection in the exclusion restriction tests (see Model 2 in Tables 3 and 4 which presents coefficient estimates, marginal effects and *t*-statistics from the baseline DCE model in which we exclude the *Optionable* variable).

In the second set of robustness tests, we use alternate definitions of the instrumental variables and include additional instrumental variables. We estimate the following eight variations of the baseline model: (i) replace *Abnormal \$volume [-1]* with *Abnormal volume [-1]*; (ii) replace *Abnormal returns [-1]* and *Abnormal \$volume [-1]* with *Abnormal returns [-3, -1]* and *Abnormal \$volume [-3, -1]*, respectively; (iii) replace *Abnormal returns [-1]* and *Abnormal \$volume [-1]* with *Abnormal returns [-3, -1]* and *Abnormal volume [-3, -1]*, respectively; (iv) replace *Abnormal returns [-1]* and *Abnormal \$volume [-1]* with *Abnormal returns [-5, -1]* and *Abnormal \$volume [-5, -1]*, respectively; (v) replace *Abnormal returns [-1]* and *Abnormal \$volume [-1]* with *Abnormal returns [-5, -1]* and *Abnormal volume [-5, -1]*, respectively; (vi) replace *Market reaction [0]* with *Market reaction [0, +1]*; (vii) include *Index stock* as an additional instrumental variable; and (viii) for M&As (earnings) include the following additional instrumental variable(s): *Bid premium*, *Number of advisors*, *Rumored deal*, and *Leveraged buy-out deal (Earnings surprise)*.²⁰ For conciseness, Panel D reports the mean *V* and *D* (and associated 95% confidence intervals) for the eight variations of the baseline model. The results suggest that the estimates of *V* and *D* are robust to alternate variable definitions and the inclusion of additional instrumental variables. For example, Table 5 Panel E reports DCE model estimates for specification (viii).

In addition, Model 3 of Tables 3 and 4 reports DCE coefficient estimates/*t*-statistics and marginal effects from the baseline DCE model in which we include M&A and earnings specific variables (i.e.,

²⁰ In specification (vi), we follow the corporate finance literature and measure abnormal announcement date returns over the [0, +1] period (e.g., Hao, 2016). In specification (vii), we include *Index stock* as an additional instrumental variable. In our baseline model, we omit the *Index stock* variable as it is correlated with several variables in our baseline model. For example, *Market capitalization* provides an alternate proxy for stock information asymmetry as small stocks are not included in the S&P 500 index and thus have higher levels of information asymmetry. In specification (viii), we include M&A (earnings) specific variables as additional instruments into our M&A (earnings) baseline model. In our baseline model for M&A and earnings announcements we include the same instrumental variables so we can make more consistent comparisons between the probability and characteristics of insider trading and its detection in M&A and earnings announcements. We also omit M&A and earnings specific variables (e.g., *Bid premium*, *Earnings surprise*) from our baseline model because such variables increase the risk of co-linearity, and are subject to missing observations (e.g., not all stocks in our sample are followed by analysts, resulting in missing earnings surprise data).

specification (vii)), respectively. The estimated drivers of insider trading and its detection for both M&As and earnings announcements are similar across Models 1 to 3.

4.6. Validation tests

We conduct two additional validation tests of the DCE model estimates. In the first, we compare our DCE model estimates of the prevalence of illegal insider trading to the number of informal inquiries conducted by the SEC/DOJ. Informal inquiries represent suspected cases of insider trading that are not brought to prosecution. They provide a sense of how much insider trading is detected but not prosecuted. While we do not have data on which announcements or stocks were the subject of informal inquiries, we do have data on the annual number of informal inquiries from the SEC.

Figure 5 plots the estimated prevalence of insider trading ahead of M&A events (solid black line) and earnings announcements (dotted black line) and the number of informal inquiries (*Informal inquiries*; solid gray line). The number of informal inquiries increases from 426 to 910 between 1996 and 2003. Subsequently, the number of informal inquiries fluctuates between 776 and 980 during the remainder of the sample period. The trends suggest that *Informal inquiries* provides support for the estimated prevalence of insider trading. First, V (M&A), V (Earnings), and *Informal inquiries* show a similar increase from 1996 to 2008, before stabilizing. Second, there is a strong positive correlation between V (M&A) and *Informal inquiries* (0.30), and between V (Earnings) and *Informal inquiries* (0.31).

< Figure 5 here >

The second validation test analyzes trading data preceding the announcements that the DCE model flags as likely to have illegal insider trading (V High) and compares it with trading data preceding announcements not flagged for insider trading (V Low). We define announcements as likely to have insider trading if the estimated probability of insider trading (V) is above its median (V High) and unlikely otherwise (V Low).²¹ We look for differences in abnormal returns and volume. Importantly, none of these trading characteristics are in the DCE equation that generates estimates of the likelihood of illegal insider trading (V ()). Therefore, any differences in the trading characteristics consistent with the expected impact of illegal insider trading would add support for the model's estimates. An absence of differences, or

²¹ Our findings illustrated in Figures 6 and 7 are robust if we use alternate definitions to define the V High and V Low samples. For example, we define announcements as likely to have insider trading if the estimated probability of insider trading is above the mean. Or alternatively, given that our full sample V for M&As (earnings) = 20% (5%), we define announcements as likely to have insider trading if the estimated probability of insider trading is in the top 20% (5%).

differences inconsistent with the expected impact of illegal insider trading would flag issues with the model's estimates.

For each stock-day in the period $[-5, +5]$ around the announcement date, we calculate abnormal returns as the difference between the daily returns and the market return.²² Daily abnormal dollar volume is calculated as the difference between the daily dollar volume minus that stock's prior month average daily dollar volume during the period $[-30, -6]$.

Figure 6 illustrates abnormal returns (Panel A) and abnormal dollar volume (Panel B) during the period $[-5, +5]$ around M&A announcements for target firms. During the pre-announcement period, there is a return run-up consistent with the expected impact of insider trading. First, each daily pre-announcement abnormal return for both the *V High* and *V Low* sample is significantly different from zero. Second, the cumulative five-day pre-announcement abnormal return run-up is economically meaningful, 4.59% for *V Low* (black dotted line) and 2.79% for *V High* (solid black line) M&As.

Furthermore, we observe large abnormal announcement date returns. The abnormal announcement date return is 28.22% for the *V High* sample, six times larger in magnitude when compared to the *V Low* sample (difference of means *t*-statistic = 25). In support of our earlier findings, we find that the probability of insider trading (*V*) is positively associated with M&As where the value of inside information is larger.

Panel B illustrates that throughout the entire pre-announcement period, daily abnormal dollar volume is significantly larger for *V Low* M&As when compared to *V High* M&As (indicated by dark gray *t*-statistic bars). Cumulative five-day pre-announcement abnormal dollar volume is large, approximately \$9 million (\$47 million) for the *V High* (*V Low*) sample. For both M&As samples, abnormal dollar volume on the announcement date is indistinguishable at approximately \$175 million. Furthermore, for both M&As samples, the daily mean abnormal dollar volume values during the $[-5, +5]$ period are significantly different from zero.

The evidence in Figure 6 Panels A and B indicate that *V High* M&As are associated with relatively lower (higher) pre-announcement (announcement date) abnormal returns and dollar volume traded. Our findings are consistent with insider trading occurring in smaller target firms where the value of inside information is larger. In support of this conclusion, the mean stock price and market capitalization of stocks in the *V High* (*V Low*) sample of M&As is approximately \$15 (\$25) and \$600 (\$2,000) million, respectively.

< Figure 6 here >

²² Our results are robust if we calculate daily abnormal returns as the difference between the daily stock return minus that stock's prior month average daily stock return during the period $[-30, -6]$.

Figure 7 plots the behavior of abnormal returns (Panel A) and dollar volume (Panel B) during the 11-day earnings event period.²³ Prior to earnings announcements, abnormal returns fluctuate around 0% for both the *V High* and *V Low* samples. However, mean *V High* abnormal dollar volume is significantly larger than the *V Low* sample throughout the pre-announcement period (indicated by light gray *t*-statistic bars). The cumulative five-day pre-announcement abnormal dollar volume is \$23 million (\$6 million) for the *V High* (*V Low*) earnings sample. On the announcement date, both abnormal returns and dollar volume are relatively larger for the *V High* sample (e.g., 3.31% versus 2.23%, and \$69 million versus \$26 million).

< Figure 7 here >

The patterns in trading activity show the expected impact of insider trading around earnings announcements for both the *V High* and *V Low* samples. For example, for the *V High* sample, the daily mean abnormal returns and volume during the $[-5, +1]$ period is significantly different from zero, and larger in magnitude when compared to the *V Low* sample. Our findings are consistent with insider trading occurring in larger firms. For earnings announcements, profitable opportunities are more likely to occur in stocks with earnings surprises (i.e., typically larger stocks, as analysts are less likely to cover and forecast earnings for smaller stocks). Consistent with this conjecture, the mean stock price and market capitalization of stocks in the *V High* (*V Low*) sample of earnings is approximately \$250 (\$100) and \$9,000 (\$6,000) million, respectively.

5. Conclusion

We use structural estimation models (detection-controlled estimation) and comprehensive hand-collected data on all US prosecuted cases of insider trading to examine the underlying prevalence of insider trading and what fraction is brought to prosecution. The structural estimation models also allow us to disentangle the drivers of insider trading and the drivers of detection/prosecution, overcoming the classic sample selection issue that is a challenge in examining illegal activity.

We find that a significant amount of insider trading occurs prior to unscheduled (M&A) and scheduled (earnings) news announcements: approximately one in five M&A events and one in 20 quarterly earnings announcements. The estimated prevalence of insider trading is at least four times larger than the number of prosecutions by the SEC/DOJ, indicating that prosecutions are only the tip of the iceberg.

²³ Figure 7 reports results for both positive and negative earnings news. For negative earnings news, we multiply abnormal returns by minus one. We use the sign of the stock return on the announcement date to determine positive and negative earnings news.

We find that insider trading is more likely when there is more liquidity which allows insiders to conceal their trades and earn higher profits. Insider trading is also more likely when the value of the inside information is higher, as measured by market reactions to the announcement of the information. Our estimates suggest that the SEC Whistleblower Program is effective in deterring a substantial amount of insider trading. We show that the conditional detection rate increases through time, particularly for insider trading ahead of earnings, partly due to the Whistleblower program and partly due to increases in the normalized SEC budget.

The overall estimated detection/prosecution rates are around 15% and do not vary considerably between M&A events and earnings announcements. Detection and prosecution is more likely when there are abnormal returns and volumes before the announcement of the information and when regulatory budgets are larger.

Our findings have a number of implications. First, given that the detection/prosecution rates vary through time, the number of prosecutions does not adequately reflect trends in the prevalence of insider trading. Similarly, a comparison of prosecuted cases to announcements with no prosecuted insider trading can lead to biased inference about the characteristics of insider trading because of the non-random detection and prosecution processes. Our findings about the determinants of insider trading and its detection could be used by regulators to more efficiently direct surveillance and enforcement effort to where it will yield the greatest results in terms of detections but also deterrence.

References

- Acharya, V. V., Johnson, T. C., (2010). More insiders, more insider trading: Evidence from private-equity buyouts. *Journal of Financial Economics* 98, 500–523.
- Admati, A., Pfleiderer, P., (1988). A theory of intraday patterns: Volume and price variability. *Review of Financial Studies* 1, 3–40.
- Ahem, K. R., (2017). Information networks: Evidence from illegal insider trading tips. *Journal of Financial Economics* 125, 26–47.
- Ahem, K. R., (2020). Do proxies for informed trading measure informed trading? Evidence from illegal insider trades. *Review of Asset Pricing Studies* 10, 397–440.
- Akey, P., Gregoire, V., Martineau, C., (2021). Price revelation from insider trading: Evidence from hacked earnings news. *Journal of Financial Economics*, forthcoming.
- Amihud, Y., (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets* 5, 31–56.
- Amin, K., Lee, C., (1997). Option trading, price discovery, and earnings news dissemination. *Contemporary Accounting Research* 14, 153–192.
- Augustin, P., Brenner, M., Subrahmanyam, M. G., (2019). Informed options trading prior to takeover announcements: Insider trading? *Management Science* 65, 5697–5720.
- Augustin, P., Brenner, M., Grass, G., Subrahmanyam, M. G., (2016). How do insiders trade? *Working paper, McGill University, New York University and HEC Montreal*.
- Back, K., Cao, H. H., Willard, G. A., (2000). Imperfect competition among informed traders. *Journal of Finance* 55, 2117–2155.
- Becker, G., (1968). Crime and punishment: An economic approach. *Journal of Political Economy* 169, 176–177.
- Bhattacharya, U., Daouk, H., (2002). The world price of insider trading. *Journal of Finance* 57, 75–108.
- Cao, C., Chen, Z., Griffin, J. M., (2005). Informational content of option volume prior to takeovers. *Journal of Business* 78, 1073–1109.
- Carlton, D., Fischel, D., (1983). The regulation of insider trading. *Stanford Law Review* 35, 857–895.
- Chan, K., Ge, L., Lin, T.C., (2015). Informational content of option trading on acquirer announcement return. *Journal of Financial and Quantitative Analysis* 50, 1057–1082.
- Collin-Dufresne, P., Fos, V., (2015). Do prices reveal the presence of informed trading? *Journal of Finance* 70, 1555–1582.
- Collin-Dufresne, P., Fos, V., (2016). Insider trading, stochastic liquidity, and equilibrium prices. *Econometrica* 84, 1441–1475.

- Comerton-Forde, C., Putniņš, T. J., (2014). Stock price manipulation: Prevalence and determinants. *Review of Finance* 18, 23–66.
- Del Guercio, D., Odders-White, E. R., Ready, M. J., (2017). The deterrent effect on the Securities and Exchange Commission’s enforcement intensity on illegal insider trading: Evidence from run-up before news events. *Journal of Law and Economics* 60, 269–307.
- Easley, D., de Prado, M. M. L., O’Hara, M., (2012). Flow toxicity and liquidity in a high-frequency world. *Review of Financial Studies* 25, 1457–1493.
- Easley, D., O’Hara, M., Srinivas, P., (1998). Option volume and stock prices: Evidence on where informed traders trade. *Journal of Finance* 53, 431–465.
- Feinstein, J. S., (1989). The safety regulation of US nuclear power plants: Violations, inspections and abnormal occurrences. *Journal of Political Economy* 97, 115–154.
- Feinstein, J. S., (1990). Detection controlled estimation. *Journal of Law and Economics* 33, 233–276.
- Feinstein, J. S., (1991). An econometric analysis of income tax evasion and its detection. *RAND Journal of Economics* 22, 14–35.
- Fernandes, N., Ferreira, M., (2009). Insider trading laws and stock price informativeness. *Review of Financial Studies* 22, 1845–1887.
- Foley, S., Karlsen, J. R., Putniņš, T. J., (2019). Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies? *Review of Financial Studies* 32, 1798–1853.
- Glosten, L. R., Milgrom, P. R., (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14, 71–100.
- Hao, Q., (2016). Is there information leakage prior to share repurchase announcements? Evidence from daily options trading. *Journal of Financial Markets* 27, 79–101.
- Holden, C. W., Subrahmanyam, A., (1992). Long-lived private information and imperfect competition. *Journal of Finance* 47, 247–270.
- Kacperczyk, M., Pagnotta, E., (2019). Chasing private information. *Review of Financial Studies* 32, 4997–5047.
- Kacperczyk, M., Pagnotta, E., (2020). Becker meets Kyle: Legal risk and insider trading. *Working paper, Imperial College London*.
- Kyle, A.S., (1985). Continuous auctions and insider trading. *Econometrica* 53, 1315–1335.
- Leland, H., (1992). Insider trading: Should it be prohibited? *Journal of Political Economy* 100, 859–887.
- Livnat, J., Mendenhall, R. R., (2006). Comparing the post-earnings announcement drift for surprises calculated from analysts and time series forecasts. *Journal of Accounting Research* 33, 177–205.
- Meulbroek, L., (1992). An empirical analysis of illegal insider trading. *Journal of Finance* 47, 1661–1699.

- Meulbroek, L., Hart, C., (1997). The effect of illegal insider trading on takeover premia. *Review of Finance* 1, 51–80.
- Patel, V., Putnins, T. P., (2020). Price discovery in stock and options markets. *Journal of Financial Markets* 47, 1–28.
- SEC (2017). FY 2017 Congressional Budget Justification. *US Securities and Exchange Commission*.
- Wang, T. Y., Winton, A., Yu, X., (2010). Corporate fraud and business conditions: Evidence from IPOs. *Journal of Finance* 65, 2255–2292.

Appendix A: Derivations for the DCE model

We define $V(\cdot)$ and $D(\cdot)$ to be monotonic link functions that map $\beta_1 x_{1i}$ and $\beta_2 x_{2i}$ to the latent probability of insider trading, and the latent probability of detected/prosecuted insider trading, respectively. We use cumulative logistic distribution functions as link functions: $V(\beta_1 x_{1i}) = \frac{1}{1+e^{-\beta_1 x_{1i}}}$, and $D(\beta_2 x_{2i}) = \frac{1}{1+e^{-\beta_2 x_{2i}}}$:

$$V(\beta_1 x_{1i}) = \Pr(L_{1i} = 1), \quad (\text{A.1})$$

$$D(\beta_2 x_{2i}) = \Pr(L_{2i} = 1 \mid L_{1i} = 1). \quad (\text{A.2})$$

Table A1 reports the probabilities of various joint outcomes (represented by cells in the table). The joint outcomes are mutually exclusive and exhaustive, so the probabilities in Table A1 sum to one.

Table A1: Two-stage DCE model probability matrix

	Insider	No-insider
Detected/prosecuted	$V(\beta_1 x_{1i})D(\beta_2 x_{2i})$	0
Not detected/prosecuted	$V(\beta_1 x_{1i})[1 - D(\beta_2 x_{2i})]$	$1 - V(\beta_1 x_{1i})$

The log likelihood of detected/prosecuted insider trading (identified in category A of Figure 1) is the log of the sum (over announcements in A) of the probabilities of that joint outcome:

$$\log L_A = \sum_{i \in A} \log [V(\beta_1 x_{1i})D(\beta_2 x_{2i})]. \quad (\text{A.3})$$

Similarly, the log likelihood of not detected insider trading and no insider trading (identified in category A^c of Figure 1) is the log of the sum (over announcements in A^c) of the probabilities of that joint outcome (the probability of not detected insider trading plus the probability of no insider trading):

$$\log L_{A^c} = \sum_{i \in A^c} \log [V(\beta_1 x_{1i})[1 - D(\beta_2 x_{2i})] + 1 - V(\beta_1 x_{1i})], \quad (\text{A.4})$$

$$\log L_{A^c} = \sum_{i \in A^c} \log [1 - V(\beta_1 x_{1i})D(\beta_2 x_{2i})]. \quad (\text{A.5})$$

Sets A and A^c constitute the universe of trading prior to announcements. Therefore, the full-sample log likelihood is:

$$\log L = \sum_{i \in A} \log [V(\beta_1 x_{1i})D(\beta_2 x_{2i})] + \sum_{i \in A^c} \log [1 - V(\beta_1 x_{1i})D(\beta_2 x_{2i})]. \quad (\text{A.6})$$

Maximum likelihood estimation involves selecting parameter vectors β_1 and β_2 such that the function $\log L$ is maximized.

Appendix B. Definitions of variables

This appendix defines the variables used in the two-equation DCE model of insider trading and detection.

Variable	Definition
Panel A: Outcome variable	
<i>Detected insider trading</i>	Dummy variable for all corporate announcements (M&A and earnings announcements) that are known to have illegal insider trading.
Panel B: Variables associated with both insider trading and detection	
<i>Market capitalization</i>	Closing price multiplied by the number of shares outstanding at the beginning of the year.
<i>Listing exchange</i>	Three dummy variables for each of the following US stock exchanges: NYSE (<i>Listing Exchange 1</i>), AMEX (<i>Listing Exchange 2</i>), and NASDAQ (<i>Listing Exchange 3</i>).
<i>Industry</i>	Four dummy variables for each of the following industries: raw materials (<i>Industry 1</i>), construction/manufacturing (<i>Industry 2</i>), services (<i>Industry 3</i>), and other (<i>Industry 4</i>).
<i>SEC Whistleblower</i>	Dummy variable following the introduction of the SEC Whistleblower Reward Program on July 21, 2010.
<i>Rumored deal</i>	Dummy variable for M&A deals that are rumored to occur based on the <i>SDC Platinum</i> flag.
<i>Leveraged buy-out deal</i>	Dummy variable for M&A deals that are leveraged buy-outs.

Appendix B. Definitions of variables (continued)

Variable	Definition
Panel C: Variables associated primarily with the likelihood of insider trading	
<i>Turnover</i>	Daily traded dollar volume in the stock divided by market capitalization, averaged over the past year.
<i>Illiquidity</i>	Daily absolute returns divided by daily traded dollar volume, averaged over the past year (Amihud, 2002).
<i>Bid-ask spread</i>	Closing bid price minus the closing offer price divided by the closing midpoint price, averaged over the past year.
<i>Market reaction</i>	Abnormal stock return on the M&A or earnings announcement date (<i>Market reaction [0]</i>). Abnormal returns are calculated as the stock return in excess of the market return. Alternative definitions used in robustness tests: Abnormal stock return on days [0, +1] around the M&A or earnings announcements (<i>Market reaction [0, +1]</i>).
<i>Bid premium</i>	Offer price minus the target firm's stock price on day $t - 20$.
<i>Earnings surprise</i>	Actual earnings per share for the current quarter minus the earnings per share in the same quarter of the previous fiscal year, divided by the share price at the end of the current quarter (Livnat and Mendenhall, 2006).
<i>Index stock</i>	Dummy variable for whether a stock is a constituent of the S&P 500.
<i>Number of advisors</i>	Total number of legal and financial advisors across the bidding and target firms.
<i>Number of bidders</i>	Total number of bidders for the same target firm.
<i>Bidder toehold</i>	Percentage of shares owned by the bidding firm in the target firm, calculated six months prior to the announcement date.
<i>Change in bidder toehold</i>	Percentage of shares purchased by the bidding firm in the target firm in the six months prior to the announcement date.
<i>Optionable</i>	Dummy variable for whether the total yearly options volume is in the top tercile across firms.

Appendix B. Definitions of variables (continued)

Variable	Definition
Panel D: Variables associated primarily with the detection/prosecution of insider trading	
<i>Abnormal returns</i>	Abnormal returns for a stock in a period of one (<i>Abnormal returns</i> $[-1]$), three (<i>Abnormal returns</i> $[-3, -1]$), or five days (<i>Abnormal returns</i> $[-5, -1]$) leading up to an announcement (M&A or earnings). Abnormal returns are calculated as the stock return in excess of the market return.
<i>Abnormal \$volume</i>	Abnormal dollar volume for a stock in a period of one (<i>Abnormal \$volume</i> $[-1]$), three (<i>Abnormal \$volume</i> $[-3, -1]$), or five (<i>Abnormal \$volume</i> $[-5, -1]$) days leading up to an announcement (M&A or earnings). Dollar volume is volume traded multiplied by stock price. Daily abnormal dollar volume is calculated as the daily dollar volume minus that stock's prior month average daily dollar volume during the period $[-30, -6]$. Similarly, we define abnormal non-dollar volume (<i>Abnormal volume</i> $[-1]$, <i>Abnormal volume</i> $[-3, -1]$, and <i>Abnormal volume</i> $[-5, -1]$).
<i>Future prosecutions</i>	Number of insider trading prosecutions filed by the SEC and DOJ in the following 12-month period, based on the date of filing the statement of allegations.
<i>Regulatory budget</i>	Spending authority of SEC and DOJ divided by the number of US common stocks. Budgets are deflated by implicit price deflators for GDP.

Figure 1. Two-stage DCE model

The figure illustrates the structure of the two-stage DCE model. Stage 1 models the characteristics of the illegal insider trading process. Stage 2 models the characteristics of the detection/prosecution process. Both stages are estimated simultaneously using maximum likelihood to select parameter values that maximize the likelihood of the observable classifications, A and A^c .

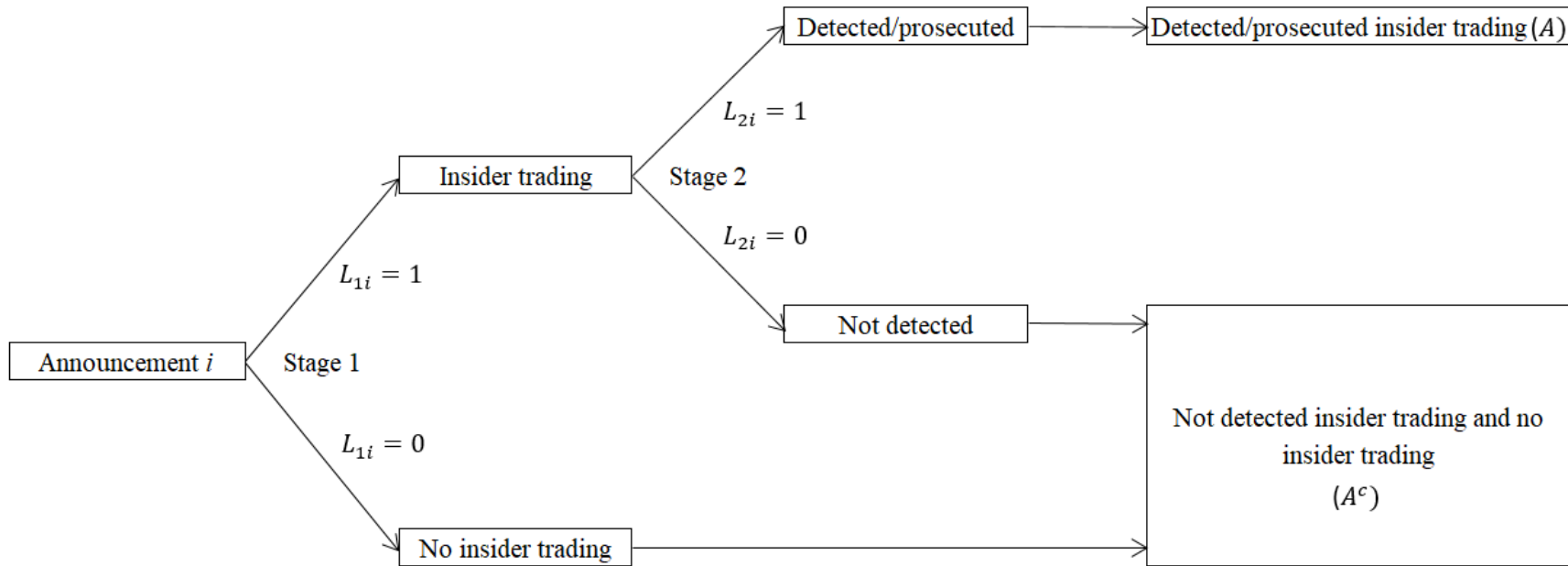
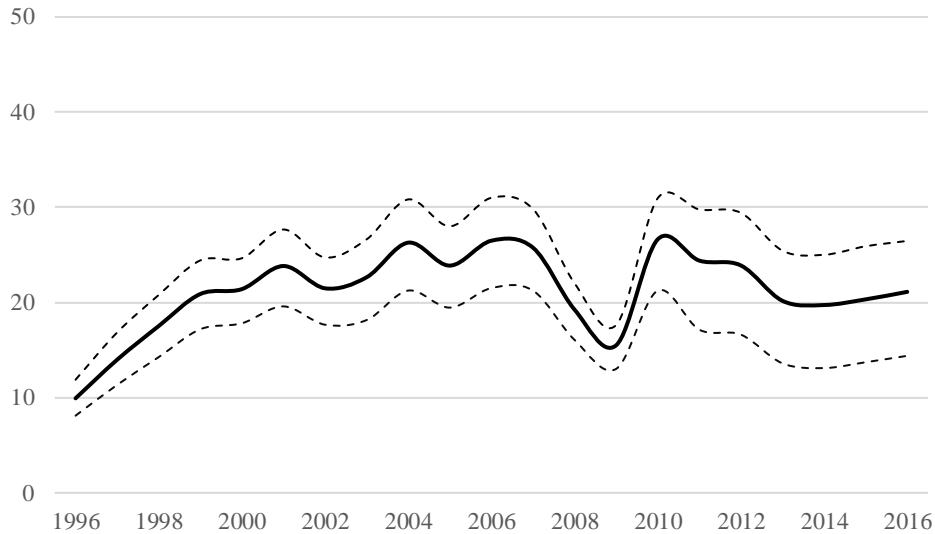


Figure 2. Insider trading and detection rates through time for M&As

This figure illustrates annual estimated rates of insider trading and detection for M&A events. Panel A illustrates the estimated prevalence of insider trading (V ; solid black line) and 95% confidence intervals (dashed black lines). Panel B illustrates the estimated probability of detection (D ; solid gray line) and 95% confidence intervals (dashed gray lines), and the percentage of announcements in which insider trading has been detected and prosecuted by the SEC/DOJ ($Detected\%$; dashed black line). Estimates for V and D are obtained from our baseline DCE model using a bootstrap procedure involving 1,000 iterations. Our sample comprises 365 M&A events in which insider trading has been detected and prosecuted by the SEC/DOJ and 6,347 M&A events in which insider trading has not been detected and prosecuted.

Panel A: Probability of insider trading (V)



Panel B: Probability of detection (D) and percentage detected by the SEC/DOJ ($Detected\%$)

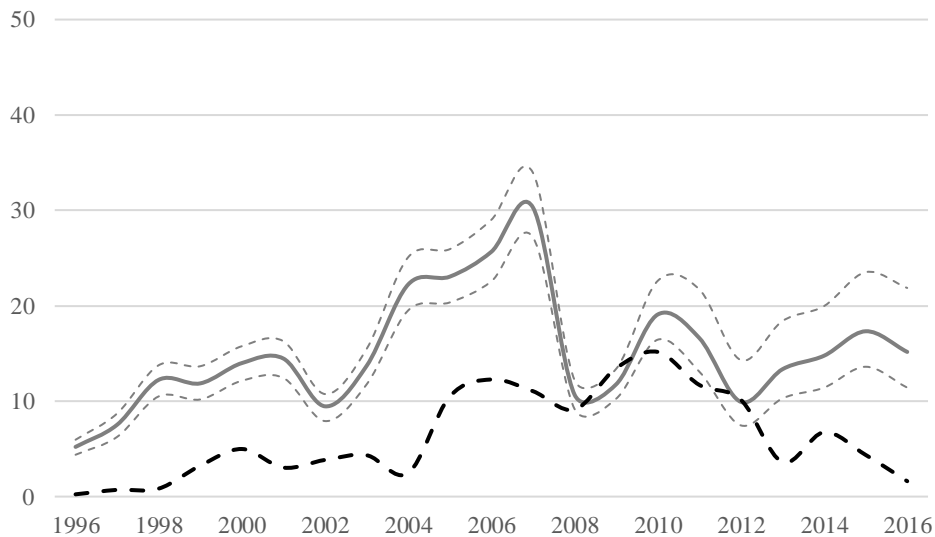
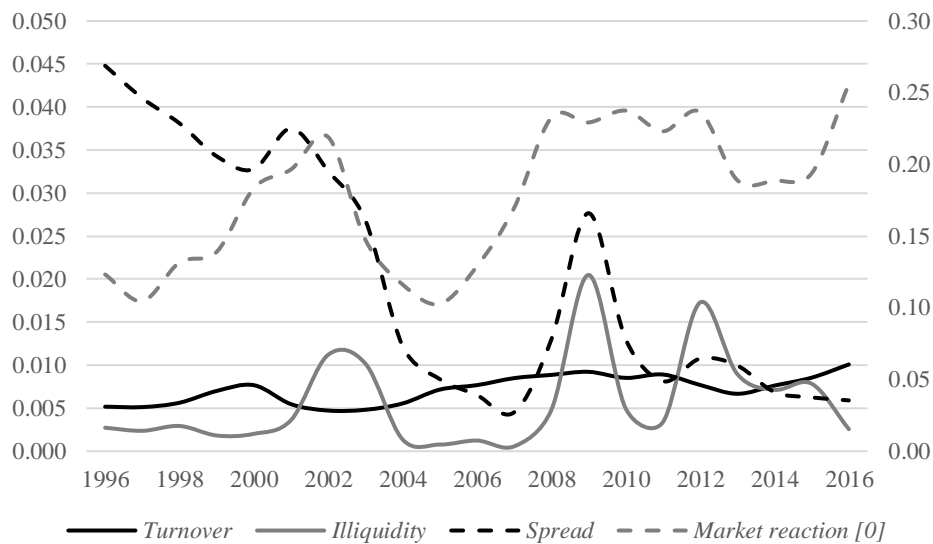


Figure 3. Instrumental variables through time

This figure illustrates selected instrumental variables through time. For target M&A firms in our sample, Panel A illustrates *Turnover* (solid black line, magnitude represented on left-hand axis), *Illiquidity* (solid gray line, magnitude multiplied by 1,000, left-hand axis), *Bid-ask spread* (dashed black line, left-hand axis), and *Market reaction [0]* (dashed gray line, right-hand axis). For earnings announcement firms in our sample, Panel B illustrates *Turnover* (solid black line), *Illiquidity* (solid gray line, magnitude multiplied by 100,000), *Bid-ask spread* (dashed black line), and *Market reaction [0]* (dashed gray line). Panel C illustrates *Market capitalization* for target M&A firms (solid black line) and for earnings announcing firms (solid gray line), magnitude expressed in millions. Panel D illustrates *Regulatory budget* (solid black line, left-hand-axis) and *Future prosecutions* (solid gray line, right-hand-axis). Variable definitions are in Appendix B. In our sample, there are 365 (88) M&A (earnings) announcements in which insider trading has been detected and prosecuted by the SEC/DOJ, and 6,347 (162,589) M&A (earnings) announcements in which insider trading has not been detected and prosecuted.

Panel A: M&As – *Turnover*, *Illiquidity*, *Bid-ask spread* and *Market reaction [0]*



Panel B: Earnings – *Turnover*, *Illiquidity*, *Bid-ask spread* and *Market reaction [0]*

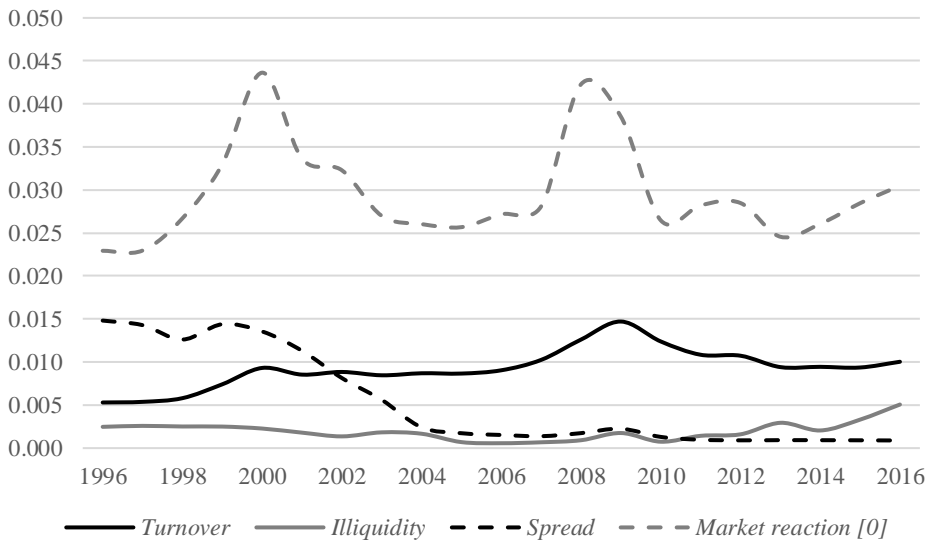
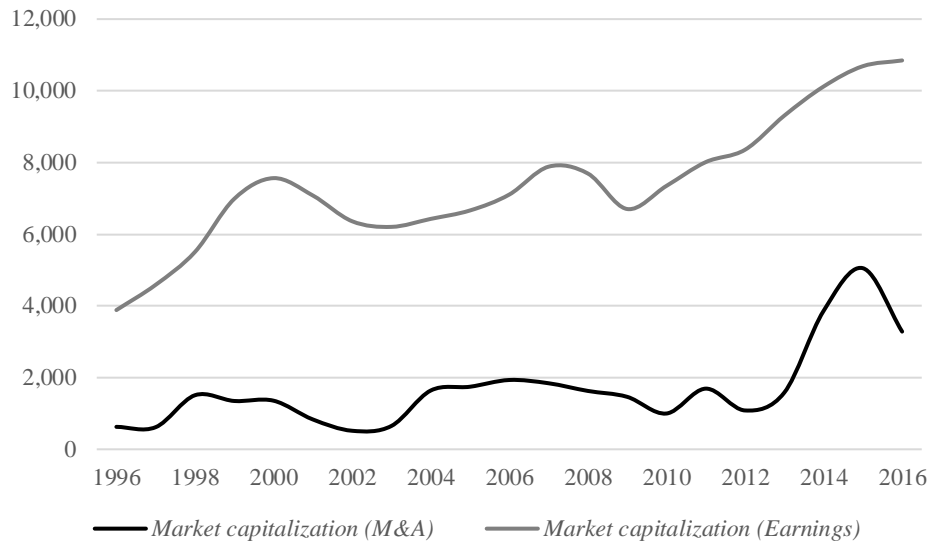


Figure 3. Instrumental variables through time (continued)

Panel C: M&As and earnings – *Market capitalization*



Panel D: *Regulatory budget* and *Future prosecutions*

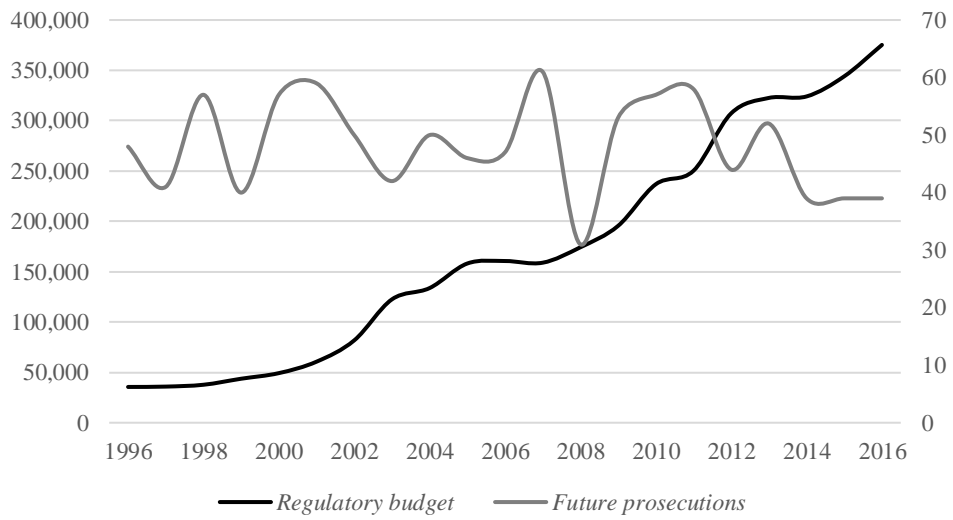
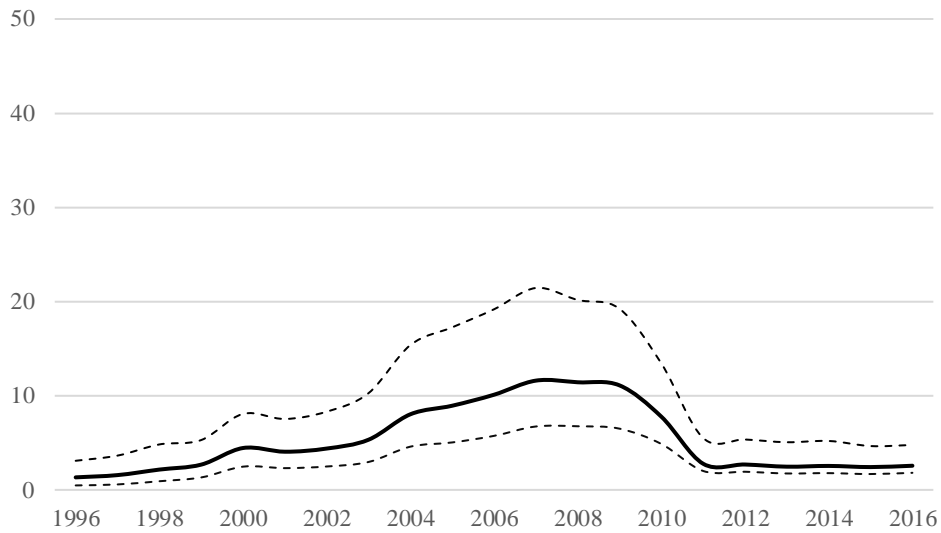


Figure 4. Insider trading and detection rates through time for earnings

This figure illustrates annual estimated rates of insider trading and detection for earnings announcements. Panel A illustrates the estimated prevalence of insider trading (V ; solid black line) and 95% confidence intervals (dashed black lines). Panel B illustrates the estimated probability of detection (D ; solid gray line) and 95% confidence intervals (dashed gray lines), and the percentage of announcements in which insider trading has been detected and prosecuted by the SEC/DOJ ($Detected\%$; dashed black line). Estimates for V and D are obtained from our baseline DCE model using a bootstrap procedure involving 1,000 iterations. Our sample comprises 88 earnings announcements in which insider trading has been detected and prosecuted by the SEC/DOJ and 162,589 earnings announcements in which insider trading has not been detected and prosecuted.

Panel A: Probability of insider trading (V)



Panel B: Probability of detection (D) and percentage detected by the SEC/DOJ ($Detected\%$)

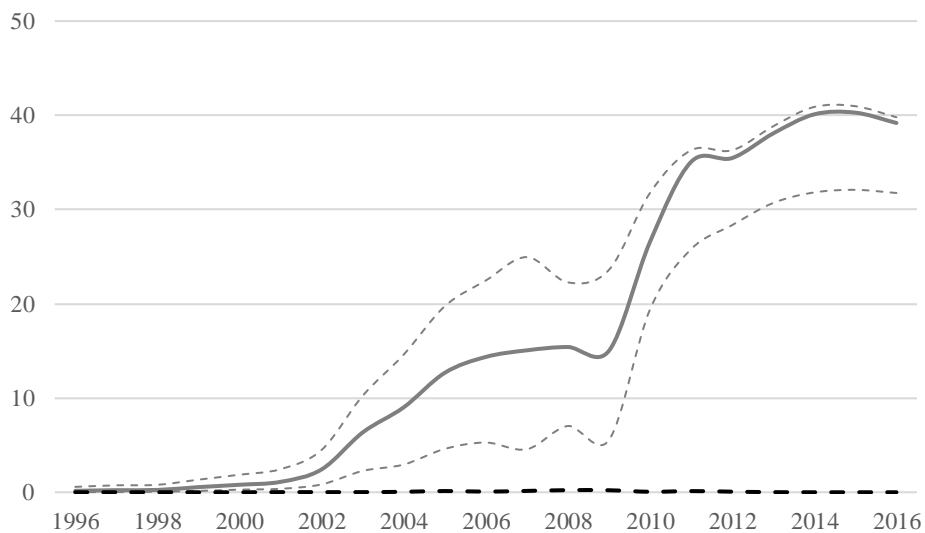


Figure 5. Insider trading versus informal inquiries through time

This figure illustrates the annual estimated prevalence of insider trading for M&A events ($V(M\&A)$; solid black line, left-hand axis), the prevalence of insider trading for earnings announcements ($V(Earnings)$; dashed black line, left-hand axis), and the number of informal inquiries undertaken by the SEC/DOJ ($Informal\ inquiries$; solid gray line, right-hand axis). Estimates for V and D are obtained from our baseline DCE model using a bootstrap procedure involving 1,000 iterations. Our sample comprises 365 (88) M&A (earnings) announcements in which insider trading has been detected and prosecuted by the SEC/DOJ and 6,347 (162,589) M&A (earnings) announcements in which insider trading has not been detected and prosecuted.

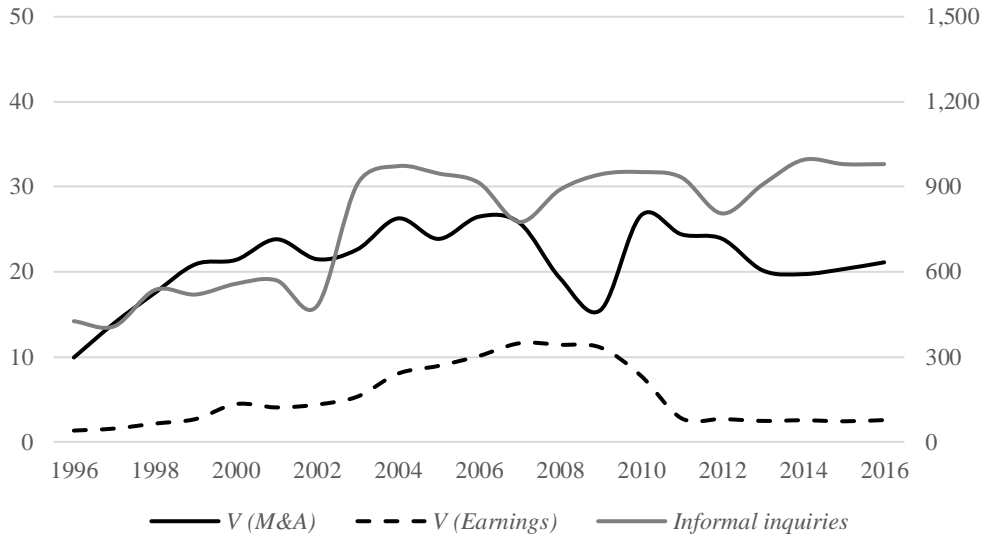
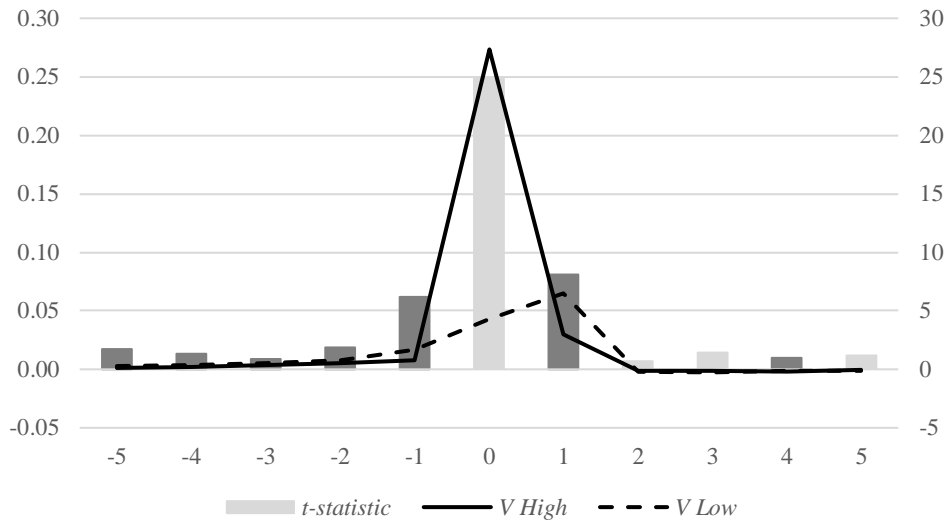


Figure 6. Abnormal returns and volume for M&As

This figure illustrates the mean abnormal behavior of returns (Panel A) and dollar volume (Panel B) in the period $[-5, +5]$ around M&A announcements (left-hand axis). The solid black line (dashed black line) represents M&A announcements in which the estimated probability of insider trading (V) is above (below) the median, $V High$ ($V Low$). Day 0 is the M&A announcement date. On the right-hand axis, shaded light gray (dark gray) bars represent difference of means Satterthwaite t -statistics when mean $V High > V Low$ ($V High \leq V Low$). Estimates for V are obtained from our baseline DCE model using a bootstrap procedure involving 1,000 iterations. Our sample comprises 365 M&A events in which insider trading has been detected and prosecuted by the SEC/DOJ and 6,347 M&A events in which insider trading has not been detected and prosecuted.

Panel A: Abnormal returns



Panel B: Abnormal dollar volume

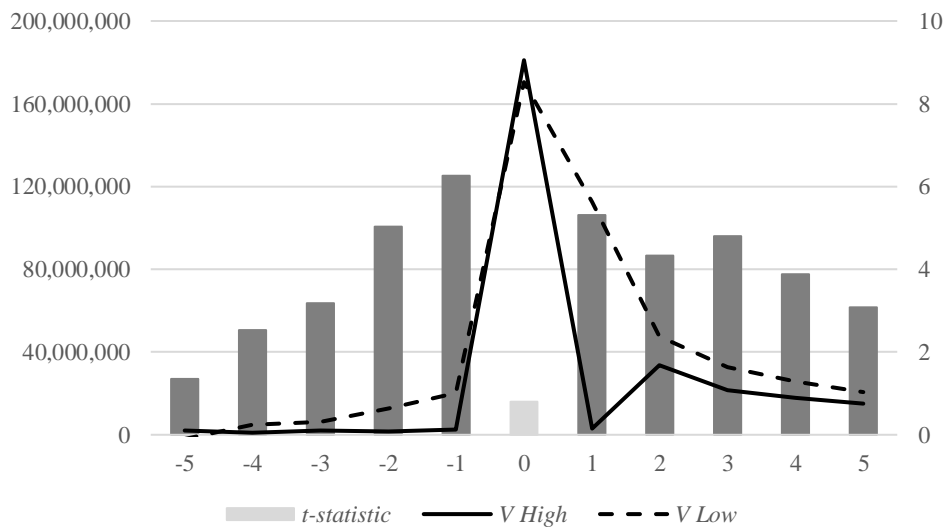
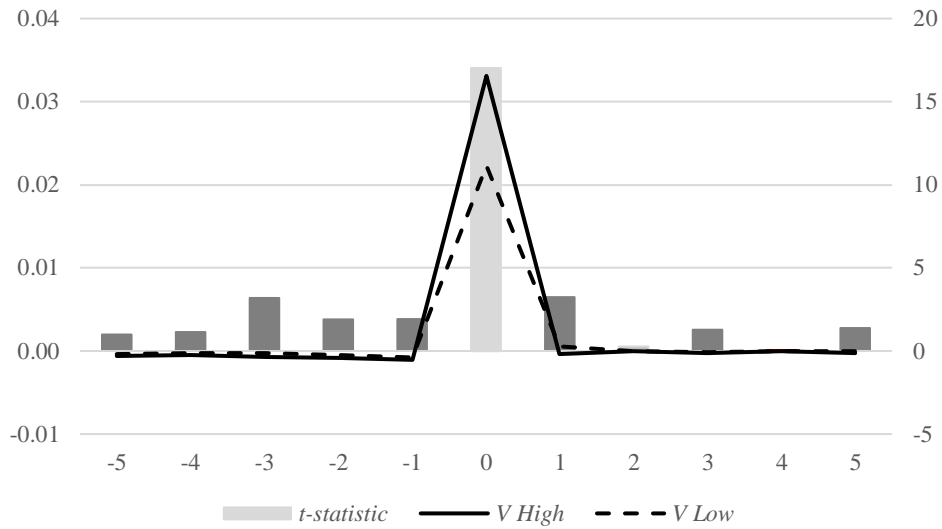


Figure 7. Abnormal returns and volume for earnings

This figure illustrates the mean abnormal behavior of returns (Panel A) and dollar volume (Panel B) in the period $[-5, +5]$ around earnings announcements (left-hand axis). The solid black line (dashed black line) represents earnings announcements in which the estimated probability of insider trading (V) is above (below) the median, $V High$ ($V Low$). Day 0 is the earnings announcement date. On the right-hand axis, shaded light gray (dark gray) bars represent difference of means Satterthwaite t -statistics when $mean V High > V Low$ ($V High \leq V Low$). Estimates for V are obtained from our baseline DCE model using a bootstrap procedure involving 1,000 iterations. Our sample comprises 88 earnings announcements in which insider trading has been detected and prosecuted by the SEC/DOJ and 162,589 earnings announcements in which insider trading has not been detected and prosecuted.

Panel A: Abnormal returns



Panel B: Abnormal dollar volume

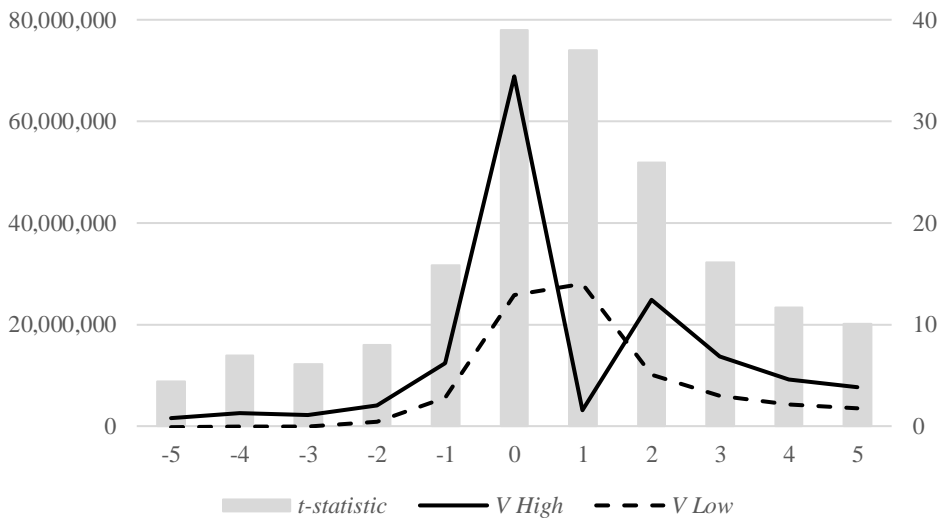


Table 1. Differences in characteristics between announcements with and without detected insider trading

This table reports differences in the mean characteristics for M&A and earnings announcements with and without insider trading prosecuted by the SEC/DOJ: *Detected* and *Not-detected*, respectively. Variable definitions are in Appendix B. Statistical significance for the difference of means is determined by the Satterthwaite *t*-statistic. ***, **, and * indicate statistical significance at 1%, 5%, and 10% levels respectively.

	M&A				Earnings		
	<i>Detected</i> (365)	<i>Not-Detected</i> (6,347)	<i>Difference</i>		<i>Detected</i> (88)	<i>Not-Detected</i> (162,589)	<i>Difference</i>
<i>Turnover</i>	0.010	0.007	0.003***	<i>Turnover</i>	0.018	0.009	0.009***
<i>Illiquidity</i> (x 10 ⁶)	0.272	4.642	-4.369***	<i>Illiquidity</i> (x 10 ⁸)	0.119	1.946	-1.827***
<i>Bid-ask spread</i>	0.008	0.025	-0.017***	<i>Bid-ask spread</i>	0.001	0.005	-0.004***
<i>Market reaction</i> [0]	0.278	0.159	0.119***	<i>Market reaction</i> [0]	0.048	0.029	0.018**
<i>Index stock</i>	0.126	0.057	0.069***	<i>Index stock</i>	0.398	0.249	0.149***
<i>Optionable</i>	0.307	0.146	0.161***	<i>Optionable</i>	0.727	0.409	0.318***
<i>Abnormal returns</i> [-1]	0.016	0.005	0.011***	<i>Abnormal returns</i> [-1]	0.001	-0.001	0.002
<i>Abnormal returns</i> [-3,-1]	0.033	0.013	0.020***	<i>Abnormal returns</i> [-3,-1]	0.003	-0.002	0.005
<i>Abnormal returns</i> [-5,-1]	0.040	0.016	0.024***	<i>Abnormal returns</i> [-5,-1]	-0.001	-0.002	0.001
<i>Abnormal \$volume</i> [-1] (× 10 ⁻³)	17,013.221	10,903.399	6,109.822	<i>Abnormal \$volume</i> [-1] (× 10 ⁻³)	900.034	6,173.233	-5,273.199
<i>Abnormal \$volume</i> [-3,-1] (× 10 ⁻³)	10,541.950	7,225.585	3,316.365	<i>Abnormal \$volume</i> [-3,-1] (× 10 ⁻³)	418.242	2,594.811	-2,176.569
<i>Abnormal \$volume</i> [-5,-1] (× 10 ⁻³)	8,447.213	5,478.417	2,968.796	<i>Abnormal \$volume</i> [-5,-1] (× 10 ⁻³)	291.725	1,537.109	-1,245.384
<i>Future prosecutions</i>	49.740	48.578	1.162**	<i>Future prosecutions</i>	48.648	48.023	0.625
<i>Regulatory budget</i> (× 10 ⁻³)	191.204	139.120	52.084***	<i>Regulatory budget</i> (× 10 ⁻³)	186.508	183.254	3.254
<i>Market capitalization</i> (× 10 ⁻⁶)	2,851.131	1,416.228	1,434.903***	<i>Market capitalization</i> (× 10 ⁻⁶)	2,832.581	754.312	2,078.269***
<i>SEC Whistleblower</i>	0.241	0.185	0.056**	<i>SEC Whistleblower</i>	0.227	0.337	-0.110***
<i>Listing exchange 1 – NYSE</i>	0.343	0.258	0.085***	<i>Listing exchange 1 – NYSE</i>	0.466	0.614	-0.148***
<i>Listing exchange 2 – AMEX</i>	0.030	0.065	-0.035***	<i>Listing exchange 2 – AMEX</i>	0.000	0.015	-0.015***
<i>Listing exchange 3 – NASDAQ</i>	0.627	0.677	-0.050*	<i>Listing exchange 3 – NASDAQ</i>	0.534	0.371	0.163***
<i>Industry 1 – Raw materials</i>	0.036	0.038	-0.002	<i>Industry 1 – Raw materials</i>	0.011	0.062	-0.051***
<i>Industry 2 – Construction/Manufacturing</i>	0.411	0.325	0.086***	<i>Industry 2 – Construction/Manufacturing</i>	0.602	0.380	0.222***
<i>Industry 3 – Services</i>	0.538	0.623	-0.085***	<i>Industry 3 – Services</i>	0.386	0.538	-0.152***
<i>Industry 4 – Other</i>	0.016	0.015	0.001	<i>Industry 4 – Other</i>	0.000	0.019	-0.019***
<i>Bid premium</i>	0.524	0.466	0.058	<i>Earnings surprise</i>	0.020	0.022	-0.002
<i>Number of advisors</i>	2.703	2.182	0.521***				
<i>Number of bidders</i>	1.055	1.107	-0.052***				
<i>Rumored deal</i>	0.090	0.051	0.039***				
<i>Leveraged buy-out deal</i>	0.132	0.093	0.038**				

Table 2. Estimated prevalence of insider trading and detection rate

This table reports estimates of the prevalence of insider trading and its detection. V is the mean probability of insider trading and D is the mean probability of detection. Estimates for V and D and their 95% confidence intervals are obtained from our baseline DCE model using a bootstrap procedure involving 1,000 iterations. Our sample comprises 365 (88) M&A (earnings) announcements in which insider trading has been detected and prosecuted by the SEC/DOJ and 6,347 (162,589) M&A (earnings) announcements in which insider trading has not been detected and prosecuted.

	M&A	Earnings
V (Insider trading)	19.76 (16.17, 22.84)	5.07 (3.12, 8.90)
D (Detection rate)	13.75 (12.02, 15.78)	14.26 (11.10, 16.63)

Table 3. Determinants of insider trading and detection in M&As

This table reports the coefficient estimates and marginal effects of the DCE model, estimated using a bootstrap procedure involving 1,000 iterations. Model 1 is the baseline model. Model 2 is the baseline model without the *Optionable* variable. Model 3 is the baseline model including the following variables: *Bid premium*, *Number of advisors*, *Number of bidders*, *Rumored deal* and *Leveraged buy-out deal*. $V(.)$ is the probability that a given announcement involves insider trading. $D(.)$ is the conditional probability of detection. Variable definitions are in Appendix B. Numbers not in brackets are the coefficient estimates. Numbers in brackets are the marginal effects (partial derivatives of the corresponding probability with respect to each of the variables, reported as a fraction of the estimated corresponding probability). Pseudo R^2 is McFadden's likelihood ratio index (one minus the ratio of the log-likelihood with all predictors and the log-likelihood with intercepts only). ***, **, and * indicate statistical significance at 1%, 5%, and 10% levels respectively, using bootstrapped standard errors.

Table 3. Determinants of insider trading and detection in M&As

Variable	Model 1		Model 2		Model 3	
	V(.)	D(.)	V(.)	D(.)	V(.)	D(.)
<i>Intercept</i>	-1.113*** (-0.718)	-1.489*** (-1.147)	-1.098*** (-0.703)	-1.535*** (-1.160)	-1.418*** (-0.926)	-1.258 (-1.084)
<i>Turnover</i>	0.181*** (0.116)		0.185*** (0.118)		0.230** (0.147)	
<i>Illiquidity</i>	-4.736*** (-3.008)		-4.738*** (-3.006)		-7.352*** (-4.705)	
<i>Bid-ask spread</i>	-0.581*** (-0.369)		-0.582*** (-0.369)		-0.454** (-0.288)	
<i>Optionable</i>	0.006 (0.004)				-0.016 (-0.011)	
<i>Market reaction [0]</i>	1.258*** (0.801)		1.260*** (0.802)		0.896*** (0.574)	
<i>Abnormal returns [-1]</i>		0.342** (0.249)		0.331** (0.250)		0.312* (0.214)
<i>Abnormal \$volume [-1]</i>		0.380*** (0.287)		0.383*** (0.288)		0.327** (0.235)
<i>Future prosecutions</i>		0.222*** (0.165)		0.219*** (0.166)		0.227** (0.161)
<i>Regulatory budget</i>		0.408** (0.294)		0.383*** (0.293)		0.697 (0.422)
<i>Market capitalization</i>	-0.613*** (-0.386)	1.480*** (1.127)	-0.622*** (-0.395)	1.509*** (1.139)	-0.307 (-0.190)	1.221** (0.946)
<i>Listing exchange 1</i>	0.785*** (0.497)	-0.827*** (-0.633)	0.797*** (0.505)	-0.850*** (-0.642)	0.505** (0.317)	-0.518 (-0.407)
<i>Listing exchange 3</i>	0.777*** (0.493)	-0.617*** (-0.465)	0.783*** (0.497)	-0.622*** (-0.470)	0.654*** (0.416)	-0.521*** (-0.379)
<i>Industry 1</i>	-0.333 (-0.209)	0.132 (0.109)	-0.340 (-0.214)	0.144 (0.111)	-0.311 (-0.197)	0.067 (0.056)
<i>Industry 2</i>	-0.259 (-0.164)	0.256 (0.193)	-0.262 (-0.165)	0.251 (0.191)	-0.689 (-0.442)	0.642 (0.408)
<i>Industry 3</i>	-0.566 (-0.358)	0.414 (0.314)	-0.573 (-0.362)	0.444 (0.315)	-0.806* (-0.516)	0.642 (0.416)
<i>SEC Whistleblower</i>	-0.243*** (-0.156)	-0.553 (-0.412)	-0.239*** (-0.153)	-0.548* (-0.414)	-0.374*** (-0.241)	-0.523* (-0.382)
<i>Bid premium</i>					4.371 (2.765)	
<i>Number of advisors</i>					0.041 (0.026)	
<i>Number of bidders</i>					-0.519*** (-0.333)	
<i>Rumored deal</i>					-0.169** (-0.109)	0.133 (0.091)
<i>Leveraged buy-out deal</i>					0.599*** (0.386)	-0.390 (-0.240)
<i>Pseudo R²</i>		35.30%		35.31%		35.80%

Table 4. Determinants of insider trading and detection in earnings announcements

This table reports the coefficient estimates and marginal effects of the DCE model, estimated using a bootstrap procedure involving 1,000 iterations. Model 1 is the baseline model. Model 2 is the baseline model without the *Optionable* variable. Model 3 is the baseline model with the *Earnings surprise* variable. $V(\cdot)$ is the probability that a given announcement involves insider trading. $D(\cdot)$ is the conditional probability of detection. Variable definitions are in Appendix B. Numbers not in brackets are the coefficient estimates. Numbers in brackets are the marginal effects (partial derivatives of the corresponding probability with respect to each of the variables, reported as a fraction of the estimated corresponding probability). Pseudo R^2 is McFadden's likelihood ratio index (one minus the ratio of the log-likelihood with all predictors and the log-likelihood with intercepts only). ***, **, and * indicate statistical significance at 1%, 5%, and 10% levels respectively, using bootstrapped standard errors.

Variable	Model 1		Model 2		Model 3	
	$V(\cdot)$	$D(\cdot)$	$V(\cdot)$	$D(\cdot)$	$V(\cdot)$	$D(\cdot)$
<i>Intercept</i>	-2.355*** (-1.975)	2.563** (1.293)	-2.349*** (-1.970)	2.654** (1.351)	-2.297*** (-1.917)	2.343 (1.175)
<i>Turnover</i>	1.146*** (0.929)		1.105*** (0.895)		1.154*** (0.930)	
<i>Illiquidity</i>	-0.719 (-0.570)		-0.786 (-0.621)		-0.797 (-0.626)	
<i>Bid-ask spread</i>	-1.151*** (-0.944)		-1.171** (-0.956)		-1.154*** (-0.941)	
<i>Optionable</i>	-0.054 (-0.048)				-0.052 (-0.045)	
<i>Market reaction [0]</i>	0.245* (0.204)		0.232** (0.194)		0.244** (0.203)	
<i>Abnormal returns [-1]</i>		-0.323 (-0.086)		-0.134 (-0.059)		-0.175 (-0.077)
<i>Abnormal \$volume [-1]</i>		0.722* (0.337)		0.667* (0.312)		0.693* (0.323)
<i>Future prosecutions</i>		-0.114 (-0.062)		-0.114 (-0.062)		-0.092 (-0.053)
<i>Regulatory budget</i>		5.129* (2.446)		4.958* (2.373)		4.990* (2.385)
<i>Market capitalization</i>	0.373*** (0.311)	2.161* (1.037)	0.341*** (0.284)	2.079* (1.002)	0.375*** (0.310)	2.073* (1.002)
<i>Listing exchange 1</i>	1.771* (1.425)	0.582 (0.311)	1.859* (1.493)	0.504 (0.264)	1.749* (1.399)	0.748 (0.398)
<i>Listing exchange 3</i>	2.034** (1.648)	0.330 (0.208)	2.120** (1.713)	0.267 (0.169)	2.012** (1.621)	0.479 (0.289)
<i>Industry 1</i>	1.902* (1.525)	-2.841*** (-1.419)	1.866* (1.503)	-2.825*** (-1.417)	2.102* (1.675)	-2.782** (-1.388)
<i>Industry 2</i>	2.875*** (2.335)	1.873** (0.929)	2.921*** (2.369)	1.821* (0.904)	2.882*** (2.327)	2.027* (1.025)
<i>Industry 3</i>	2.927*** (2.373)	0.051 (0.056)	2.974*** (2.410)	0.067 (0.060)	2.939*** (2.369)	0.293 (0.187)
<i>SEC Whistleblower</i>	-1.090*** (-0.874)	5.510** (2.884)	-1.119*** (-0.891)	5.725** (3.011)	-1.112*** (-0.884)	5.219** (2.739)
<i>Earnings surprise</i>					-0.094 (-0.077)	
<i>Pseudo R²</i>		70.69%		71.06%		70.65%

Table 5. Robustness tests

This table examines the robustness of our DCE model estimates. V is the mean probability of insider trading, and D is the mean probability of detection. V and D and their 95% confidence intervals are estimated using a bootstrap procedure with 1,000 iterations. Panel A reports the mean V and D (and associated 95% confidence intervals) from the baseline DCE model. Panel B reports the mean V and D (and associated 95% confidence intervals) for the nine specifications in which we relax the exclusion restrictions on individual instrumental variables. For example, in the first specification we estimate the baseline model without the *Turnover* variable, in the second we exclude the *Illiquidity* variable, and so forth. Panel C reports the mean V and D (and associated 95% confidence intervals) from the baseline DCE model without the *Optionable* variable. Panel D reports the mean V and D (and associated 95% confidence intervals) in the eight specifications in which we use alternate definitions for instrumental variables and/or include additional instrumental variables. The eight specifications include: (i) replace *Abnormal \$volume [-1]* with *Abnormal volume [-1]*; (ii) replace *Abnormal returns [-1]* and *Abnormal \$volume [-1]* with *Abnormal returns [-3,-1]* and *Abnormal \$volume [-3,-1]*, respectively; (iii) replace *Abnormal returns [-1]* and *Abnormal \$volume [-1]* with *Abnormal returns [-3,-1]* and *Abnormal volume [-3,-1]*, respectively; (iv) replace *Abnormal returns [-1]* and *Abnormal \$volume [-1]* with *Abnormal returns [-5,-1]* and *Abnormal \$volume [-5,-1]*, respectively; (v) replace *Abnormal returns [-1]* and *Abnormal \$volume [-1]* with *Abnormal returns [-5,-1]* and *Abnormal volume [-5,-1]*, respectively; (vi) replace *Market reaction [0]* with *Market reaction [0,+1]*; (vii) include *Index stock* as an additional instrumental variable; and (viii) for M&As (earnings) include the following additional instrumental variable(s): *Bid premium*, *Number of advisors*, *Number of bidders*, *Rumored deal*, and *Leveraged buy-out deal (Earnings surprise)*. For M&As (earnings), Panel E reports the mean V and D (and associated 95% confidence intervals) from the baseline DCE model augmented with additional instrumental variable(s): *Bid premium*, *Number of advisors*, *Number of bidders*, *Rumored deal*, and *Leveraged buy-out deal (Earnings surprise)*. Variable definitions are detailed in Appendix B. The sample includes 365 (88) M&A (earnings) announcements in which insider trading has been detected and prosecuted by the SEC/DOJ, and 6,347 (162,589) M&A (earnings) announcements in which insider trading has not been detected and prosecuted.

	M&A	Earnings
Panel A		
V	19.76 (16.17, 22.84)	5.07 (3.12, 8.90)
D	13.75 (12.02, 15.78)	14.26 (11.10, 16.63)
Panel B		
V	21.24 (15.33, 25.43)	6.28 (3.21, 10.33)
D	13.53 (11.44, 20.03)	13.97 (11.08, 17.04)
Panel C		
V	19.86 (16.62, 22.76)	5.12 (3.16, 8.69)
D	13.45 (11.96, 15.27)	14.21 (11.39, 16.79)
Panel D		
V	19.14 (9.21, 23.03)	5.36 (3.20, 9.46)
D	14.82 (11.73, 34.56)	14.00 (10.44, 16.47)
Panel E		
V	19.47 (14.38, 23.33)	5.25 (3.17, 9.25)
D	14.17 (11.13, 18.67)	14.15 (10.73, 17.08)