The prevalence and underpinnings of closing price manipulation

Carole Comerton-Forde and Tālis J. Putniņš*
Discipline of Finance, Faculty of Economics and Business, University of Sydney, NSW 2006, Australia

This version: 15 October 2008

Abstract

We empirically analyze the prevalence and economic underpinnings of closing price manipulation and its detection. We find that stocks with high levels of information asymmetry and mid to low levels of liquidity are most likely to be manipulated. A significant proportion of manipulation occurs on month-end and quarter-end days. An increase in government regulatory budgets would increase the rate of prosecution and decrease the amount of manipulation. We estimate that between 1.7% and 1.9% of closing prices are manipulated suggesting that for each prosecuted instance of closing price manipulation there are between 400 and 440 instances that remain undetected or not prosecuted. We also observe differences in manipulation and detection rates across the US and Canadian stock exchanges.

JEL classification: G14

Keywords: manipulation, closing price, high-closing, detection controlled estimation

---

* Corresponding author: Tālis Putniņš Email: T.Putnins@econ.usyd.edu.au; Phone: +61 2 9351 3915; Fax: +61 2 9351 6461.
We thank the Australian Stock Exchange, the Australian Research Council and Securities Industry Research Centre of Asia-Pacific for funding (ARC Linkage Project LP0455536). We thank the Securities Industry Research Centre of Asia-Pacific and Reuters for providing access to data used in this study. We are grateful for the comments of Robert Bartels, Jonathan Batten, Ryan Davies, Jonathan Feinstein, Doug Foster, Richard Gerlach, Petko Kalev, Si Li, Pamela Moulton, James Rydge, Tom Smith, Terry Walter, Tracy Yue Wang and seminar participants at the 2008 Financial Management Association Annual Meeting, the 2008 European Financial Management Association Annual Conference, the University of Sydney 2nd Annual Microstructure Meeting, the University of Technology Sydney and Macquarie University.

1. Introduction

Closing price manipulation, the illegal act of aggressively buying or selling stock at the end of a trading day in order to push the closing price to an artificial level, is detrimental to stock markets and their participants. Manipulation discourages participation and causes investors to trade in alternative markets. This decreases the liquidity of markets known to be manipulated, thereby increasing trading costs. Consequently, manipulation can lead to an increase in the cost of capital, making firms reluctant to list their shares in markets known for manipulation. Manipulation impairs price discovery through reduced order flow and distorts prices from their natural levels. As a result market efficiency is reduced. For these reasons, understanding closing price manipulation is of great importance to academics, exchanges and regulators. The objective of this paper is to analyze what factors drive manipulation and estimate how prevalent it is in equity markets.

People manipulate closing prices for many reasons. Mutual fund net asset values (NAV), and thus fund performance, are often calculated using closing prices. The performance of a fund determines its ranking relative to competitors and is also commonly used as the basis for fund manager remuneration. With these clear incentives it comes as little surprise that some fund managers manipulate closing prices.[1] Closing prices have also been manipulated in order to profit from large positions in derivatives on the underlying stock[2] and by brokers attempting to alter their customers' inference of

[1] This type of manipulation is commonly conducted on the last day of a reporting period such as a month-end or quarter-end. See Carhart et al. (2002), Bernhardt and Davies (2005). This practice is also known as "marking the close", "painting the tape", "high closing", "marking up" or "portfolio pumping".
[2] See, for example, Kumar and Seppi (1992) and Ni et al. (2005).

their execution ability.[3] Manipulation has occurred during pricing periods for seasoned equity issues and takeovers, to maintain a stock's listing on an exchange with minimum price requirements, to avoid margin calls, and on stock index rebalancing days for a stock to gain inclusion in an index.

Our understanding of the pervasiveness and underpinnings of closing price manipulation is limited by the fact that only some non-random fraction of manipulation is detected and prosecuted by market regulators. Closing price manipulation is perceived by market participants to be common, but how common is common? Similarly, many questions regarding the underpinnings of manipulation are unanswered. For example, fund managers have been prosecuted for manipulating closing prices on quarter-ends but is closing price manipulation more likely on quarter-ends than on other days?

The paper makes two main contributions. First, we quantify the extent to which various factors drive manipulation and its detection and second, we estimate the prevalence of manipulation in equity markets. We use a sample of actual manipulation cases and methodology that explicitly takes into consideration that only a non-random subset of manipulation is detected. Our sample of prosecuted manipulation cases is from four US and Canadian stock exchanges - New York Stock Exchange (NYSE), American Stock Exchange (AMEX), Toronto Stock Exchange (TSX) and TSX Venture Exchange (TSX-V).

We find that stocks with high levels of information asymmetry and mid to low levels of liquidity are most likely to be manipulated. A significant proportion of manipulation occurs on month-end and quarter-end days suggesting fund managers are responsible for a considerable fraction of manipulation. We also find that stock price

[3] See, for example, Hillion and Suominen (2004).

volatility deters manipulation by drawing the attention of regulators. We estimate that for each prosecuted instance of closing price manipulation between 400 and 440 instances of manipulation remain undetected or not prosecuted and that this rate differs substantially across exchanges. We also find that larger government regulatory budgets increase the rate of prosecution and significantly deter manipulation. Therefore increased government regulatory budgets are likely to enhance market integrity.

Our estimates of what fraction of manipulation remains undetected are crucial in evaluating the effectiveness of regulation and deciding whether or not current regulatory effort is sufficient in detecting and deterring manipulation. The insights into what drives manipulation have implications for improving the efficiency with which scarce regulatory resources are utilized in detecting and deterring manipulation. This study provides an instrument to calculate the probability of manipulation that is not detected or not prosecuted. It can be used to study undetected manipulation that is otherwise not observable and therefore allows alerting parameters of market surveillance systems to be better defined. Our estimates of the frequency of closing price manipulation and where it is most likely to occur also help quantify how harmful manipulation is to market efficiency and social welfare.

2. Empirical model of manipulation and detection

Many violations of laws and regulations, such as manipulation and insider trading, are not detected and prosecuted. Analyzing the detected fraction can lead to substantial biases in inference about the characteristics or frequency of violations. However, this problem is overlooked or inadequately addressed in much of the empirical literature.

Biases in inference about the characteristics of market misconduct arise when the set of detected cases systematically differ from all violations because of non-random detection. The biases become particularly problematic when some aspect of the detection process is related to what is being examined. For example, Aggarwal and Wu (2006) analyze a sample of "pump-and-dump" manipulation cases litigated by the SEC. In a "pump-and-dump" scheme a manipulator takes a long position in a stock, inflates the price through aggressive trading or by releasing false information and then profits from selling the stock at the inflated price. If cases of manipulation that cause large changes in the price of a stock are more likely to be detected and litigated by the SEC, then the inferences of Aggarwal and Wu about the effect of manipulation on prices or what types of stocks are more likely to be manipulated are potentially significantly biased. The difficulty in estimating the underlying rate of violations when only some fraction is detected is more obvious – if undetected violations cannot be observed, how can one infer what fraction goes undetected?

The econometric problems caused by incomplete detection are well documented by Feinstein (1990, 1991) who develops a detection controlled estimation (DCE) methodology to overcome these problems. This methodology allows inference about undetected violations, which are not directly observable. DCE models have been applied to the regulation of nuclear power plants (Feinstein (1989)), income tax evasion (Feinstein (1990, 1991)), compliance with environmental protection legislation (Brehm and Hamilton (1996), Helland (1998)) and false positives in mammograms (Kleit and Ruiz (2003)). The idea behind DCE is simple – jointly estimating models of the detection and violation processes explicitly allows for the possibility of incomplete detection. In its simplest form, a DCE model is a system of two equations – one modeling violation and the other modeling detection conditional on violation having occurred.

2.1 The model setup

We modify the basic DCE model by representing the detection of closing price manipulation as a two stage process. Therefore our model consists of three stages. The first stage models the probability that the closing price of a particular stock on a particular date is manipulated and the second two stages model the probability that a particular manipulation is detected.

The reason for modeling detection as a two stage process is as follows. Closing price manipulation is often detected by a regulator when price and volume movements trigger 'alerts' in automated market surveillance systems. This is particularly the case when a pattern emerges of several alerts generated from trades made by a particular broker, in a particular stock or on a particular day. We refer to this as *direct* detection. Once a trader has been detected for manipulating prices, further investigation of their trading records often reveals other instances of manipulation, attempted manipulation or conspiring manipulators that would have otherwise remained undetected. Also, some instances of manipulation that do not trigger alerts in market surveillance systems are brought to the attention of the regulator by complaints from market participants. We refer to detection of manipulation that does not trigger price or volume alerts in surveillance systems as *indirect* detection. Examples of indirect detection can be found in our manipulation sample: instances where day-end returns are zero or even negative despite the manipulator's intent to inflate the closing price. These instances represent unsuccessful attempts at manipulation. We model direct and indirect detection separately because their empirical characteristics are quite different. For example, directly detected manipulation is likely to have a large abnormal return on the day of manipulation whereas indirectly detected manipulation will not.

It should be noted that implicit in our model of detection is also prosecution. Although the two processes could be modeled separately such a model is likely to suffer from identification problems due to the lack of observable variables that affect one process but not the other. To separately identify detection and prosecution would require generally unavailable variables such as whether incriminating telephone conversations were recorded or whether incentives and gain to the manipulator can be convincingly demonstrated. Therefore we model detection *and* prosecution as a single process and simply refer to this as detection, consistent with other DCE models in the literature. The limitations of doing so are discussed together with the results.

The propensity for stock-day *i*, that is, a particular stock on a particular day, to have its closing price manipulated is modeled as a continuous latent variable, Y_{1i}^*, that is a function of market-, stock- and time-specific attributes, X_{1i}.

$$Y_{1i}^* = X_{1i}\beta_1 + \varepsilon_{1i} \tag{1}$$

$$Y_{1i} = \begin{cases} 1 & \text{(manipulated)} \\ 0 & \text{(not)} \end{cases} \quad \text{if} \quad \begin{cases} Y_{1i}^* > 0 \\ Y_{1i}^* \leq 0 \end{cases} \tag{2}$$

Y_{1i} is the actual binary variable for whether the closing price of stock-day *i* has been manipulated. Y_{1i} cannot be directly observed if detection is incomplete. Instead, we observe stock-days that have been manipulated *and* detected.

Conditional on manipulation having occurred, the propensity for stock-day *i* to be directly detected by a regulator is modeled as a continuous latent variable, Y_{2i}^{*}, that is a function of market-, stock- and time-specific attributes, X_{2i}.

$$Y_{2i}^{*} = X_{2i}\beta_2 + \varepsilon_{2i} \tag{3}$$

$$Y_{2i} = \begin{cases} 1 \ \text{(directly detected)} & \text{if } Y_{2i}^{*} > 0 \\ 0 \ \text{(not)} & \phantom{\text{if }} Y_{2i}^{*} \le 0 \end{cases} \tag{4}$$

Similarly, Y_{2i} is the actual binary variable for whether the manipulated closing price *i* is directly detected.

Conditional on manipulation having occurred and not being directly detected, the propensity for stock-day *i* to be indirectly detected is modeled as a continuous latent variable, Y_{3i}^{*}, that is a function of market-, stock- and time-specific attributes, X_{3i}.

$$Y_{3i}^{*} = X_{3i}\beta_3 + \varepsilon_{3i} \tag{5}$$

$$Y_{3i} = \begin{cases} 1 \ \text{(indirectly detected)} & \text{if } Y_{3i}^{*} > 0 \\ 0 \ \text{(not)} & \phantom{\text{if }} Y_{3i}^{*} \le 0 \end{cases} \tag{6}$$

Y_{3i} is the actual binary variable for whether the manipulated closing price *i* is indirectly detected.

< INSERT FIGURE 1 HERE >

This model is illustrated graphically in Figure 1. As can be seen, a sample of stock-days falls into two disjoint sets, A and A^c. The first consists of stock-days that have been manipulated and either directly or indirectly detected. The second consists of stock-days that have either not been manipulated or have been manipulated but have escaped both direct and indirect detection. It is important to recognize that Y_{1i}, Y_{2i} and Y_{3i} cannot be separately observed. We observe sets A and A^c and based on these we derive the likelihood for the sample explicitly taking into consideration the unobservability.

This DCE model is similar to Heckman-style models used to deal with selection bias in that both explicitly model the process causing the sample to be a non-random subset of the population in order to allow unbiased estimation. However, Heckman-style models are not suited to application on incomplete detection problems because one of the outcomes of the selection process, undetected or not prosecuted manipulation, cannot be directly observed as it would be in a Heckman-style application (non-respondents in survey data for example).

An assumption implicit in the specification of this model is that error terms for the manipulation, direct detection and indirect detection propensities are not cross-sectionally or serially correlated. Although there are reasons why this may not hold, the likely scenarios leading to violation of this assumption are not expected to cause a significant bias in the parameter estimates (as would be the case in certain other models) and this assumption is consistent with the existing DCE literature. An example of how serial correlation may arise is if a stock's closing price is manipulated over several consecutive days during a pricing period. The fact that our sample is comprised of many stocks with stock specific events such as pricing periods likely to be independent across stocks makes our sample analogous to repeated measures data. The effect of serially correlated errors therefore is expected to "average out" and not bias the estimators (see Kennedy (2003) for example). Cross-sectional correlation of errors may arise if, for example, a fund manager decides to manipulate several stocks in their portfolio on the same day. We account for this and similar scenarios with the inclusion of some specific time dummy variables as explained in the following section.

2.2 Estimation

We use the maximum likelihood method proposed by Poirier (1980) and Feinstein (1990) to estimate this model. We define *M*( ), *D*( ) and *I*( ) to be monotonic link functions that link $X_{1i}\beta_1$, $X_{2i}\beta_2$ and $X_{3i}\beta_3$, to latent probabilities for manipulation, direct detection and indirect detection respectively.[4] That is,

$$M\left(X_{1i}\beta_1\right) = \Pr\left(Y_{1i}=1\right) \tag{7}$$

$$D\left(X_{2i}\beta_2\right) = \Pr\left(Y_{2i}=1|Y_{1i}=1\right) \tag{8}$$

$$I\left(X_{3i}\beta_3\right) = \Pr\left(Y_{3i}=1|Y_{1i}=1,Y_{2i}=0\right) \tag{9}$$

[4] In our implementation of this model the link functions are cumulative logistic distribution functions, that is, $M\left(X_{1i}\beta_1\right) = \frac{1}{1+e^{-X_{1i}\beta_1}}$, $D\left(X_{2i}\beta_2\right) = \frac{1}{1+e^{-X_{2i}\beta_2}}$ and $I\left(X_{3i}\beta_3\right) = \frac{1}{1+e^{-X_{3i}\beta_3}}$. The disturbance terms, ε_{1i}, ε_{2i} and ε_{3i}, are from independent logistic distributions with mean zero and variance $\frac{\pi^2}{3}$ (scale parameter of one). In our robustness tests we examine alternate distributions for the disturbance term.

In order to observe a detected manipulation, the stock-day must have been manipulated and either directly or indirectly detected.[5] Therefore the likelihood that a particular stock-day *i* is from set A, that is, the set of detected manipulation is:

$$L_{A\,i} = M(X_{1i}\beta_1)D(X_{2i}\beta_2) + M(X_{1i}\beta_1)[1-D(X_{2i}\beta_2)]I(X_{3i}\beta_3) \tag{10}$$

The log-likelihood of the entire set of detected manipulation stock-days (set A), is therefore:

$$\log L_A = \sum_{i\in A} \log\{M(X_{1i}\beta_1)D(X_{2i}\beta_2)+M(X_{1i}\beta_1)[1-D(X_{2i}\beta_2)]I(X_{3i}\beta_3)\} \tag{11}$$

Similarly, when we observe a stock-day for which manipulation has not been detected either no manipulation has occurred or manipulation has occurred but the manipulation has escaped both direct and indirect detection. Therefore the likelihood that a particular stock-day *i* has no detected manipulation, that is, the stock-day falls into set A^c is:

$$L_{A^c i} = [1-M(X_{1i}\beta_1)] + M(X_{1i}\beta_1)[1-D(X_{2i}\beta_2)][1-I(X_{3i}\beta_3)] \tag{12}$$

The log-likelihood of the entire set of observations with no detected manipulation (set A^c) is therefore:

$$\log L_{A^c} = \sum_{i\in A^c} \log\{[1-M(X_{1i}\beta_1)]+M(X_{1i}\beta_1)[1-D(X_{2i}\beta_2)][1-I(X_{3i}\beta_3)]\} \tag{13}$$

The log-likelihood of all observations is the sum of the log-likelihoods of each of the two groups. To allow estimation with data collected from endogenous stratified sampling

[5] We make the simplifying assumption of no false detection, that is, the probability of detecting and prosecuting manipulation given that no manipulation has occurred is zero. This seems reasonable considering the strength of evidence required to prosecute closing price manipulators.

(choice-based sampling) we add weights to the observations and the resulting weighted maximum-likelihood estimator (due to Manski and Lerman (1977)) becomes:

$$\log L = w_A \sum_{i \in A} \log\{M(X_{1i}\beta_1)D(X_{2i}\beta_2) + M(X_{1i}\beta_1)[1-D(X_{2i}\beta_2)]I(X_{3i}\beta_3)\}$$
$$+ w_{A^c} \sum_{i \in A^c} \log\{[1-M(X_{1i}\beta_1)] + M(X_{1i}\beta_1)[1-D(X_{2i}\beta_2)][1-I(X_{3i}\beta_3)]\} \qquad (14)$$

where $w_A = \tau / s$, $w_{A^c} = (1-\tau)/(1-s)$ and τ and s are the fractions of stock-days with prosecuted manipulation in the population and sample respectively.

Maximum likelihood estimation maximizes the likelihood of the sample (equation 14) through an iterative process, allowing consistent estimation of the coefficients for the factors that affect manipulation and detection, β_1, β_2 and β_3.

2.3 Alternate models

Identification is a potential problem of all DCE models. Our model decomposes a single datum, detected manipulation, into three components, manipulation, direct detection and indirect detection. The identification issue arises because initially we do not know if this decomposition can be uniquely performed. Intuitively, we need to be able to ascertain whether a sample of cases for which the rate of detected manipulation is low is characterized by a low rate of manipulation or low rates of detection. We also need to ascertain for cases of detected manipulation whether they were directly or indirectly detected.[6]

Identification requires predictor variables uniquely associated with each stage. That is, we need at least one variable that predicts manipulation but not direct or indirect

[6] For a more formal discussion of the identification issue see Feinstein (1990).

detection, at least one variable that predicts direct detection but not manipulation or indirect detection and at least one variable that predicts indirect detection but not manipulation or direct detection. Whilst this condition is theoretically satisfied by our three-equation DCE model we must treat estimates with caution because identification can, in practice, still be a problem. Identification also depends on the strength of the predictors upon which identification is reliant and the amount of variation in the explanatory variables.

We define an alternate two-equation model, similar to the original DCE model used in Feinstein (1989, 1990) because such a model is expected to have fewer problems with identification. This model allows detection to be the result of direct or indirect detection but, in contrast to the three-equation model, it makes no effort to distinguish between the two. The equations and likelihood function for this model are provided in Appendix A.

Another potential issue in DCE models is correlated errors. Brehm and Hamilton (1996) show with Monte Carlo simulations that using an uncorrelated error DCE model when errors are correlated may lead to modest underestimation of the most significant coefficients and increased error around those estimates. A reason why the errors may be correlated is expectations simultaneity that is not incorporated into the model. Two examples of this are a regulator that is more likely to investigate stock-days that have a higher probability of being manipulated or manipulators that choose to manipulate stocks with a lower probability of investigation.

To overcome the potential problem associated with correlated errors and allow for more sophisticated behavior we define a third model with expectations simultaneity.

Based on our three-equation modified DCE model we replace the equations of the propensities for direct and indirect detected with the following:

$$Y_{2i}^{*} = X_{2i}\beta_2 + M(X_{1i}\beta_1)\delta_2 + \varepsilon_{2i} \qquad (15)$$

$$Y_{3i}^{*} = X_{3i}\beta_3 + M(X_{1i}\beta_1)\delta_3 + \varepsilon_{3i} \qquad (16)$$

In this model the probability that a regulator investigates a stock-day for manipulation and therefore the propensity for manipulation to be detected in a stock-day depends on the regulator's assessment of the probability of manipulation, $M(X_{1i}\beta_1)$. The full set of equations and likelihood function are provided in Appendix A.

3. Variables and model specifications

Most variables that we consider primarily influence either manipulation or detection, but also have an indirect effect on the second process due to the interaction between manipulators and regulators. Regulators, to some extent, anticipate the behavior of manipulators and manipulators anticipate the behavior of regulators. For example, if fund managers manipulate closing prices at quarter-ends then a primary determinant of the probability of manipulation is whether or not a day is a quarter-end. A regulator that is aware of this association is more likely to investigate suspicious trading on quarter-end days and therefore whether or not a day is a quarter-end is a secondary determinant of the probability of detecting manipulation. Another example is that if regulatory budget size affects the detection rate, the probability of manipulation will depend on the particular regulatory jurisdiction because manipulators are likely to take into consideration the probability of being caught.

<INSERT TABLE I HERE>

All variables are defined in Table I. We discuss variables according to their primary association - first those associated with manipulation, then detection and lastly both manipulation and detection.

We use market capitalization, turnover, bid-ask spread and closing price as measures of liquidity, although at no stage are all four variables included in a model at the same time. The liquidity of a stock affects how easy it is to manipulate its closing price. Less liquid stocks generally have less depth in the order book and consequently trades have a more substantial price impact. In addition, the manipulator of a low turnover stock has to compete with fewer trades to control the price and is more likely to be successful in making the last trade of the day and setting the closing price. This intuition is consistent with the implication of Hillion and Suominen's (2004) model of closing price manipulation that illiquid stocks will be more frequently manipulated.

We include two variables to measure the degree of information asymmetry in a stock. The first is the number of analysts' forecasts of a stock's earnings and the second is whether or not the stock is included in a broad market index.[7] Allen and Gale (1992) demonstrate the theoretical possibility of profitable stock price manipulation under a rational expectations framework. The basis of their argument is information asymmetry. Investors are uncertain whether a large trader who buys the stock does so because he

[7] Whether or not a stock is a constituent of an index may be associated with effects other than those that occur through information asymmetry. Kumar and Seppi (1992) show that cash settled options (as in the case of futures on indices) provide a profitable manipulation strategy. An Australian prosecution case (Australian Securities Commission v Nomura International PLC – 29 ACSR 473) provides supporting evidence that such a manipulation strategy is possible.

knows it is undervalued or because he intends to manipulate the price. The models of manipulation presented by Kumar and Seppi (1992) and Aggarwal and Wu (2006) are constructed on the same basis. The implication of this is that manipulation is more likely in stocks with higher levels of information asymmetry.

Whilst liquidity and information asymmetry reflect the degree of difficulty in successfully manipulating a stock, we also include variables to capture various motivations for manipulation. Both theoretical and empirical evidence suggests that stock prices are manipulated to profit from options listed on the underlying stock or from futures contracts on indices, particularly in the period immediately prior to expiration.[8] Therefore we include an indicator variable for whether or not a stock has listed options and a second indicator variable for whether such a stock with listed options is in its last trading day prior to expiry of the options.

Fund managers are known to manipulate closing prices at the end of reporting periods such as on the last day of a month or a quarter (Carhart et al. (2002)).[9] Therefore, we include indicator variables for whether or not a day is the last in a month or a quarter. Closing prices are also known to have been manipulated to avoid margin calls and to maintain a stock's listing on an exchange with a minimum price requirement. These incentives for manipulation are more likely to occur in periods in which a stock's price is

[8] Jarrow (1994) demonstrates by example that a derivative security allows for market manipulation trading strategies that would otherwise not exist. Kumar and Seppi (1992) develop a model where the manipulator takes a position in the futures market and then manipulates stock prices at expiry to profit from the futures position. Empirical studies on the effect of expiration-days on the underlying stock prices (Stoll and Whaley (1987), Chamberlain et al. (1989), Stoll and Whaley (1991)) generally find that effects of manipulation can be found in the last hour before expiration and that the price effect is reversed in the first half hour of trading after expiration. Ni et al. (2005) present evidence that on option expiration dates the closing prices of stocks with listed options cluster at option strike prices and attribute this to closing price manipulation.

[9] Bernhardt et al (2007) develop a theoretical model of a mutual fund manager's investment decision and prove that fund managers have incentives to use short-term price impacts to manipulate closing prices at reporting period ends.

decreasing. We include a variable to measure the price trend of stocks to capture these two and other similar motivations related to the price level.

We examine two opposing effects of volatility on manipulation by including the standard deviation of daily returns. Hillion and Suominen (2004) model the behavior of brokers that manipulate closing prices to alter their customers' inference about their execution ability. An implication of their model is that stock price volatility increases the likelihood of manipulation because broker ability is more valuable in the presence of higher volatility. However, it is also possible that volatility attracts the attention of regulators and therefore deters manipulation.

The variables associated primarily with detection and prosecution include government regulator budget, number of filed closing price manipulation prosecutions as well as various indicators of abnormal trading activity that are likely to concern a regulator. Government regulatory budgets, in our case the budgets of the US Securities and Exchange Commission and the Ontario Securities Commission, determine the amount of resources available to conduct investigations and prepare cases for prosecution. Therefore, larger regulatory budgets are likely to be associated with greater capacity to prosecute manipulation. Stock exchanges also have responsibility for manipulation surveillance, so government regulatory budget only measures part of the total amount spent on regulation.[10] The number of filed closing price manipulation prosecutions measures the effectiveness and experience of the regulator in detecting and prosecuting closing price manipulation.

[10] A potential problem with this measure of regulator budget is that it is endogenous, i.e. government regulatory budgets are increased in times or countries where manipulation is more widespread. The consequence of this potential endogeneity is to underestimate regulation's deterrence effect on manipulation.

Closing price manipulation is often detected by a regulator when abnormal trading triggers ‘alerts’ in automated market surveillance systems – described earlier as *direct* detection. The measures of abnormal trading characteristics that we use are abnormal returns, abnormal volume, and price reversion from the closing price to the price the following morning. These trading characteristics are influenced by manipulation for the following reasons. First, the aim of a closing price manipulator is to cause changes in the price of a stock. Second, to affect the closing price, the manipulator trades or releases information which is likely to induce additional trading by speculators or arbitrageurs. Finally, an examination of actual closing price manipulation cases documented in litigation records suggests that manipulation is often carried out by aggressive trading in the last minutes before the close creating liquidity imbalances. Given overnight for new orders to enter the market and resolve the liquidity imbalance, prices often revert towards their original levels the following morning as demonstrated by Carhart et al. (2002).

Many instances of manipulation, however, do not create the abnormal trading characteristics that trigger the alerts of regulators, but may be detected through investigations of other instances of manipulation or investor complaints – discussed earlier as *indirect* detection. Indirect detection is more likely if manipulation is occurring on nearby days in the same stock or in other stocks on the same day as the probability of an alert and subsequent investigation is higher. For this reason we include measures of the aggregate levels of abnormal trading in a stock through time and on a particular day across all stocks as factors that affect the probability of indirect detection. In the two-equation model the direct and indirect detection variables are combined into a single detection equation, thereby overcoming the potential problem of weak identification of direct and indirect detection.

<INSERT TABLE II HERE>

Table II specifies the variables that are used in each of the equations. We address in two ways the fact that many variables influence both manipulation and detection. In models without expectations simultaneity (models one and two) some variables are included in both equations, for example, regulator budget and number of manipulation prosecutions. This is because these variables measure the capacity and effectiveness of the regulator and, at the same time, affect the manipulator's perceived probability of being caught. On the other hand, the measures of abnormal trading that are primary determinants of the probability of detection are not included as determinants of manipulation because the act of manipulation itself influences these variables and therefore their values can only be observed ex-post the manipulation not ex-ante. None of the primary determinants of manipulation are included in the detection equation. In this regard the regulators in the first two models are somewhat naïve in that they do not take advantage of all the information available to them about the determinants of manipulation. In these models the regulator treats all "alerts", that is all observations of abnormal trading, equally rather than devoting more resources to investigating particular alerts, such as on quarter-end days or in illiquid stocks for example.

The third model addresses the secondary associations of variables by incorporating the probability of manipulation as a determinant of the probability of detection. Regulators are modeled as being sophisticated, that is they are aware of the probability of manipulation and use this to alter their detection process. The manipulators, as in the first two models, also take into consideration the government regulator's budget and previous manipulation prosecutions when deciding whether or not to manipulate.

Conceptually, identification of these models can be though of as occurring in the following simplified way. Suppose the larger the day-end price increase associated with a manipulated closing price the greater the probability that manipulation is detected and prosecuted. The sliver of observations with very large day-end price increases have a high probability that any manipulation that is present gets detected and therefore are used to identify the determinants of manipulation. Knowing the determinants of manipulation the sliver of observations that are likely to have been manipulated are used to identify the other factors that influence detection. Of course, this process is not sequential as in this simplified description, but rather, simultaneous.

4. Data

We construct samples of prosecuted closing price manipulation cases (events) and stock-days in which no manipulation is detected or prosecuted (non-events) using endogenous stratified sampling. That is, due to the rare nature of events, we collect all available events and only a fraction of non-events.

We manually collect all of the closing price manipulation cases detected and prosecuted by market regulators in the US and Canada in the period 1 January 1997 to 1 January 2006. We systematically identify the cases from searches of the litigation releases and filings of the market regulators SEC, OSC, RS, IDA, MFDA, IIROC, NYSE

Reg and AMEX DRC[11] and searches of the legal databases Lexis, Quicklaw and Westlaw. We also obtain lists of the case names and filing dates from the appendices of SEC annual reports of all the instances of market manipulation against which the SEC took legal action in the fiscal years 1997 to 2005. We manually examine the litigation releases of each case in this list to identify instances of closing price manipulation. For cases in which insufficient details are provided by the market regulators we obtain court records and filings through the Administrative Office of the US Courts using the PACER service.

We eliminate cases from our sample if insufficient information is available to determine which stocks were manipulated on which days; if they are in over-the-counter markets, are instruments other than common stock, do not involve trade-based techniques, do not have trade and quote data available; or if they do not have at least three months of trading history prior to the start of manipulation.[12] The final sample of detected and prosecuted closing price manipulation is comprised of 160 instances of manipulation with complete data. These 160 instances of a stock being manipulated on a particular day are obtained from six independent manipulation cases, each containing multiple instances of closing price manipulation. These instances of closing price manipulation involve 29 stocks from four exchanges (AMEX, NYSE, TSX, TSX-V). The case names, alleged misconduct and legal outcome are described in Appendix B.

---

[11] The full names of these regulators are US Securities and Exchange Commission (USA), Ontario Securities Commission (Canada), Market Regulation Services Inc. (Canada), Investment Dealers Association (Canada), Mutual Funds Dealers Association (Canada), Investment Industry Regulatory Organization of Canada (Canada), NYSE Regulation Inc. (USA) and AMEX Division of Regulation and Compliance (USA) respectively.

[12] Although cases in which insufficient information is available to determine the manipulated stock and date cannot be included in the manipulation sample, they are included in the population count of prosecuted manipulation. Consequently, these cases affect the model estimation through their influence on the weights in equation 14.

The sample in which manipulation is not detected and prosecuted is obtained by taking for each manipulated stock-day all stock-days on the corresponding exchange in a period of one month up to and including the manipulation date. After eliminating stock-days with missing or erroneous data this sample includes 1,199,777 observations. We obtain stock intra-day trade and quote data as well as the options and index composition data from a *Reuters* database maintained by the *Securities Industry Research Centre of Asia-Pacific* (*SIRCA*). We obtain the remaining data from *Thomson's Datastream* and the websites of the regulators. We apply various normalizing transformations to the data as documented in Table I.

<INSERT TABLE III HERE>

Table III reports the means, standard deviations and medians of the variables for the sample of detected and prosecuted instances of closing price manipulation and the sample in which manipulation is not detected and prosecuted. These summary statistics are intended to provide and indication of the magnitude and dispersion of these variables to allow a quantitative interpretation of the coefficient estimates reported in the following section. As suggested previously, when motivating the use of detection controlled estimation, a simple comparison of the means between the two samples may lead to biases in inference about the effects of these variables on manipulation and detection.

Ignoring these potential biases for now, we notice that the difference in means and medians between the two samples is as expected for most variables. The sample of detected and prosecuted manipulation instances involves less liquid stocks (lower market capitalization, turnover, closing prices and larger spreads), lower levels of institutional following (less analysts forecasts and index constituency), higher volatility and are more often month-end and quarter-end days. The detected manipulation sample is also associated with lower government regulatory budgets, higher abnormal returns, more price reversion and higher abnormal volume, as well as higher aggregate levels of the abnormal trading variables on other days in the same stock.

5. Results

First we present the estimated model coefficients and discuss the determinants of manipulation and detection. Second, we use our models to estimate the frequency of manipulation and detection. Finally, we report results of robustness tests.

5.1 The determinants of manipulation and detection

The log-likelihood functions of the models are maximized to yield estimates of the variable coefficients. Due to the large number of potential variables in each model a stepwise variable selection procedure is used. Starting with just the constant terms, variables that result in the largest log-likelihood increases are added with each iteration and the model re-estimated until additional variables would not yield a significant improvement in the log-likelihood. However, for robustness we also estimate models with alternate sets of variables including those not deemed to be significant by the stepwise procedure. We use various starting values to ensure convergence to a consistent set of parameters.

The estimated coefficients are reported in Table IV. Since the estimated coefficients of probabilistic models are difficult to interpret, we also report the marginal effect of each variable on the dependant variable (probability of manipulation, direct and indirect detection) in brackets.[13] For continuous variables, the marginal effects may be interpreted as the percentage change in the probability of either manipulation, direct or indirect detection for a one percent change in the value of the independent variable.

<INSERT TABLE IV HERE>

Table IV shows that government regulator budget has a strong effect on both manipulation and detection. Across all three models higher government regulator budgets increase in the probability of detecting and prosecuting manipulation and also decrease the probability of manipulation. The latter effect is likely to be because increased regulator capacity has a deterrence effect on manipulation. A 1% increase in a government regulator's real budget per stock is estimated to result in a 3.1% decrease in the amount of closing price manipulation in addition to a 3.0% increase in the rate of prosecution. The former estimate is the average marginal effect across the three models and the later is from model 2 because only this model aggregates detection.

The coefficients of both variables that proxy for the level of information asymmetry (number of analysts forecasts and index stock indicator) suggest that stocks with higher levels of information asymmetry are more likely to be manipulated. A 10% reduction in the number of analysts' forecasts is estimated to increase the probability of

[13] Marginal effects are calculated as $\frac{\partial \Pr}{\partial X} = \frac{\beta^{*} e^{\beta^{*} X}}{(1+e^{\beta^{*} X})^{2}}$, where Pr is the estimated probability of manipulation, direct and indirect detection, β^{*}, are the coefficient estimates and X are the actual observed variable values. They are reported as a percentage of Pr, the estimated probability of manipulation, direct and indirect detection. Marginal effects are calculated for each observation and then averaged over the entire sample.

manipulation by approximately 4%. This finding holds across all three models and these two variables make the largest contribution towards maximizing the likelihood. This suggests that information asymmetry is among the most important determinants of manipulation. This result is consistent with the key assumption of the theoretical models of Allen and Gale (1992), Kumar and Seppi (1992) and Aggarwal and Wu (2006).

The coefficients of the two measures of liquidity in Table IV, market capitalization and turnover, are positive which at first seems like an unexpected result - more liquid stocks have a higher probability of being manipulated. However, this interpretation is not correct. The measures of liquidity are highly correlated with the asymmetric information proxies so highly liquid stocks, by also being low information asymmetry stocks, are already given a low probability of manipulation by the information asymmetry variables. The positive coefficients of the liquidity variables suggest that the most illiquid stocks are not favored by manipulators but, at the same time, unless the magnitude of the effect of liquidity on manipulation is larger than the magnitude of information asymmetry on manipulation, manipulators also do not favor highly liquid stocks. To confirm this interpretation we re-estimate the models replacing the continuous variables, market capitalization and turnover, with quintile dummy variables. We find that the probability of manipulation is highest for stocks in the third and fourth quintiles where the first quintile is defined as having the highest values of market capitalization and turnover and the fifth quintile has the lowest values. Therefore manipulators generally prefer stocks that are slightly less liquid than average but are not among the most illiquid stocks.

A possible explanation of the previous result is that highly liquid stocks are difficult to manipulate because of the high levels of trading activity and depth in the order book and also the associated low levels of information asymmetry. Very illiquid stocks are also not favored by manipulators because they often lack the incentives or scale of potential profits that are present in middle-range and highly liquid stocks. For example, fund managers are, in general, likely to hold relatively liquid stocks and even if they do hold small illiquid stocks they generally represent a small proportion of their portfolios. Therefore it is unlikely that manipulating the closing prices of very illiquid stocks will give fund managers much benefit in overstating their portfolio's performance. Similarly, derivative instruments are less frequently available on very illiquid stocks and these stocks are typically not included in major indices. Finally, when brokers manipulate stocks to alter their customers inference of their execution ability this is more likely to occur for large clients and large orders, which would seldom occur in the most illiquid stocks.

The results in Table IV also show that stocks are significantly more likely to be manipulated on month-end and quarter-end days. Carhart et al. (2002) present evidence that manipulation of stock prices on month-end and quarter-end days is largely attributable to fund managers. Therefore our results suggest that fund managers are responsible for a significant proportion of all manipulation. On the other hand, whether or not a stock has listed options and whether or not a stock with options is trading in the last day before the options expire do not have a particularly strong effect on the probability of manipulation. Similarly, whether a stock's price has been increasing or decreasing over a period of one month (trend) does not appear to have a significant effect on manipulation.

Stock price volatility reduces the likelihood of manipulation. A 10% increase in the standard deviation of daily returns is estimated to decrease the probability of manipulation by 8%. This finding is consistent with the explanation that volatility attracts the attention of regulators and therefore deters manipulation. Hillion and Suominen's (2004) model of brokers that manipulate closing prices to alter their customers' inference about their execution ability predicts that volatility increases the likelihood of manipulation. Our finding is not necessarily inconsistent with Hillion and Suominen because there are many other reasons why people manipulate closing prices and it could be that these dominate the effects of brokers attempting to alter perceptions of their execution ability.

Turning to the variables that affect detection, in all three models the abnormal trading characteristic variables (abnormal return, reversion and abnormal volume) increase the probability of detection. Indirect detection appears to be influenced by abnormal trading, or manipulation, in the same stock in a period of a few days either side. In particular manipulation of stocks that have a number of overnight price reversions in a period of two weeks has an increased probability of being indirectly detected. The regulator notices the abnormal pattern of price reversions and upon investigation reveals instances of manipulation that did not trigger alerts in surveillance systems. However, it appears that abnormal trading in stock cross-section on a particular day does not increase the likelihood of indirect detection of manipulation on that day. One explanation may be that on any particular day, at most, a small proportion of stocks are manipulated and the effect of this is negligible in cross-section. When both direct and indirect detection processes are combined into a single detection process, as in the two-equation model, similar results are observed regarding the effect of the abnormal trading characteristic variables on the probability of detection.

An unexpected result is observed in the third model with regard to the effect of the probability of manipulation, M( ), on the probability of detection. The results suggest that *ceteris paribus*, that is after controlling for things such as the effect of abnormal trading characteristics on detection, the probability of detection decreases as the probability of manipulation increases - although this result has low statistical significance. Viewing the interaction between manipulators and regulators as a strategic game, one way to interpret this result is that manipulators are the more strategic party and that regulators are not taking advantage of all available information. Support for this interpretation is that as previously described, manipulators react significantly to changes in regulatory budget and volatility suggesting some degree of strategic behavior. However, alternate explanations exist such as the implicit role of prosecution in our model of detection. It may be the case that where manipulation is more likely to occur it is more difficult to prosecute and consequently the probability of detection *and* prosecution in such circumstances decreases. For example, the probability of manipulation is higher on month-end days but prosecution of manipulation on such days may be more difficult due to the legitimate reasons to trade at the close on these days.

The exchange indicator variables are included in all equations as controls for differing levels of manipulation and detection in each of the two countries (US and

Canada) or in different exchanges within the one country.[14] Industry indicator variables on the other hand are not included in the final models as they are generally not statistically significant.

5.2 The frequency of manipulation and detection

The three models of manipulation and detection allow estimation of the underlying rate of manipulation (detected and not detected) and the fraction of manipulation that remains undetected. Denoting the parameter estimates by β_1^*, β_2^* and β_3^* an application of Bayes's law for the three-equation models demonstrates that for stock-days with no detected or prosecuted manipulation (set A^c) the probability of an undetected manipulated closing price is:

$$\Pr\left(Y_{1i}=1 \mid Y_{2i}=0, Y_{3i}=0\right) = \frac{M\left(X_{1i}\beta_1^*\right)\left[1-D\left(X_{2i}\beta_2^*\right)-D\left(X_{2i}\beta_2^*\right)I\left(X_{3i}\beta_3^*\right)\right]}{1-M\left(X_{1i}\beta_1^*\right)D\left(X_{2i}\beta_2^*\right)-M\left(X_{1i}\beta_1^*\right)\left[1-D\left(X_{2i}\beta_2^*\right)\right]I\left(X_{3i}\beta_3^*\right)} \quad (17)$$

For the two-equation model this probability is:

$$\Pr\left(Y_{1i}=1 \mid Y_{2i}=0\right) = \frac{M\left(X_{1i}\beta_1^*\right)\left[1-D\left(X_{2i}\beta_2^*\right)\right]}{1-M\left(X_{1i}\beta_1^*\right)D\left(X_{2i}\beta_2^*\right)} \quad (18)$$

These estimates of the probability of manipulation where manipulation has not been prosecuted are useful in efficiently allocating regulators' resources, particularly when resources are increased and it becomes possible to investigate additional cases of

[14] Although their coefficients are highly statistically significant they cannot, in isolation, be used to infer differences in the overall levels of manipulation and detection between the exchanges as much of that is explained by the systematic differences in the explanatory variables across the exchanges. For this reason the levels of manipulation and detection by market are examined separately in the following subsection.

suspected manipulation. These probability estimates can also be used to study the characteristics of closing price manipulation that is not detected.

The fraction of undetected manipulation in the population as a whole can be consistently estimated as:

$$\left(\frac{N}{Tn}\right)\sum_{i\in A^c}\Pr\left(Y_{1i}=1|Y_{2i}=0,Y_{3i}=0\right) \tag{19}$$

where T is the total number of observations in the population (the sum of the number of observations in set A and set A^c), N is the population number of observations in set A^c and n is the sample number of observations in set A^c.

Models 1 and 2 estimate the fraction of undetected closing price manipulation in the population as 1.7% and 1.9% of all stock-days respectively. The rate of detected and prosecuted manipulation in the population (number of observations in set A divided by T) is 0.004%. This suggests that there are many more instances of manipulation not prosecuted than prosecuted manipulations. In fact, these estimates suggest that only about 0.2% of all manipulation is prosecuted. For every prosecuted closing price manipulation approximately 400 to 440 manipulations remain either undetected or not prosecuted. Here the limitation of modeling detection and prosecution together becomes clear – we cannot infer what fraction of the not prosecuted manipulation would have been detected. Adding the rates of prosecuted and not prosecuted manipulation, the underlying rate of manipulation in the population is estimated at 1.7% to 1.9% of stock-days – not significantly different from the rate of manipulation that is not prosecuted due to the negligible fraction of prosecuted manipulation.

<INSERT TABLE V HERE>

Table V presents estimates of the frequency of manipulation and detection by exchange. Some interesting differences between exchanges and countries are observed. These estimates suggest that manipulation is much more pervasive on the NYSE than any of the other exchanges. We find the large magnitude of this difference surprising and therefore conduct further analysis to validate this result. Throughout our sample period trading on the Canadian exchanges occurred with decimal pricing, whereas the US exchanges switched from fractional to decimal pricing within our sample.[15] The change to decimal trading systems affected spreads and liquidity, which in turn affect manipulation and detection. To ensure that the surprisingly large differences in manipulation rates are not caused by effects from the pre-decimalization period we re-estimate the models and rates using only the post-decimalization portion of data. In doing so, we remove from our sample 32 instances of closing price manipulation from the NYSE and AMEX. We find that the previous result, that the NYSE has a higher rate of manipulation, continues to hold. However, the frequency of closing price manipulation on the NYSE and AMEX in the post-decimalization period is lower than in the pre-decimalization period. More precisely, the estimated rates of manipulation (and corresponding multipliers) on the NYSE and AMEX in the post-decimalization period are 0.81% (multiplier of 317) and 0.021% (multiplier of 1.3) respectively, compared to 2.97% and 0.065% for the full sample period. There is no significant change for the Canadian exchanges in the later time period.

[15] All Canadian stock exchanges switched from fractional to decimal trading systems on 15 April 1996, whereas the NYSE and AMEX began phasing in decimal trading systems from 28 August 2000 and completely switched to decimal trading on 29 January 2001.

As additional robustness tests of this result we add country interaction effects with key explanatory variables, add dummy variables for the pre-decimalization period, examine manipulation rate estimates through time and compare the microstructure conditions across the exchanges. We conclude that the rate of manipulation on the NYSE has been declining through time and on average in our sample period is higher than the manipulation rates of the other exchanges.

The two smaller exchanges in each country by market capitalization, AMEX and TSX-V, have an equal or lower rate of closing price manipulation than the corresponding larger exchange. This result is not consistent with Aggarwal and Wu (2006) who conclude that most manipulation occurs more frequently in small and illiquid markets. However, Aggarwal and Wu (2006) use a sample of prosecuted manipulation cases and do not address the bias introduced by not incorporating the detection process in their analysis. It appears that this is the reason why our finding differs and highlights the importance of controlling for detection characteristics. The multiplier reported in Table V, that is the number of manipulation instances that remain undetected or not prosecuted for every prosecuted instance, is considerably smaller for AMEX and TSX-V relative to the corresponding larger exchange. It ranges from about five for the TSX-V to about 871 for the NYSE. So, whilst it may be true that there are more prosecuted manipulation cases in the smaller exchanges, our results suggest that this is because the proportion of manipulation detected in small exchanges is considerably higher. The underlying rate of manipulation is in fact, on average, greater on the larger exchanges.

This finding is consistent with many of the motivations for manipulation as discussed previously (brokers and fund managers generally hold relatively liquid stocks) as well as our finding that the third and fourth, not fifth, quintiles of liquidity are the preferred targets of manipulators.

Further, the results in Table V suggest a difference in the detection rates across the two countries. The Canadian exchanges have detection rates several times greater than the US exchanges. Considering that the budget of the SEC per stock is considerably larger than that of the OSC this suggests that the OSC is more focused or efficient in detecting and prosecuting closing price manipulation.[16]

As a final note, parameter values and estimates of the underlying rates of manipulation and detection should be interpreted cautiously. These are estimates from statistical techniques that rely on certain statistical assumptions. The most important of these, as discussed in Feinstein (1989, 1990, 1991), are the assumptions required to identify manipulation from detection.

5.3 Robustness tests

We examine the robustness of our results to several factors, among the most important of which are changes in the distribution from which the disturbance term is assumed to be drawn. In our initial implementation of the three DCE models the disturbance terms, ε_{1i}, ε_{2i} and ε_{3i}, are assumed to be drawn from independent standard logistic distributions with probability density function $f(\varepsilon) = \frac{e^{-\varepsilon}}{\left(1+e^{-\varepsilon}\right)^2}$. To test the sensitivity of our results to this assumption we estimate the models using four alternate

[16] The same does not apply for other types of violation. Bhattacharya (2006) reports that the SEC prosecutes 10 times more cases per firm for all securities laws violations than the OSC prosecutes.

disturbance term distributions. The alternate distributions are modifications of the standard logistic distribution with fatter tails, thinner tails, a right skew and a left skew.[17]

The model estimates under the alternate disturbance term distributions are reported in Table VI. The marginal effects of the independent variables are very similar under the different disturbance term distributions. Overall this suggests that our results are not overly sensitive to the assumed distribution of the disturbance term.

We also examine the robustness of our results to changes in the sample composition, the time period from which the sample is drawn, different model specifications and alternate variable definitions. To test the sensitivity to the particular sample and time period we split our data into two sub-samples, first by time (earliest half of the data and latest half of the data) and then randomly, and estimate the model separately on each sub-sample. We also re-estimate our model using only post-decimalization data. We test alternate model specifications by including the variables from Table I that are left out of the reported models. We examine the sensitivity of the results to the way the variables are measured by replacing variables with their alternate definitions given in Table I. We find that the main results hold in each of these robustness tests and therefore we do not report these results.

[17] The fat and thin tailed distributions are equal mixtures of a standard logistic distribution and a logistic distribution with larger or smaller scale parameter respectively. Their probability density functions are given by $f(\varepsilon) = \frac{e^{-\varepsilon}}{2\left(1+e^{-\varepsilon}\right)^2} + \frac{e^{-\varepsilon/s}}{2s\left(1+e^{-\varepsilon/s}\right)^2}$ with s=2 for the fat tailed distribution and s=0.5 for the thin tailed distribution. The right and left skew distributions are generalized logistic distributions with probability density $f(\varepsilon) = \frac{be^{-\varepsilon}}{\left(1+e^{-\varepsilon}\right)^{(b+1)}}$ and b=3 for the right skew distribution and b=0.5 for the left skew distribution.

6. Conclusions

Using methodology that explicitly takes into consideration that only detected and prosecuted manipulation is directly observable we examine the determinants of manipulation and its detection. Stocks with high levels of information asymmetry and mid to low levels of liquidity are most likely to be manipulated. The probability of manipulation is higher on month-end and quarter-end days suggesting fund managers account for a significant proportion of manipulation. Stock price volatility deters manipulation by attracting the attention of regulators. Larger government regulatory budgets increase the rate of prosecution and significantly deter manipulation. These insights give a better understanding of closing price manipulation and have important implications for efficiently utilizing scarce regulatory resources.

We also find that only a small fraction of manipulation is detected and prosecuted. For each instance of prosecuted closing price manipulation there are between 400 and 440 instances of manipulation remain undetected or not prosecuted and this rate differs substantially across exchanges. Overall manipulation is more common on larger exchanges but is detected at a significantly higher rate on small exchanges. The Canadian regulator appears to be more efficient at prosecuting closing price manipulation than the US regulator.

The findings of this study highlight the pervasiveness of manipulation relative to the number of prosecuted cases suggesting manipulation is a serious issue for exchanges and regulators. This problem can be reduced by allocating additional resources to regulation. The results suggest that a 1% increase in government regulatory budgets would result in an estimated 3.1% decrease in the amount of closing price manipulation and a 3.0% increase in the rate of prosecution.

Further, this study provides an instrument to calculate the probability of manipulation that has not been prosecuted. This instrument can also be used to study the characteristics of undetected manipulation that are otherwise unobservable and therefore allows alerting parameters of market surveillance systems to be better defined. Our estimates of the frequency of closing price manipulation and its drivers help understand its effect on market efficiency and its social harm, and therefore contribute to the debate on whether such practice should be illegal (Kyle and Viswanathan (2008)). Future research should consider the market and economic impacts of manipulation, for example, how much does manipulation distort prices and what harm is caused by these distorted closing prices?

References

Aggarwal, Rajesh K., and Guojun Wu, 2006, Stock market manipulations, *Journal of Business* 79, 1915-1953.

Allen, Franklin and Douglas Gale, 1992, Stock-price manipulation, *The Review of Financial Studies* 5, 503-529.

Bernhardt, Dan, and Ryan J. Davies, 2005, Painting the tape: Aggregate evidence. *Economics Letters* 89, 306-311.

Bernhardt, Dan, Ryan J. Davies, and Harvey Westbrook, 2007, Smart fund managers? Stupid money?, Unpublished working paper (January 2007), Available at SSRN: http://ssrn.com/abstract=428088.

Bhattacharya, Utpal, 2006, Enforcement and its impact on cost of equity and liquidity of the market, Unpublished working paper (May 2006), Available at SSRN: http://ssrn.com/abstract=952698.

Brehm, John and James T. Hamilton, 1996, Noncompliance in environmental reporting: Are violators ignorant, or evasive, of the law?, *American Journal of Political Science* 40, 444-77.

Carhart, Mark, Ron Kaniel, David Musto, and Adam Reed, 2002, Leaning for the tape: Evidence of gaming behavior in equity mutual funds, *Journal of Finance* 57, 661–693.

Chamberlain, Trevor W., C. Sherman Cheung, and Clarence C.Y. Kwan, 1989, Expiration-day effects of index futures and options: Some Canadian evidence, *Financial Analysts Journal* 45, 67–71.

Feinstein, Jonathan S., 1989, The safety regulation of U.S. nuclear power plants: Violations, inspections, and abnormal occurrences, *The Journal of Political Economy* 97, 115-154.

Feinstein, Jonathan S., 1990, Detection controlled estimation, *Journal of Law and Economics* 33, 233-276.

Feinstein, Jonathan S., 1991, An econometric analysis of income tax evasion and its detection, *Rand Journal of Economics* 22, 14-35.

Helland, Eric, 1998, The enforcement of pollution control laws: Inspections, violations, and self-reporting, *Review of Economics and Statistics* 80, 141-153.

Hillion, Pierre, and Matti Suominen, 2004, The manipulation of closing prices, *Journal of Financial Markets* 7, 351–375.

Jarrow, Robert A., 1994, Derivative security markets, market manipulation, and option pricing theory, *The Journal of Financial and Quantitative Analysis* 29, 241-261.

Kennedy, Peter, 2003, A guide to econometrics, Fifth edition (MIT Press, Cambridge, MA).

Kleit, Andrew N. and James F. Ruiz, 2003, False positive mammograms and detection controlled estimation, *Health Services Research* 38, 1207-1228.

Kumar, Praveen, and Duane J. Seppi, 1992, Futures manipulation with cash settlement, *Journal of Finance* 47, 1485–1502.

Kyle, Albert S. and S. Viswanathan, 2008, How to define illegal price manipulation, forthcoming *American Economic Review Papers and Proceedings*.

Manski, Charles F, and Steven R. Lerman, 1977, The estimation of choice probabilities from choice based samples, *Econometrica* 45, 1977-1988.

Ni, Sophie X., Neil D. Pearson, and Allen M. Poteshman, 2005, Stock price clustering on option expiration dates, *Journal of Financial Economics* 78, 49-87.

Poirier, Dale J., 1980, Partial observability in bivariate probit models, *Journal of Econometrics* 12, 209-217.

Stoll, Hans R., and Robert E. Whaley, 1987, Program trading and expiration-day effects. *Financial Analysts Journal* 43, 16–28.

Stoll, Hans R., and Robert E. Whaley, 1991, Expiration-day effects: What has changed? *Financial Analysts Journal* 47, 58–72.

Appendix A: Alternate models

A1. Two-equation (standard DCE) model of manipulation and detection

Using the same notation as for the three-equation model and omitting much of the explanation the two-equation model of manipulation and detection is as follows.

$$Y_{1i}^{*} = X_{1i}\beta_1 + \varepsilon_{1i} \quad (20)$$

$$Y_{1i} = \begin{cases} 1 \ \text{(manipulated)} & \text{if} \ Y_{1i}^{*} > 0 \\ 0 \ \text{(not)} & Y_{1i}^{*} \le 0 \end{cases} \quad (21)$$

$$Y_{2i}^{*} = X_{2i}\beta_2 + \varepsilon_{2i} \quad (22)$$

$$Y_{2i} = \begin{cases} 1 \ \text{(detected)} & \text{if} \ Y_{2i}^{*} > 0 \\ 0 \ \text{(not)} & Y_{2i}^{*} \le 0 \end{cases} \quad (23)$$

$$M(X_{1i}\beta_1) = \Pr(Y_{1i} = 1) \quad (24)$$

$$D(X_{2i}\beta_2) = \Pr(Y_{2i} = 1 | Y_{1i} = 1) \quad (25)$$

$$\log L_A = \sum_{i \in A} \log\{M(X_{1i}\beta_1) D(X_{2i}\beta_2)\} \quad (26)$$

$$\log L_{A^c} = \sum_{i \in A^c} \log\{[1 - M(X_{1i}\beta_1)] + M(X_{1i}\beta_1)[1 - D(X_{2i}\beta_2)]\} \quad (27)$$

$$\log L = w_A \sum_{i \in A} \log\{M(X_{1i}\beta_1) D(X_{2i}\beta_2)\} + w_{A^c} \sum_{i \in A^c} \log\{[1 - M(X_{1i}\beta_1)] + M(X_{1i}\beta_1)[1 - D(X_{2i}\beta_2)]\} \quad (28)$$

A2. Three-equation model of manipulation and detection with expectations simultaneity

$$Y_{1i}^{*} = X_{1i}\beta_1 + \varepsilon_{1i} \quad (29)$$

$$Y_{1i} = \begin{cases} 1 \ \text{(manipulated)} & \text{if} \ Y_{1i}^{*} > 0 \\ 0 \ \text{(not)} & Y_{1i}^{*} \le 0 \end{cases} \quad (30)$$

$$M(X_{1i}\beta_1) = \Pr(Y_{1i} = 1) \quad (31)$$

$$Y_{2i}^{*} = X_{2i}\beta_2 + M(X_{1i}\beta_1)\delta_2 + \varepsilon_{2i} \quad (32)$$

$$Y_{2i} = \begin{cases} 1 \ \text{(directly detected)} & \text{if} \ Y_{2i}^{*} > 0 \\ 0 \ \text{(not)} & Y_{2i}^{*} \le 0 \end{cases} \quad (33)$$

$$D\left(X_{2i}\beta_2, M\left(X_{1i}\beta_1\right)\delta_2\right) = \Pr\left(Y_{2i}=1 | Y_{1i}=1\right) \tag{34}$$

$$Y_{3i}^{*} = X_{3i}\beta_3 + M\left(X_{1i}\beta_1\right)\delta_3 + \varepsilon_{3i} \tag{35}$$

$$Y_{3i} = \begin{cases} 1 \ \text{(indirectly detected)} \\ 0 \ \text{(not)} \end{cases} \text{if} \begin{cases} Y_{3i}^{*} > 0 \\ Y_{3i}^{*} \leq 0 \end{cases} \tag{36}$$

$$I\left(X_{3i}\beta_3, M\left(X_{1i}\beta_1\right)\delta_3\right) = \Pr\left(Y_{3i}=1 | Y_{1i}=1, Y_{2i}=0\right) \tag{37}$$

Representing $M\left(X_{1i}\beta_1\right)$ by M , $D\left(X_{2i}\beta_2, M\left(X_{1i}\beta_1\right)\delta_2\right)$ by D and $I\left(X_{3i}\beta_3, M\left(X_{1i}\beta_1\right)\delta_3\right)$ by I :

$$\log L_A = \sum_{i \in A} \log\left\{MD + M\left[1-D\right]I\right\} \tag{38}$$

$$\log L_{A^c} = \sum_{i \in A^c} \log\left\{\left[1-M\right] + M\left[1-D\right]\left[1-I\right]\right\} \tag{39}$$

$$\log L = w_A \sum_{i \in A} \log\left\{MD + M\left[1-D\right]I\right\} + w_{A^c} \sum_{i \in A^c} \log\left\{\left[1-M\right] + M\left[1-D\right]\left[1-I\right]\right\} \tag{40}$$

Appendix B: Summary of manipulation cases

Case	Exchange	Alleged misconduct	Outcome
In the Matter of Schultz Investment Advisors, Inc. and Scott Schultz, Administrative Proceeding File No. 3-12136	NYSE	Closing price manipulation by a fund manager at the ends of reporting periods. This was done to inflate reported performance and subsequently collect more management fees.	Settlement, fine and suspension.
In the Matter of Spear, Leeds & Kellogg, L.P., Administrative Proceeding File No. 3-11189 / In the Matter of Baron Capital Inc. et al., Administrative Proceeding File No. 3-11096	NYSE	Closing price manipulation by a substantial shareholder during the pricing-period for a company acquisition.	Settlement and fines.
SEC v. Competitive Technologies, Inc. et al., Civil Action No. 304 CV 1331 JCH (District of Connecticut)	AMEX	A prolonged multi-faceted scheme carried out by several brokers, former brokers and company CEO in an attempt for the manipulators to enrich themselves and avoid margin calls.	Conviction by a federal jury and settlements.
In the Matter of RT Capital Management Inc. et al.	TSX	Closing price manipulation of several stocks by several fund managers at the ends of reporting periods to inflate reported performance. This enabled the fund to collect more management fees and earned the fund managers greater remuneration.	Settlement, fines and suspensions.
In the Matter of John Andrew Scott (OOS 2003-010) / In the Matter of Linda Grace Malinowski (OOS 2003-011) / In the Matter of Matthew Philip Linden (OOS 2003-012)	TSX	Manipulation of closing prices over a period of consecutive days in relation to an application for price protection.	Settlements, fines and suspensions.
In the Matter of Alfred Simon Gregorian (DN 2006-003) / In the Matter of Research Capital Corporation (DN 2006-005)	TSX-V	Manipulation of the closing prices of a single stock over a period of time to create a misleading appearance of strength and stability in the market for the company's shares.	Settlement, fines and suspensions.

Table I
Definitions of variables

This table defines the variables used in the models of manipulation and detection. The third column reports the transformation that is applied to the raw data to normalize and scale the variables. Where required by the transformation negative values are multiplied by negative one before and after applying the transformation.

Variable	Definition	Transform
Panel A: Variables associated with both manipulation and detection		
Detected manipulation	Indicator variable for any of the 160 instances of detected and prosecuted closing price manipulation described in the section 4. Detected manipulation sample	
Exchange	Four indicator variables for each of the exchanges American Exchange (AMEX), New York Stock Exchange (NYSE), Toronto Stock Exchange (TSX), TSX Venture Exchange (TSX-V).	
Industry	Ten indicator variables based on the Industry Classification Benchmark (ICB) to represent the industry of a stock. The industries are (1) oil and gas, (2) basic materials, (3) industrials, (4) consumer goods, (5) health care, (6) consumer services, (7) telecommunications, (8) utilities, (9) financials and (10) technology.	
Panel B: Variables associated primarily with manipulation		
Market capitalization	Share price multiplied by the number of ordinary shares in issue. The amount on issue is updated whenever new tranches of stock are issued or after a capital change. Calculated on the first day of each month in millions of US dollars.	$\log(\cdot)$
Turnover	Median number of trades per day in the stock in the previous month. Alternative definitions used for robustness tests:	$\ln(\cdot)$
	Turnover (2) - median daily traded US dollar volume in previous month.	$\log(\cdot)$
	Turnover (3) - median daily US dollar volume traded in previous month divided by market capitalization.	$\log(\cdot)$
Spread	The median of the past month's daily mean proportional spreads for that stock. Daily mean proportional spreads are calculated as the equal weighted mean of the difference between the best bid and ask divided by the bid-ask midpoint price at every quote update and trade.	$\sqrt{\cdot}$
Closing price	The price of the last trade before the market closes at 16:00. The closing time is adjusted on days where the close is delayed.	$\sqrt[3]{\cdot}$
Institutional	Institutional following defined as the total number of IBES analyst forecasts of that financial year's earning s per share (EPS). Calculated on the first day of each month in number of forecasts.	$\sqrt{\cdot}$
Index stock	Indicator variable for whether a stock is a constituent of either of the indices Standard and Poor's (S&P) 500 or S&P/TSX Composite Index (TSE 300 Index prior to May 2002). Calculated on the first day of each month.	
Optionable	Indicator variable for whether the stock has options trading on it that expire within a month.	
Option expiry	Indicator variable for whether an optionable stock is in its last day of trade before options on that stock expire.	
Trend	Close to close return over previous calendar month.	$\sqrt[3]{\cdot}$
Month-end	Indicator variable for the last trading day of a month.	
Quarter-end	Indicator variable for the last trading day of a quarter, that is, the last trading days in the months of March, June, September and December.	
Volatility	Standard deviation of daily returns calculated from closing prices over the previous month.	$\sqrt{\cdot}$

Table I (continued)

Variable	Definition	Transform
	Panel C: Variables associated primarily with detection	
Prosecutions	Number of closing price manipulation prosecutions filed by the market regulators in that country in the previous year (rolling one year window) based on the date of filling the statement of allegations.	
Regulator budget	Budget of the principal government regulator divided by the number of common stocks for which the regulator is responsible. The principal regulator for AMEX and NYSE is the US Securities and Exchange Commission (SEC) and for TSX and TSX-V it is the Ontario Securities Commission (OSC). Budgets are taken from the annual reports of the regulators for each regulator's financial year, deflated by the OECD published CPI of the corresponding country and converted to US dollars. The units of this variable are '00,000s of US dollars in real (August 1998) terms per common stock.	
Abnormal return (AR)	Abnormal day-end return calculated as return from bid-ask midpoint 30 minutes before close to closing price (or in the absence of any trades in the last 30 minutes then midpoint at time of last trade to closing price) less that stock's previous month's median value. Alternative definitions used for robustness tests: AR2 - as per AR1 but using last 60 minutes of trading in place of 30. AR3 - abnormal daily return calculated as close to close return less that stock's previous month's median value.	$\sqrt{\cdot}$ $\sqrt{\cdot}$ $\sqrt{\cdot}$
Reversion (RV)	Overnight price reversion calculated as return from closing price to next morning's 11am bid-ask midpoint price.	$\sqrt{\cdot}$
Abnormal volume (AV)	Abnormal day-end dollar volume, V, relative to benchmark daily traded dollar volume. Calculated as ((V/Turnover2)*100) where V is calculated as traded US dollar volume in the last 30 minutes before close less that stock's previous month's median value. Turnover2 is the median daily traded US dollar volume in that stock in the previous month. Alternative definitions used for robustness tests: AV2 - as per AV but using last 60 minutes of trading in place of 30. AV3 - abnormal daily volume calculated as per AV but using daily traded dollar volume in place of the last 30 minute traded dollar volume.	$\sqrt[4]{\cdot}$ $\sqrt[4]{\cdot}$ $\sqrt[4]{\cdot}$
AR time-series	Abnormal day-end return aggregated over a period of time for a particular stock. Calculated as median value of AR for that stock in a two week period starting seven days back in time and ending seven days forward in time. Alternative definitions used for robustness tests: AR2 time-series – as per AR time-series but using AR2 in place of AR. AR3 time-series – as per AR time-series but using AR3 in place of AR.	$\sqrt{\cdot}$ $\sqrt{\cdot}$ $\sqrt{\cdot}$
RV time-series	Reversion aggregated over a period of time for a particular stock. Calculated as median value of RV for that stock in a two week period starting seven days back in time and ending seven days forward in time.	$\sqrt{\cdot}$
AV time-series	Abnormal day-end volume aggregated over a period of time for a particular stock. Calculated as median value of AV for that stock in a two week period starting seven days back in time and ending seven days forward in time. Alternative definitions used for robustness tests: AV2 time-series – as per AV time-series but using AV2 in place of AV. AV3 time-series – as per AV time-series but using AV3 in place of AV.	$\sqrt[4]{\cdot}$ $\sqrt[4]{\cdot}$ $\sqrt[4]{\cdot}$
AR cross-section	Abnormal day-end return aggregated in stock cross-section. Calculated as median value of AR for all stocks on the corresponding exchange on that day. Alternative definitions used for robustness tests: AR2 cross-section – as per AR cross-section but using AR2 in place of AR. AR3 cross-section – as per AR cross-section but using AR3 in place of AR.	
RV cross-section	Reversion aggregated in stock cross-section. Calculated as median value of RV for all stocks on the corresponding exchange on that day.	
AV cross-section	Abnormal day-end volume aggregated in stock cross-section. Calculated as median value of AV for all stocks on the corresponding exchange on that day. Alternative definitions used for robustness tests: AV2 cross-section – as per AV cross-section but using AV2 in place of AV. AV3 cross-section – as per AV cross-section but using AV3 in place of AV.	$\sqrt{\cdot}$ $\sqrt{\cdot}$ $\sqrt{\cdot}$

Table II
Specification of models

This table defines which variables are used in each of the equations for the three models. Model 1 is a three-equation modified detection controlled estimation (DCE) model, Model 2 is a standard two-equation DCE model and Model 3 is a modified three-equation DCE model with expectations simultaneity. M( ) is the probability of manipulation, D( ) is the conditional probability of direct detection (conditional probability of detection in the standard two-equation DCE model) and I( ) is the conditional probability of indirect detection. Variables are defined in Table I. The symbol + indicates a variable is included as a factor in the corresponding probability.

Variable	Model 1			Model 2		Model 3		
	M( )	D( )	I( )	M( )	D( )	M( )	D( )	I( )
Exchange	+	+	+	+	+	+	+	+
Industry	+	+	+	+	+	+	+	+
Market capitalization	+			+		+		
Turnover	+			+		+		
Spread	+			+		+		
Closing price	+			+		+		
Volatility	+			+		+		
Institutional	+			+		+		
Index stock	+			+		+		
Optionable	+			+		+		
Option expiry	+			+		+		
Trend	+			+		+		
Month-end	+			+		+		
Quarter-end	+			+		+		
Prosecutions	+	+	+	+	+	+	+	+
Regulator budget	+	+	+	+	+	+	+	+
Abnormal return (AR)		+			+		+	
Reversion (RV)		+			+		+	
Abnormal volume (AV)		+			+		+	
AR time-series			+		+			+
RV time-series			+		+			+
AV time-series			+		+			+
AR cross-section			+		+			+
RV cross-section			+		+			+
AV cross-section			+		+			+
M( )							+	+

Table III
Summary statistics

This table reports summary statistics for variables used in the model of manipulation and detection. The variables are defined in Table I. *Raw data* are actual observed values whereas *Normalized and scaled* are values after applying normalizing transformations to the variables. *Detected manipulation* refers to the sample of stock-days in which manipulation has been detected and prosecuted by a regulator (Yes) and the sample of stock-days without detected and prosecuted manipulation (No). Medians and standard deviations (*Std dev*) are not reported for dichotomous variables.

Variable	Detected manipulation	Raw data			Normalized and scaled		
		Mean	Std dev	Median	Mean	Std dev	Median
Panel A: Variables associated with both manipulation and detection							
Exchange (AMEX)	Yes	0.18			0.18		
	No	0.16			0.16		
Exchange (TSX)	Yes	0.54			0.54		
	No	0.18			0.18		
Exchange (TSX-V)	Yes	0.09			0.09		
	No	0.03			0.03		
Exchange (NYSE)	Yes	0.19			0.19		
	No	0.64			0.64		
Panel B: Variables associated primarily with manipulation							
Market capitalization	Yes	475	1,642	70.2	2.08	0.54	1.85
	No	3,215	13,331	290	2.47	0.98	2.46
Turnover (2)	Yes	928,255	4,605,416	77,196	4.91	0.63	4.89
	No	7,620,469	28,043,541	249,563	5.48	1.27	5.40
Spread	Yes	2.83	1.85	2.54	1.59	0.55	1.59
	No	1.99	3.45	0.70	1.12	0.86	0.84
Closing price	Yes	11.5	13.6	6.89	2.04	0.66	1.90
	No	19.1	23.7	14.21	2.37	0.89	2.42
Institutional	Yes	1.65	4.56	0.00	0.62	1.13	0.00
	No	3.87	5.95	1.00	1.30	1.48	1.00
Trend	Yes	4.33	21.8	2.15	0.34	2.30	1.29
	No	0.29	17.1	0.53	0.10	2.03	0.81
Volatility	Yes	4.22	2.50	3.80	1.97	0.59	1.95
	No	3.29	3.58	2.27	1.65	0.75	1.51
Index stock	Yes	0.10			0.10		
	No	0.16			0.16		
Optionable	Yes	0.10			0.10		
	No	0.30			0.30		
Option expiry	Yes	0.00			0.00		
	No	0.01			0.01		
Month-end	Yes	0.37			0.37		
	No	0.05			0.05		
Quarter-end	Yes	0.23			0.23		
	No	0.02			0.02		

Table III (continued)

Variable	Detected manipulation	Raw data			Normalized and scaled		
		Mean	Std dev	Median	Mean	Std dev	Median
Panel C: Variables associated primarily with detection							
Prosecutions	Yes	0.39	0.66	0.00	0.39	0.66	0.00
	No	0.80	1.04	0.00	0.80	1.04	0.00
Regulator budget	Yes	61.9	77.6	6.83	0.62	0.78	0.07
	No	147	85.5	165	1.47	0.86	1.65
Abnormal return (AR)	Yes	1.21	2.16	0.83	0.71	1.11	0.91
	No	0.03	2.21	0.00	0.01	0.97	0.00
Reversion (RV)	Yes	1.66	4.10	1.74	0.93	1.49	1.32
	No	-0.12	3.42	0.00	-0.03	1.34	0.00
Abnormal volume (AV)	Yes	78.0	199	13.2	1.65	1.82	1.91
	No	10.1	82.5	0.00	0.33	1.42	0.00
AR time series	Yes	0.23	0.79	0.16	0.21	0.72	0.40
	No	0.01	1.07	0.00	0.00	0.61	0.00
RV time series	Yes	1.24	1.52	1.11	0.81	0.93	1.06
	No	-0.10	1.64	0.00	-0.03	0.87	0.00
AV time series	Yes	25.2	141	0.00	0.42	1.60	0.00
	No	0.98	8.10	0.00	0.13	1.00	0.00
AR cross-section	Yes	-0.01	0.09	0.00	-0.01	0.09	0.00
	No	0.00	0.07	0.00	0.00	0.07	0.00
RV cross-section	Yes	-0.29	0.54	-0.18	-0.29	0.54	-0.18
	No	-0.06	0.49	0.00	-0.06	0.49	0.00
AV cross-section	Yes	0.86	2.63	0.00	0.34	0.89	0.00
	No	0.20	1.66	0.00	0.15	0.68	0.00
Observations	Yes	160					
	No	1,199,777					

Table IV
Model estimation

This table reports the results of the estimation of the models. Model 1 is a three-equation modified DCE model, Model 2 is a standard two-equation DCE model and Model 3 is a modified three-equation DCE model with expectations simultaneity. M( ) is the probability of manipulation, D( ) is the conditional probability of direct detection (detection in the standard two-equation DCE model) and I( ) is the conditional probability of indirect detection. Variables are defined in Table I. 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 percentage of the estimated corresponding probability). Significance at the 10%, 5% and 1% levels is indicated by *, ** and *** respectively.

	Model 1			Model 2		Model 3		
Variable	M( )	D( )	I( )	M( )	D( )	M( )	D( )	I( )
Constant	-2.19*	-11.1***	-333***	1.61	-13.7***	-2.86***	-9.66***	-329***
Regulator budget	-2.33***	1.50***	71.1**	-3.88***	3.04***	-3.32***	1.82***	71.7***
	(-2.29)	(1.47)	(68.0)	(-3.80)	(2.85)	(-3.31)	(1.77)	(68.8)
Institutional	-0.375***			-0.398***		-0.386***		
	(-0.369)			(-0.390)		(-0.384)		
Index stock	-0.970***			-0.910***		-1.11***		
	(-0.953)			(-0.893)		(-1.10)		
Market capitalization	0.854***			0.624***		0.886***		
	(0.840)			(0.612)		(0.882)		
Turnover (2)	0.276***			0.134		0.341***		
	(0.272)			(0.132)		(0.339)		
Month-end	1.64***			1.56***		1.61***		
	(1.61)			(1.53)		(1.60)		
Quarter-end	1.97***			2.13***		2.34***		
	(1.93)			(2.09)		(2.33)		
Volatility	-0.823***			-0.626***		-0.864***		
	(-0.809)			(-0.614)		(-0.860)		
Abnormal return (AR)		0.822***			0.544***		0.822***	
		(0.809)			(0.510)		(0.803)	
Reversion (RV)		0.234***			0.175***		0.225***	
		(0.230)			(0.164)		(0.220)	
Abnormal volume (AV)		0.950***			0.780***		1.02***	
		(0.934)			(0.731)		(1.00)	
AR time series			3.84***		0.012			1.52**
			(3.67)		(0.0120)			(1.45)
RV time series			59.9***		1.47***			54.1***
			(57.3)		(1.38)			(51.9)
AV time series			-7.72***		0.117*			-7.07***
			(-7.39)		(0.110)			(-6.79)
Exchange (AMEX)	-4.40***	5.06***	178***	-5.22***	5.80***	-2.51***	2.38***	178***
	(-4.32)	(4.98)	(170)	(-5.12)	(5.44)	(-2.50)	(2.32)	(171)
Exchange (TSX)	-5.87***	5.91***	242***	-9.68***	10.6***	-5.47***	4.35***	247***
	(-5.77)	(5.81)	(232)	(-9.49)	(9.95)	(-5.45)	(4.24)	(237)
Exchange (TSX-V)	-6.27***	-0.015	263***	-8.85***	9.51***	-5.60***	-2.23	261***
	(-6.16)	(-0.0152)	(251)	(-8.68)	(8.92)	(-5.57)	(-2.18)	(250)
M( )							-7.57**	-9.98
Observations		1,199,937		1,199,937			1,199,937	
Log likelihood		-2,659		-2,713			-2,642	

Table V
Estimated frequency of manipulation and detection by exchange

This table reports estimates of the frequency of manipulation and detection from Model 1 (three-equation modified DCE model). *NYSE* is the New York Stock Exchange, *AMEX* is the American Stock Exchange, *TSX* is the Toronto Stock Exchange and *TSX-V* is the TSX Venture Exchange. *Fraction detected* and *Fraction undetected* are the fraction of detected closing price manipulation and the estimated fraction of undetected or not prosecuted manipulation in the population respectively. *Multiplier* is *Fraction undetected* divided by *Fraction detected* and estimates the number of manipulations that remain undetected or not prosecuted for every prosecuted manipulation. *Manipulation rate* is the sum of *Fraction detected* and *Fraction undetected.*

Exchange	Fraction detected	Fraction undetected	Multiplier	Manipulation rate
NYSE	0.0034%	2.96%	871	2.97%
AMEX	0.0066%	0.058%	9	0.065%
TSX	0.0058%	0.094%	16	0.100%
TSX-V	0.0041%	0.019%	5	0.023%

Table VI
Robustness tests

This table reports the results of Model 1 (the three-equation modified DCE model from Table V) estimated under alternate distributions for the disturbance term. *Fat tails* and *Thin tails* are equal mixtures of a standard logistic distribution and a logistic distribution with larger or smaller scale parameter respectively. *Right skew* and *Left skew* are generalized logistic distributions with values of the skew parameter set to produce right and left skew distributions respectively. M( ) is the probability of manipulation, D( ) is the conditional probability of direct detection and I( ) is the conditional probability of indirect detection. Variables are defined in Table I. 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 percentage of the estimated corresponding probability). Significance at the 10%, 5% and 1% levels is indicated by *, ** and *** respectively.

	Fat tails			Thin tails			Right skew			Left skew		
Variable	M( )	D( )	I( )	M( )	D( )	I( )	M( )	D( )	I( )	M( )	D( )	I( )
Constant	0.545	-22.6***	-334***	-1.15	-10.6***	-154***	-0.804*	-3.63***	-70.4	-2.68	-23.2***	-335***
Regulator budget	-5.74***	3.43***	73.3***	-2.28***	1.32***	35.0**	-0.845***	0.554***	16.8	-5.03***	3.02***	72.3***
	(-2.85)	(1.71)	(35.3)	(-2.27)	(1.32)	(33.4)	(-2.2352)	(1.32)	(44.0)	(-2.49)	(1.50)	(34.8)
Institutional	-0.692***			-0.328***			-0.160***			-0.705***		
	(-0.343)			(-0.327)			(-0.4245)			(-0.349)		
Index stock	-1.80***			-0.925***			-0.382***			-2.03***		
	(-0.899)			(-0.922)			(-1.0111)			(-1.00)		
Market capital	1.56***			0.807***			0.328***			1.64***		
	(0.776)			(0.804)			(0.8678)			(0.815)		
Turnover (2)	0.597***			0.314***			0.059			0.619***		
	(0.296)			(0.313)			(0.1573)			(0.306)		
Month-end	3.05***			1.54***			0.623***			3.16***		
	(1.52)			(1.54)			(1.6478)			(1.56)		
Quarter-end	4.27***			2.11***			0.741***			4.15***		
	(2.12)			(2.10)			(1.9608)			(2.05)		
Volatility	-1.72***			-0.791***			-0.146***			-1.75***		
	(-0.857)			(-0.789)			(-0.3860)			(-0.866)		
Abnormal return (AR)		1.64***			0.826***			0.340***			1.63***	
		(0.819)			(0.823)			(0.813)			(0.814)	
Reversion (RV)		0.523***			0.272***			0.095***			0.521***	
		(0.261)			(0.271)			(0.228)			(0.259)	
Abnormal volume (AV)		1.65***			0.840***			0.499***			1.71***	
		(0.826)			(0.837)			(1.19)			(0.853)	
AR time series			-0.330			-0.362			0.782			-0.002
			(-0.159)			(-0.345)			(2.04)			(-0.0012)
RV time series			57.3***			24.1***			13.6**			57.4***
			(27.6)			(23.0)			(35.7)			(27.6)
AV time series			-5.62***			-1.51*			-1.38***			-7.58***
			(-2.70)			(-1.44)			(-3.62)			(-3.65)
Exchange dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	No	No	No	No	No	No	No	No	No	No	No	No
Observations		1,199,937			1,199,937			1,199,937			1,199,937	
Log likelihood		-2,644			-2,654			-2,735			-2,651	

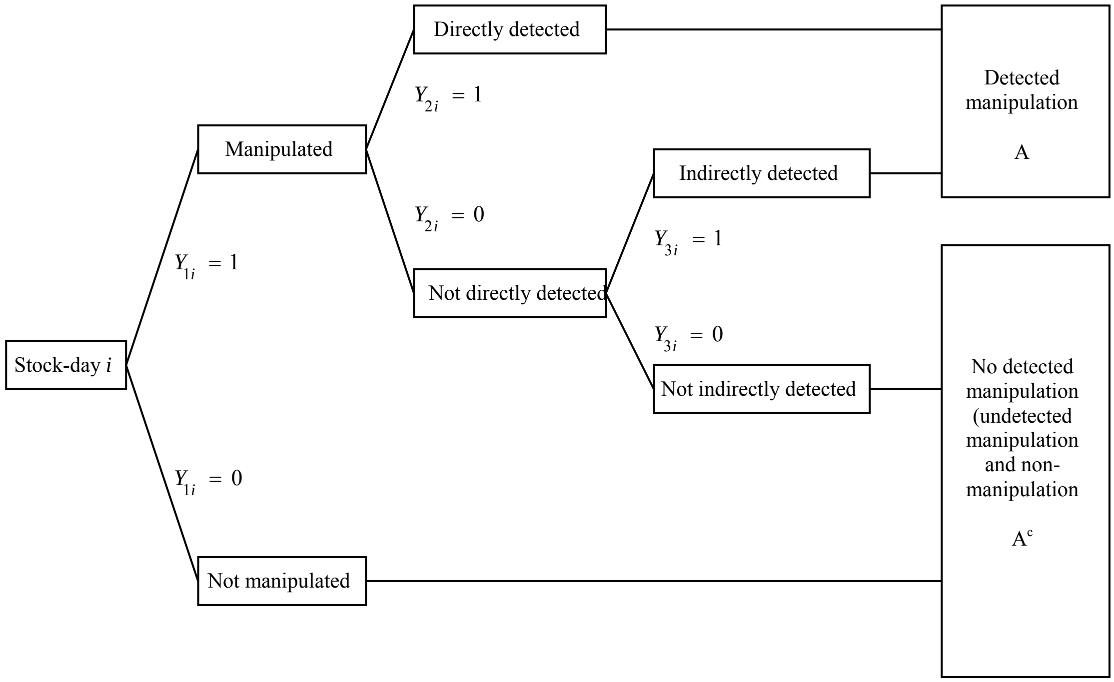

Figure 1. Modified detection controlled estimation model.