Product Differentiation, Benchmarking, and Corporate Fraud*

Audra Boone† William Grieser† Rachel Li‡ Parth Venkat§

March 15, 2019

Abstract

We find that product market differentiation is an economically meaningful and statistically robust predictor of accounting fraud. To establish identification, we exploit a firm's similarity with competitors issuing an IPO as a shock to the firm's product differentiation. Further, we control for several measures of competition and exploit large tariff reductions to show that product market differentiation plays a unique role in determining fraudulent behavior, relative to other characteristics of competition. Finally, using cross sectional variation in firm complexity, institutional monitoring, and analyst attention, as well as shocks to a firm's information environment, we provide evidence that lower differentiation disciplines managers by facilitating fraud detection through a benchmarking channel.

Keywords: Corporate Fraud, Product Market Competition, Product Differentiation, Corporate Governance

*We thank Donald Bowen, Joey Engelberg, Will Gerken, Umit Gurun, Gerard Hoberg, Zack Liu, Michelle Lowry, Tanakorn Makaew, Gonzalo Maturana, Jordan Neyland, Joerg Picard, Vesa Pursiainen, Nathan Swem, Xinxin Wang, Jared Wilson, participants at the 2018 FMA Asia Annual Conference, the 2018 FMA Annual Meeting, the 2018 Australasian Banking and Finance Conference, the 2018 New Zealand Finance Conference, and the 2019 Midwest Finance Association Conference, as well as seminar participants at Drexel University, Clemson University, Southern Methodist University, Texas Christian University, Universidad de los Andes, University of Nevada Las Vegas, and the U.S. Securities and Exchange Commission for helpful comments. †Department of Finance, Neeley School of Business Texas Christian University, e-mail: audra.boone@tcu.edu; w.grieser@tcu.edu. ‡Southern Methodist University, e-mail: qingqiul@smu.edu §U.S. Securities and Exchange Commission. The SEC disclaims responsibility for any private publication or statement of any SEC employee or Commissioner. This article expresses the authors views and does not necessarily reflect those of the Commission, the Commissioners, or other members of the staff.

I Introduction

Financial reporting facilitates efficient resource allocation by providing information about a firm's financial position, performance, and relevant circumstances to stakeholders. When firms fraudulently report their financials, they undermine this process by destroying the trust essential to a well-functioning financial system (Greenspan, 2008). Such miscreant behavior can erode firm value (Karpoff et al., 2008; Dyck et al., 2010), impose negative externalities (Kedia and Philippon, 2009), discourage stock market participation (Guiso et al., 2008), and distort investors' allocation decisions (Giannetti and Wang, 2016; Gurun et al., 2018). These consequences highlight the importance of understanding the factors shaping managers' incentives to commit fraud and the ability of monitors to detect such behavior.

In this paper, we explore the relation between financial misreporting and product market competition, which is widely considered one of the strongest forces affecting managerial behavior (Smith, 1776; Shleifer and Vishny, 1997). Product market competition can influence fraudulent conduct through competing forces. First, competition can pressure managers to manipulate financial statements in an attempt to boost their perceived performance prior to acquisitions, capital raising events, or stock option expiration dates (e.g. Shleifer, 2004; Bergstresser and Philippon, 2006). Alternatively, competition can discipline managers by aligning managerial incentives with shareholder interests (Hart, 1983; Schmidt, 1997), or by enriching the information environment, thus facilitating monitors' ability to assess performance and detect divergent behavior (Tirole, 2010). In essence, competition can limit the opportunity for managers to engage in financial misreporting, thereby enhancing discipline.

Despite the intuitive theoretical relations between competition and fraud, empirically identifying these effects has proved challenging. Part of the difficulty arises because traditional empirical measures obfuscate the multifaceted nature of competition, and impose challenges in comparing values across industries.[1] Given that competition is predicated on rivals producing marketable substitutes, we emphasize the role of product market differentiation in influencing fraud. To measure product differentiation, we use pairwise product similarity scores developed by Hoberg and Phillips (2010, 2016). Unlike traditional measures of competition that presume all firms have static and equal relationships within industries, product similarity scores provide firm-specific, time-varying competitor networks that capture the degree of competition between firms. These features enable us to circumvent another difficulty in identifying the effects of competition on fraud by separating firm-level aspects of managerial behavior from aggregate industry characteristics.

[1]For instance, it is unclear how to compare a high concentration ratio in a highly differentiated industry to a low concentration ratio in a homogeneous industry.

Our analyses reveal that the incidence of fraud is significantly lower for firms with a less differentiated product mix. Specifically, a one standard deviation increase in average product similarity score is associated with a 14.8-23.7% decrease in the rate of settled SEC enforcement actions and fraud-driven class action lawsuits.[2] These findings are consistent with product market competition imposing discipline, on average. Moreover, we find that product market differentiation has a relatively large economic effect when compared to virtually any predictor of corporate fraud previously explored in the literature.

One concern is that product similarity may be associated with pervasive differences in the rate of observed fraud across industries (e.g. Povel et al., 2007; Wang et al., 2010), which are imperfectly controlled for by our inclusion of industry and industry-year fixed effects. However, firm-level product similarity scores enable us to exploit within-industry heterogeneity that is plausibly outside of a firm's control. To this end, we exploit the similarity score of competitors undergoing an IPO as a potential shock to a firm's ex post product differentiation in an instrumental variable framework. Importantly, a firm has little, if any, control regarding which rivals choose to issue an IPO, the timing of that decision, and whether the rival IPO will result in a higher or lower similarity score. Using two-stage least squares, we find corroborating evidence that product market similarity exhibits a strong disciplining effect. Importantly, the variation in these tests come from within-industry. Thus, while we can not entirely rule out omitted variables bias, these findings help mitigate concerns over selection effects and industry heterogeneity.

It is important to note that product market similarity is based on 10-K descriptions, and therefore captures publicly-available information about a firm's differentiation rather than the firm's total differentiation with both public and private competitors. Next, we illustrate that this informational content of product similarity captures a particularly meaningful relationship between competition and fraud that is not explained by traditionally-used measures of competition. First, we show that including a battery of alternative measures as controls has a negligible effect on the coefficient estimates for product market similarity.[3] Furthermore, we fail to find a robust statistical relationship between these alternative competition measures and fraud when examined separately. To further explore the unique role that product market differentiation plays relative to other measures of competition, we consider the potential disciplining channels.

One possibility is that intense competition can facilitate investors' control of the agency

[2] Donelson et al. (2017) suggest the combination of AAER and class action lawsuits provide the most accurate and complete measure of corporate fraud that is currently available.

[3] These measures include industry-wide profit margin, sales concentration, top-4 firm market share, number of competitors, product market fluidity, and the extent firms mention competition in their 10-Ks.

problem by reducing managerial slack. Specifically, competition can reduce managerial rent extraction (Ades and Di Tella, 1999; Nickell, 1996), reduce the benefits of shirking (Hart, 1983), and enhance bankruptcy risk, thus forcing managers to work harder to survive (Schmidt, 1997). To explore this channel, we exploit large tariff reductions, which increase the intensity of foreign competition (Fresard, 2010), but do not directly affect the quality or quantity of readily available information through financial disclosures. Thus, tariffs provide an apt setting to analyze changes in competition that are likely independent of the informational aspect of 10-K based product differentiation. Our analysis reveals that tariff reductions have a mild pressure effect on fraudulent behavior. Further, product similarity continues to exhibit a strong disciplining effect in the presence of large tariff reductions. These findings further substantiate the notion that product similarity highlights a unique aspect of competition.

Another possibility is that information conveyed by similar product market rivals provides context for industry conditions, thus enabling more precise signals of firm-specific performance (Holmstrom, 1982; Nalebuff and Stiglitz, 1983). In turn, the ability to benchmark against comparable firms facilitates monitoring for investors, regulators, and auditors. If enhanced monitoring increases the possibility of getting caught, managers could rationally respond by committing less fraud, leading to a disciplining effect. To explore this channel, we perform several cross-sectional tests. First, we study how firm complexity impacts the relationship between product market differentiation and fraud. On the margin, it is likely more difficult to detect abnormal behavior for complex firms operating in many segments than for firms with a simple organizational structure (e.g. Cohen and Lou, 2012).[4] We proxy for complexity using the number of unique SIC codes that a firms product mix spans, and split firms into quartiles based on this proxy. Consistent with the benchmarking channel, we find that the disciplining effect of product similarity increases monotonically across complexity quartiles, after controlling for firm size. Indeed, the coefficient estimate is more than four times as large for the most complex quartile when compared to the least complex quartile. This finding suggests that having less differentiated rivals generates a larger marginal impact on the ability to detect corporate fraud for complex firms.

Next, we consider the marginal impact of product market similarity on the monitoring of two particularly important corporate governance actors with varying degrees of reliance on public financial statements: institutional investors and sell-side analysts. Institutional investors have access to managers and internal information relevant to fraud (Shleifer and Vishny, 1986; Dyck et al., 2010), whereas "analysts ... [have] to delve through details of

[4]Cohen and Lou (2012) find that financial markets incorporate information at a slower pace for complicated firms.

financial reports and industry trends to uncover misrepresentations" (Dyck et al., 2010). Access to inside information can potentially reduce the reliance on public signals of firm performance. Consistent with this notion, we find that product market similarity has a muted effect for firms with high institutional ownership concentration and an enhanced effect for firms with greater analyst coverage. Further exploration reveals that the pattern for analyst coverage is driven by the post- Reg FD period (i.e. after analysts' reliance on publicly available information increased). These findings are consistent with publicly available benchmarks playing a greater role for monitors that rely more heavily on publicly available information.

As a final test to illustrate the information channel, we utilize IPOs and M&As of a firms rivals as a shock to the firms information environment. IPOs increase the publicly available financial information of existing competitors, which in turn enhances the ability to assess, compare, and scrutinize related firms own financial statements (e.g. Bauguess et al., 2018). Similarly, M&A activity generates attention in the merging firms industries due to potential spillover effects on rivals, customers, and suppliers (e.g. Fee and Thomas, 2004), and due to potential ensuing acquisition activity (e.g. Song and Walkling, 2000). We find that higher IPO and M&A activity of rival firms is associated with a higher incidence of detected fraud. Further, this increase in detection is significantly more pronounced for firms with less similar rivals, prior to the IPO (i.e., the effect is greater for ex ante less disciplined firms). This finding suggests these events enhance monitors effectiveness in detecting fraud.

A typical concern with empirical studies of corporate fraud is that only detected, rather than all committed, fraud is observed (Dyck et al., 2013; Dimmock and Gerken, 2012). That is, empirical measures of fraud capture the joint outcome of a firm committing fraud and being caught. If greater product market similarity (lower differentiation) enhances detection or increases pressure to commit fraud, we would observe more cases being uncovered by auditors, regulators, and investors. However, our finding that firms with lower product market differentiation have lower rates of detected fraud indicates that managers of these firms either engage in less fraud or are less likely to be caught. Lower product market differentiation (i.e. higher similarity) with competitors should be more informative regarding common shocks to production costs and demand (e.g. Tirole, 2010), which govern firm performance and financial reporting incentives. This argument is consistent with evidence suggesting that benchmarking informs boards regarding CEO ability (Murphy, 1986) as well as market- and industry-wide conditions when determining CEO pay (Oyer, 2004).[5] Indeed, it is difficult to ascertain a plausible explanation for why the presence of similar rivals would

[5]Also consistent with the benchmarking effects of competition, Hsu et al. (2017) find that analysts produce more accurate for forecasts for firms that face more intense competition.

decrease outsiders ability to detect reporting manipulations. Thus, our findings suggest that managers rationally respond to enhanced detection rates by committing less fraud. We find corroborating evidence for the commission effect according to the bivariate probit (partial observability) model highlighted in Wang et al. (2010).

In falsification tests, we replace settled fraud-driven misstatements with unintentional accounting restatements and fail to observe a discernable pattern with product market similarity. This finding increases our confidence that we have identified an economically meaningful link between fraud and product market similarity in our primary analysis. Additionally, our primary results remain qualitatively similar to a variety of non-linear model specifications, as well variations in our set of control variables and the time period used in our estimation. Our results are also robust to a) variations in the level of winsorization, b) to removing outliers, rather than winsorizing, and c) eliminating any winsorization of the data.Further, our results are not sensitive to the specific construction of our independent variable. In particular, our results are qualitatively unchanged when we average product similarity scores across a firms top 5, 10 or 15 closest rivals, rather than averaging across the firms entire competitor network, when we use a precision weighted average, and when we use the number of rivals above a given percentile of product market similarity (75th, 90th and 95th).

In summary, we find a strong and robust link between product market differentiation and corporate fraud. Indeed, our estimates suggest that product market similarity has an economically larger effect on fraud than any factor, other than firm size, previously documented in the literature. Our initial results suggest that product market similarity imposes a strong disciplining effect on financial reporting misconduct. Further, while none of our follow-up analyses provides incontrovertible evidence in isolation, the preponderance of evidence suggests that the disciplining effect stems through a benchmarking channel. These results indicate that market-based mechanisms, particularly through enhanced information environment, play an important role in both the incentive to commit fraud and the ability of external parties to uncover fraudulent activities.

II Literature Review

Our paper relates to the literature examining the effect of various measures of competition on managerial discipline. On one hand, competition can diminish conflicts of interest by incentivizing managerial effort (Nickell, 1996) or by reducing resources available for rent extraction (Ades and Di Tella, 1999; Schmidt, 1997). On the other hand, competition has been argued to pressure managers to distort the perceived performance relative to rivals

(Shleifer, 2004; Tirole, 2010; Andergassen, 2016). Until recently, only coarse industry-level measures of competition have been available to researchers, which has introduced challenges in identifying the potential effect of these opposing forces. Indeed, the existing empirical evidence on the link between competition and fraud is often contradictory and inconclusive (e.g. Holmstrom, 1982; Nalebuff and Stiglitz, 1983; Karaoglu et al., 2006). We shed light on this relationship by exploiting newly developed, firm-level measures of product differentiation that allow us to conduct more powerful tests. Consistent with a disciplining channel of competition, we document that product market similarity is strongly associated with a lower incidence of fraud.

In addition, our work suggests that benchmarking is an important factor to consider in studying competition, as it enhances information, and therefore facilitates monitoring ability (Badertscher et al., 2013). In turn, this benchmarking channel can influence managerial behavior. More specifically, as predicted by Holmstrom (1982) and Nalebuff and Stiglitz (1983), to the extent that firms are similarly impacted by the competitive landscape, more direct competition increases information about the firms that could help reduce moral hazard problems. Indeed, empirical work indicates that a information generated from peers is an important determinant the cost of capital (Shroff et al., 2017) and can influence analyst coverage and forecast accuracy (Hsu et al., 2017). We study whether this effect applies to financial fraud. Our results suggest that having similar rivals can facilitate information acquisition, which is consistent with survey evidence by Dichev et al. (2013) on factors that help detect financial misrepresentation. Thus, our work suggests an avenue through which external parties, such as short-sellers and analysts, can obtain information useful in the detection of fraud (Dyck et al., 2010).

Our analysis also complements empirical work by Wang et al. (2010) who show that industry-level information affects fraud detection. Our results indicate that information contained in firm-specific product markets and unique competitor networks leads to substantial within-industry heterogeneity in fraud detection, after controlling for the industry-wide measures of information outlined in Wang et al. (2010). Other work by Balakrishnan and Cohen (2013) investigates the interplay between traditional measures of competition and the industry-level incidence of restatements, rather than misreporting and fraud. Our results also suggest that competition matters, but we focus on a particular dimension of competition (product differentiation) that facilitates exploration of the benchmarking channel. We show that the benchmarking ability brought about by firms with a similar product mix holds after controlling for various measures of industry concentration. Further, we fail to find a direct relationship between product differentiation and financial restatements, which we find to be reassuring for the disciplining channel of product market similarity.

Our work also relates to the literature on corporate governance and corporate fraud (e.g. Beasley, 1996; Farber, 2005; Khanna et al., 2015). Several papers propose that corporate governance mechanisms are endogenous responses to the cost and benefits of different internal governance mechanisms, as well as external monitoring from entities such as sell-side analysts, banks, or institutions (Gillan et al., 2011). Our findings suggest an alternate source of external discipline: product market competition, which compliments recent work suggesting that product market competition can substitute for other formal corporate governance mechanisms (Giroud and Mueller, 2010; Chhaochharia et al., 2016). The remainder of the paper is organized as follows. Section II discusses the data and our set of control variables. Section III contains the empirical measures of competition. Section IV covers the results, and Section V concludes.

III Data

We follow recent empirical work of Donelson et al. (2017) by defining corporate accounting fraud as, the intentional, material misstatement of financial statements that causes damages to investors. Donelson et al. (2017) advocate using a combination of public and private enforcement actions through AAER and class action lawsuits to capture financial reporting fraud to mitigate measurement error. While regulatory enforcement is important, other participants, such as the media, industry regulators, and employees, serve as important actors in this arena (Dyck et al., 2010).

We obtain AAER data for the sample period 1996-2010. According to the Center for Financial Reporting and Management, the U.S. Securities and Exchange Commission (SEC) issues AAERs during, or at the conclusion of, an investigation against a company, an auditor, or an officer for alleged accounting or auditing misconduct. The AAER dataset provides information on the nature of the misconduct, the named individuals, and the entities involved, as well as their effect on the financial statements. The misstatement investigations in our sample occur because of bribery, fraud, inflated assets, financial reporting related enforcement actions concerning civil lawsuits brought by the in federal court, and orders concerning the institution and/or settlement of administrative proceedings.

We construct our sample of class action lawsuits following the work of Choi et al. (2009), Griffin et al. (2004), Jayaraman and Milbourn (2009), and Thompson and Sale (2003). We start by downloading all class action lawsuits from the SCAC hosted by Stanford University for 1996 through 2011 and scan each filing to only include 10-b5 class action lawsuits, which eliminates those lawsuits that occur for non-financial reasons.[6]

We define each firm-year as an AAER year, a SCAC year, both, or neither. Our primary independent variable, fraud, is a binary variable equal to one for all firm years in which there is an AAER or SCAC. We exclude firms in the financial and utilities industries and firms headquartered outside the United States. Further, we drop ADRs, firms with assets less than $1M, and firms with missing assets or sales items in Compustat. Our final sample of corporate fraud events includes 935 firm-years that are affected by AAER misstatements in at least one quarterly or annual financial statement from 322 unique firms from 1996 to 2010. In addition, our sample includes 311 class action lawsuits affecting 299 firms from 1996 to 2011. In total, our sample contains 498 firms and 1,217 firm-years, flagged as years with fraudulent reporting. These figures are very closely in line with those of (Dyck et al., 2010). As shown in Table 1, the overall incidence of fraud in our sample is 1.9%.

To construct our set of control variables, we follow work in the finance and accounting literature related to corporate fraud (variable definitions are reported in Table A.1). We include predictors of accounting misstatements from Dechow et al. (2011), which include Richardson et al. (2005) (RSST) accruals, change in accounts receivable (ΔAR), change in inventory (ΔInventory), the percentage of soft assets (% Soft Assets), change in cash sales (ΔCash Sales), change in ROA (ΔROA), change in employees (ΔEmployees), and a dummy for security issuance (d_Security Issue).

The variable RSST accruals measures the change in non-cash net operating assets, including both working capital accruals and long-term operating capital. Bergstresser and Philippon (2006) show that changes in accounts receivable (ΔAR) and change in inventory (ΔInventory) are associated with incentives to improve sales growth and gross profit margin. A firms soft assets as a percentage of total assets (% soft assets) is associated with more discretion for earnings management. We define % soft assets as total assets minus property plant and equipment and cash and cash equivalents, all scaled by total assets. Change in cash-based sales (ΔCash Sales) excludes accrual-based sales to measure the portion of sales that are not subject to discretionary accrual management. Change in ROA (ΔROA) controls for changes in earnings growth. The variable ΔEmployees is the percentage change in employees less the percentage change in total assets. This measure is associated with labor costs and must be expensed as incurred. Reducing the number of employees can boost a firms short-term financial performance by immediately lowering expenses. Finally, we in-

[6] Karpoff et al. (2017) note the importance of additional checks of the sources to ensure that they contain instances of fraud.

clude a dummy variable (d_Security Issue) equal to one for firm years in which a firm issues debt or equity, which can increase incentives to manage earnings (Rangan, 1998). We refer to specifications including only the controls from Dechow et al. (2011) as the Dechow set of controls.

We also include specifications that contain proxies for monitoring mechanisms and corporate opaqueness, which could potentially influence the marginal impact of our proposed benchmarking channel. We include Institutional Ownership, the natural log of the number of analysts covering a firms stock (Ln Num Analysts), research and development expenses (R&D), and industry stock return r-squared (Ind R2).[7] To construct the industry r-squared, we follow Wang and Winton (2014) and first regress each firms daily stock returns on the weighted-average daily market return and the weighted-average daily industry return. Then, we take the average r-squared for each firm in a given three-digit SIC code. Managers may feel pressured to commit fraud when they require capital from outside sources (Teoh et al., 1998; Wang and Winton, 2014). Thus, we include the Whited and Wu (2006) Index for financial constraints.[8]

We include the natural log of total assets (ln assets) as a measure of firm size. We also include the variable book leverage, which is defined as long and short-term debt over total assets. Highly levered firms have greater probabilities of financial distress, which has been shown be associated with financial misreporting (e.g. Healy and Wahlen, 1999). Alternatively, leverage can have a disciplining effect by either mitigating agency issues between managers and shareholders (Grossman and Hart, 1982), or providing an additional source of external monitoring vis-a-vis debtholders.

Product differentiation is likely related to relative performance evaluation (RPE). Firms with less product market differentiation might naturally have better benchmarks, and therefore, be more prone to RPE, which could pressure some managers to cut corners or misstate earnings to outperform benchmarks (Cheng, 2011). This effect would work against our hypothesis and findings. Thus, to increase the power of our tests, we control for RPE following the work of Wang and Winton (2014) who construct an indicator variable RPE. First, the authors estimate the following regression equation:

$$prob(\text{CEO Turnover}_{i,t-1}) = \gamma_1 RP^{+}_{i,t} + \gamma_2 RP^{-}_{i,t} + \epsilon_{i,t} \qquad (1)$$

[7] To handle observations with missing R&D, we follow the method outlined in Koh and Reeb (2015) and replace each missing observation with the industry year average and include a dummy variable for whether the firm has missing R&D (R&D dummy).

[8] In unreported analysis, we use an alternative proxy for equity finance needed (EFN) defined by Demirg-Kunt and Maksimovic (1998) as ROA/(1-ROA), which measures a firms asset growth rate in excess of the maximum internally financeable growth rate. We find qualitatively similar results.

where $RP^{+}_{i,t}$ is equal to relative performance when relative performance is above 0, and zero otherwise; and $RP^{-}_{i,t}$ is equal to relative performance when relative performance is below 0, and 0 otherwise. Relative performance is measured as the difference in performance between firm i and the weighted average of firm is rivals according to its three-digit SIC code. Following Wang and Winton (2014), we estimate equation (1) separately for each industry (three-digit SIC) and define the binary variable RPE equal one for industries where $\hat{\gamma}_2 < 0$. We refer to specifications that include all our control variables as the full set of controls.

Table 1 provides the number of observations, mean, standard deviation, 10th percentile, and 90th percentile value for our control variables. We estimate all specifications for both winsorized and non-winsorized data. Estimates obtained from winsorized data (1% in each tail) are reported in the Internet Appendix.

III.1 Product Differentiation

For our measure of product differentiation, we use the product similarity score developed by Hoberg and Phillips (2010, 2016). Using textual analytics, the measure captures the relatedness of a firms product market with all other firms that publicly file annual reports with the SEC in a given year. The process involves vectorizing the product market vocabulary from the business description from each firms 10-K, according to a dictionary the authors develop. They then assign pairwise similarity scores based on the cosine similarity between two firms' vectorized product market descriptions. The cosine similarity is higher when the product market descriptions between the two firms are more similar. The measure ranges from 0 (no similarity) to 1 (perfect similarity).

We also make use of the text-based network industry competitors (TNIC) that Hoberg and Phillips define as a byproduct of their product similarity score. The TNIC competitor set includes all firms with a similarity score above a given threshold.[9] Importantly, TNIC allows the flexibility for each firm to have its own distinct set of competitors. This intransitive feature better reflects economic reality, and it allows us to exploit granularity in the data that is not possible using measures created from standard transitive industry classifications, such as SIC codes, which implicitly presume that all firms have an equal relationship within industries and are unrelated across industries. In particular, we can isolate firm specific

[9]Hoberg and Phillips (2016) only provide with pairwise similarities above the threshold, c, so that the TNIC set matches the same degree of coarseness as the three-digit SIC code classification scheme. In other words, both TNIC-3 and three-digit SIC codes would result in approximately the same number of firm pairs being deemed competitors.

features of competition from industry-level characteristics in a regression framework.

The TNIC approach also improves upon some basic inaccuracies of other classification schemes. For example, the Coca-Cola Company and PepsiCo are not considered competitors according to their four-digit SIC code, or their Fama-French 48 industry classification, but have a high similarity score (80th percentile). Furthermore, TNIC industry classifications are updated annually, which provides more flexibility and accuracy in empirical design. For example, when Exxon sold its retail gas stations in 2008, this event was reflected by the change in its competitor set (TNIC) and average product similarity score (from 0.035 to 0.012). However, the divestment from Exxon was not reflected by a change in its SIC code or other industry classifications. As a result, the level of competition that Exxon faced according to SIC code-based Hirschman-Herfindahl Index (HHI) measures did not change in response to its large divestment. The measurement error imposed by traditional competition measures can bias results and limit the power to detect existing relationships between fraud and various aspects of competition.

Using the TNIC competitor classification and product similarity scores, we create our main variable of interest; Average Similarity Score, as the average pairwise similarity score of all competitors within a firms TNIC-3 classification in each year. As shown in Table 1, the firms in our sample have 49 competitors on average, with an Average Similarity Score of 0.03 above the threshold set by Hoberg and Phillips (2016).

A potential issue with the Average Similarity Score based on the TNIC classification is that it only includes firms over a certain threshold of similarity. While imposing this threshold allows us to focus on closely-related rivals, there can be substantial variation in the number of competitors being averaged across for each firm. The wide variation in both the number of competitors each firm has and the degree of similarity with each competitor, can obfuscate the association between fraud and product differentiation. Two firms, for example, could have the same average product market similarity scores for different reasons. One firm could have several moderately close rivals, while another firm could have a mix of some nearly identical rivals and some that are barely related. While both firms could have the same average product similarity score, we would expect the firm with the near identical rivals to provide more precise information about a firm's competitive landscape and factors affecting performance.

To address such concerns, we implement a series of alternate methods for aggregating product similarity scores. Rather than averaging across all competitors in a firm's TNIC, we average across the top 5, 10 or 15 closest competitors. This process creates more homogeneity by utilizing the same number of competitors for each firm and focuses on each firm's closest rivals, which should provide the greatest information externalities. As an alternative approach, we count of the number of competitors each firm has that are in the top percentile (95th, 90th and 75th) of the overall distribution of similarity scores across all firms in the sample. This process allows us to count the number of rivals that each firm has that are very similar relative to the complete cross-section of firms.

Additionally, we develop a measure that emphasizes the degree of similarity between rivals. In particular, rivals provide signals about similar firms, with greater similarity between two rivals producing a less noisy signal. It follows that both the similarity with a given rival, as well as the number of rivals, impact the total signal provided by a firms product market competitors. If signal noise is normally distributed, then there is an inverse squared relationship between product market similarity and the quality of the signal. We define a measure of precision as:

$$precision_{it} = \left(\frac{1}{N_i} \sum_{j=1}^{N_i} \frac{1}{(1 - score_{i,j,t})^2} \right)^{0.5} \tag{2}$$

where N_i is the number of competitors in firm is TNIC, and $score_{i,j,t}$ is the product similarity score between firm i and competitor j in year t.[10] Higher precision is indicative of a greater signal provided by a firms product market rivals. Rather than an average, we also create a sum of the similarity scores, which would be higher when a firm has more rivals that are more similar.

Table A.2 of the Internet Appendix reports the correlations for our main measure, Average Similarity Score, and the alternative similarity score measures noted above. These measures are highly correlated with each other and with the main measure that averages across all competitors (around 75%), which mitigates concerns regarding the distribution of similarity scores driving the results.

[10]We thank Jerry Hoberg for suggesting this measure.

IV Empirical Results

IV.1 Competition and Fraud: Pressure vs. Discipline

IV.1.1 Product Differentiation

In this section, we discuss results from firm-level regressions that examine the association between corporate fraud and product market differentiation. We first explore associations in a standard panel data framework before exploring an instrumental variables approach.

We report OLS estimates for the association between average product similarity score (Average Similarity Score) and corporate fraud in Table 2.[11] The firm-year is the unit of observation in all reported specifications in this section. The specification in Column 1 only includes year fixed effects. Column 2 includes the natural log of total assets (Ln Assets) as well as the Dechow set of controls (i.e. accruals, change in accounts receivable (ΔAR), change in Inventory (ΔInventory), the percentage of soft assets (% Soft Assets), change in cash sales (ΔCash Sales), change in ROA (ΔROA), change in employees (Δ Employees), and a dummy for security issuance (d_Security Issue)).

In Column 3, we also include R&D, a dummy for positive R&D, the natural log of the number of analysts (Ln number analyst), Institutional Ownership, the Whited-Wu Index, Industry Stock Return R-squared, a flag for relative performance evaluation (RPE flag), and the number of competitors based on TNIC classification (TNIC NCOMP). Including Institutional Ownership results in a large drop in the number of observations and does not appear to have a meaningful effect on the detection of fraud. Furthermore, inclusion of Institutional Ownership only seems to intensify the relationship between fraud and Average Similarity Score. Considering these issues, we drop Institutional Ownership from the remaining specifications. We also exclude the RPE flag due as it is an industry level measure, which is absorbed by the industry fixed effects. Thus, the specification from Column 4 is our primary specification throughout the remainder of our analysis. Henceforth, we refer to the specification of control variables in Column 4 as our Full set of control variables. All explanatory variables are lagged by one year.

The granularity of our data enables us to control for unobserved heterogeneity at the industry and industry-year level. The specification in Column 4 includes industry (three-digit SIC code) and year fixed effects, and the specification in Column 5 includes industry-year fixed effects. The inclusion of fixed effects improves upon existing studies that are typically

[11] In untabulated analysis, we estimate this relationship with probit and logit specifications and find similar results.

unable to account for unobserved industry heterogeneity because the variables they deploy are often constructed at the industry level. In particular, inclusion of industry or industry-year fixed effects accounts for pervasive differences in the propensity to commit fraud across industries and helps to mitigate the effects of large industry shocks explained by factors not controlled for in our initial specifications. The t-statistics are calculated from standard errors clustered by three-digit SIC code.[12]

Throughout all specifications, the coefficient estimate for Average Similarity Score exhibits a very consistent, economically meaningful, and statistically significant, mitigating effect on fraud. A one standard deviation increase in Average Similarity Score (0.023) is associated with a roughly 0.48 percentage point decline in the rate of fraud. That is, a one standard deviation increase in average product similarity score is associated with a decline in the rate of fraud from 1.8% to 1.3%. Thus, the results suggest that product similarity has a large economic effect. Indeed, firm size is the only predictor of fraud that we have found documented in the literature to have a larger economic relation to fraud than Average Similarity Score.[13] These results are robust to several different sample periods (i.e., before and after Sarbanes Oxley) and to the inclusion of controls that proxy for external monitoring, such as the number of analysts and the degree of institutional ownership.[14]

IV.1.2 Addressing Industry Heterogeneity

A potential concern with our initial findings is that product market similarity could be related to pervasive differences in fraudulent activity across industries that have been documented in the literature (e.g. Povel et al., 2007; Wang et al., 2010). While our inclusion of industry and industry-year fixed effects should partially mitigate this concern, we cannot perfectly control for differences in industry characteristics that are related to fraud. For instance, CEOs that are more likely to commit fraud might select into industries based on characteristics that are either directly or incidentally associated with product market differentiation. More specifically, an overconfident manager could be more likely to flaunt rules (commit fraud) and choose product markets that are innovative and risky (i.e. more likely to be highly differentiated). Alternatively, in mature industries, firms could have settled into well-differentiated positions and simultaneously commit less fraud for fear of damaging

[12]Our results are robust to clustering at broader industry classifications (e.g., two-digit SIC) and at the firm level. It is also robust to the alternate construction of the independent variable as described in Section III.

[13]We report results in the Internet Appendix for a specification with all variables standardized for expositional convenience and ease of comparison.

[14]We also control for financial statement comparability developed by De Franco et al. (2011) and we obtain similar results.

valuable brands or reputations.

To mitigate concerns of omitted variable bias, we exploit IPOs of a firm's rivals as a shock to a given firm's level of product market differentiation - rival activity that is plausibly outside of the firm's control. In particular, we implement two-stage least squares regressions using the product similarity scores between firm i and its rivals that issue an IPO in year t as an instrument for firm is degree of product market differentiation in year $t+1$. Importantly for our setting, the change in differentiation due to rival IPOs occurs after the focal firm has selected into an industry.

For each of our instruments to be valid, it should be related to a firm's product market similarity, but not directly related to a firm's propensity to commit fraud prior to selecting into a given industry. The first criterion requires that our instrumental variable exhibits a strong relation to a firms degree of product differentiation. Unless the similarity between firm i and its rivals issuing an IPO is exactly equal to the average similarity between firm i and its preexisting public rivals, this activity would change firm i's average product market similarity. Thus, the relevance condition should hold by definition.

The second criterion (exclusion restriction) requires that product market similarity with rivals' issuing an IPO is unrelated to the focal firm's propensity to commit fraud, other than through its effect on the focal firm's ex post degree of product differentiation. One potential concern is that rival IPOs have been shown to occur in waves (e.g. Lowry and Schwert, 2002), which could be endogenously related to industry opportunities and fraudulent behavior. However, we use the similarity score with a firm's rivals issuing an IPO as opposed to the existence or number of rival IPOs. In addition, we directly control for the number of rival IPOs and capital raised, and include industry fixed effects to better isolate the changes happening directly through a change in average score. Rival IPOs increase (decrease) a firm's average similarity score if the similarity between the issuing and focal firm is greater (smaller) than the focal firm's ex ante average similarity score. A firm has little, if any, control regarding which rivals choose to issue an IPO, the timing of that decision, and whether the rival IPO will result in a higher or lower similarity score. Thus, the exclusion restriction is likely only violated if issuing rivals have knowledge of the focal firm's propensity to commit fraud and then time their IPOs based jointly on that knowledge and their product market similarity to the focal firm.

To conduct this analysis, we collect a sample of IPOs from Thomson Reuters SDC platinum financial securities database from 1996-2012. For each pairwise observation of competitors, i and j, we flag whether firm j underwent an IPO in year t. We then create our instrumental variable, Rival-IPO Similarity, equal to firm is Average Similarity Score with all rival firms issuing an IPO in year t. Already public rivals in year t are excluded from this calculation. Next, we create a variable, Num Competitor IPO, that is the total number of rivals firm i competes with that underwent an IPO in year t. We control for the number of rival IPOs to help isolate the effect due to changes in Average Similarity Score, rather than the extent of rival IPO activity. For robustness we also create an indicator variable, Competitor IPO, that is equal to 1 if any of a firms rivals underwent an IPO in year t, and 0 otherwise. On average, there are 3.2 rival IPOs per firm-year in our sample, with a median of 0 (43% of firms have at least one rival IPO). Contingent on having at least one IPO rival, each firm has an average of 7.6 rivals launching IPOs.

In our first stage results, reported in Column 1-3 of Table 3, we find a strong positive relationship between Rival-IPO Similarity and Average Similarity Score. The positive sign indicates that rivals undergoing an IPO that are more similar to firm i, increases firm is overall similarity score, on average. The smallest F-statistic that we observe in the first stage is 35.52 (5.96^2) and all others are above 70.96 (8.42^2). These F-statistics are all substantively larger than 10 (the typical rule of thumb threshold), so it does not appear that we have a weak instrument problem. The reported t-statistics are calculated using standard errors clustered at the three-digit SIC code (SIC3).

In Column 4-6 of Table 3, we report the second stage results of two-stage least squares regressions using Rival-IPO Similarity as an instrument for Average Similarity Score. We find strong corroborating evidence that the incidence of corporate fraud is significantly lower for firms operating in less differentiated product markets. In particular, the coefficient estimates range from 0.495-0.670 across all specifications, suggesting a consistent and economically meaningful effect. Importantly, these findings persist with the inclusion of industry fixed-effects in Column 6, which further helps to mitigate endogeneity concerns.

The coefficient estimates in the IV analysis are roughly twice as large as the OLS coefficients. The larger estimates could imply that there are omitted variables working against our observed effect in our initial analyses, and that the actual impact of product market differentiation on fraud is indeed larger than our initial estimates suggest. Alternatively, the larger coefficient estimates could be capturing a local average treatment effect. That is, the larger partial effect could be concentrated in firms with rival IPO activity. However, the estimates are in line with those of the top quartile in our complexity analysis presented in Table 6. While we cannot entirely rule out the potential for omitted variables to jointly determine a firms fraudulent reporting and the similarity score of rivals who undergo IPOs, the IV results are suggestive of a causal relationship between product differentiation and corporate fraud.

IV.1.3 Alternative Measures of Competition

In this section, we illustrate that product differentiation captures a particular aspect of competition not explained by traditionally-used measures. Our alternative measures of competition include those widely utilized in prior literature, such as: HHI (Hirschman, 1945; Herfindahl, 1950), profit margin (Bain, 1951), and the sales concentration ratio of the largest four firms in an industry (Heflebower, 1957).

The HHI based on SIC code is the most extensively used measure of competition in studies related to product market competition. The HHI for industry j is calculated as:

$$\text{HHI}_j = \sum_{i=1}^{N_j} (MS_i)^2 \tag{3}$$

Where MS_i is the sales-based markets share of firm i in industry j, and N_j is the number of firms in industry j. HHI has a maximum value of 1 that is attained if a single firm makes up an entire industry, and a minimum value of $1/N_j$ if each firm has equal weight in industry j. HHI was originally designed to measure concentration in the U.S. steel industry, a relatively homogeneous industry. Thus, this measure can better capture the competitive landscape where industries are well defined (e.g. Faccio and Zingales, 2017). It is less useful, however, in instances where firms have diversified baskets of differentiated products and are therefore more difficult to delineate. To allow for more accurately defined product markets, we also create a TNIC based HHI. When constructing each firms TNIC-HHI, we can weight the sales of each rival by the firms product market similarity with that rival. Thus, a firms more similar rivals receive greater weight in its TNIC-HHI.

Classic models of competition suggest that competition becomes more intense as more firms offer marketable substitutes(Tirole, 1988), which motivates our inclusion of the number of competitors according to a firms TNIC or three-digit SIC codes. As additional measures of market power, we include profit margin (Bain, 1951) and the sales concentration ratio of the top four firms in an industry (Heflebower, 1957). We also include newer measures of competition such as product market fluidity, which captures competitive pressure from potential entrants that captures each firms ex ante competitive threats (Hoberg et al., 2014). Product market fluidity isolates threats based on the actions by rival firms, rather than changes of the firm itself. Finally, for a subset of our firms, we use a firm-level based on the frequency that firms discuss competition, or its variants, in their 10-Ks (Li et al., 2013). This measure captures the degree of competitive threats perceived by the focal firm.

In each column of Table 4, we estimate a specification that includes our full set of control variables from Column 4 of Table 2. In Columns 1-3, sales-based HHI using three-digit SIC codes does not appear to have a meaningful relationship with fraud. In Column 4, we include a sales-based Herfindahl-Hirschman Index (HHI) calculated from a firms TNIC, rather than primary SIC code. This specification allows us to explore whether the apparent lack of power exhibited by the HHI in relation to corporate fraud is driven by the use of SIC codes to define competitor networks, or by the lack of a strong relationship between market concentration and fraud. In Column 5, we also control for the natural log of the number of competitors each firm has according to its TNIC. Column 6 contains product market fluidity from Hoberg et al. (2014). Finally, in Column 7, we include the Competition 10-K measure based on Li et al. (2013). Notably, the relationship between Average Similarity Score on fraud remains consistent in both significance and magnitude across all columns. The key takeaway from this analysis is that the alternative measures do not appear to affect the association between fraud and product differentiation. We explore several combinations of control variables and different sample periods and find that these results are not sensitive to model specification.

In Table 11, we report estimates for the relationship between corporate fraud and other measures of competition, excluding Average Similarity Score. We perform this exercise to ascertain whether the traditional measures have an association in the absence of product differentiation, which could be capturing some of the variation of these traditional measures. For each measure of competition, we include our full set of control variables from Table 2 (described in section II).[15] We estimate the specifications in Table 11 without the inclusion of industry fixed effects to improve the odds of finding a statistical relationship. The number of rivals in a firms primary three-digit SIC industry (Ln NCOMP SIC3) is the only competition variable that exhibits a statistically significant relation to fraud (10% level) in Table 11. Interestingly, Top 4 Concentration no longer exhibits a relation to fraud when Average Similarity Score is not included in the same regression.

In untabulated results, we find that including industry (SIC3) fixed effects attenuates the point estimates of alternative measures of competition in Table 4 even further. This result highlights one potential reason for a lack of strong evidence between product market characteristics and corporate fraud documented in the literature. Furthermore, Table A.2 of the Internet Appendix reveals a weak correlation between traditional measures of competition and product similarity scores. Overall, the evidence suggests that product market similarity captures a different aspect of competition not explained by traditional measures. As these

[15] We also estimate specifications without control variables and report results in the Internet Appendix. The results are substantively very similar. Note, we include firm size as a control in all specifications since size is strongly related to measures of competition and is a strong predictor of fraud (see Buzby, 1975; Reynolds and Francis, 2000; Graham et al., 2005).

alternative measures are all designed to capture the degree of competition in an industry in various ways, our results suggest that there is something unique about the relation between fraud and product market similarity.

IV.1.4 Fraud Detection vs. Commission

One concern with empirical studies on fraud is that only detected fraud rather than all committed fraud is observed (Dyck et al., 2013; Dimmock and Gerken, 2012). The empirical measures of fraud captures the joint outcome of a firm committing fraud and being caught. This can make studies difficult to interpret because partial effects can be due to changes in the probability of detection and / or changes in commission rates. Lower product market differentiation could enhance any combination of the two pairs: detection or evasion and / or pressure or discipline.

If product market differentiation decreases the amount of fraud committed or decreases detection rates without impacting commission rates, we would observe less detected fraud for highly differentiated firms. However, we find the opposite, that firms with higher product market differentiation have higher rates of detected fraud. This indicates that managers of differentiated firms either are less likely to be caught or engage in less fraud. In our opinion, it is difficult to ascertain a plausible explanation for why the presence of similar rivals would decrease outsiders ability to detect reporting manipulations. On the other hand, lower product market differentiation with competitors should be more informative regarding common shocks to production costs and demand (e.g. Tirole, 2010) which govern firm performance and financial reporting incentives. This argument is consistent with evidence suggesting that benchmarking informs boards regarding CEO ability (Murphy, 1986) as well as market- and industry-wide conditions when determining CEO pay (Oyer, 2004). Thus, because we do not believe higher differentiation could make it easier to detect fraud (it should have no effect or a positive effect on detection) and because we find more detected fraud among more differentiated firms, the only possibility is that managers of differentiated firms commit more fraud.

The next important question is why differentiated firms commit more fraud. There could be a direct impact of differentiation on fraud commission or an indirect impact through increased detection probability. Our hypothesis is consistent with the later, that differentiation makes concealing fraud easier, thus differentiated managers are more likely to commit fraud. This is equivalent to saying that similarity makes detection easier, and thus managers rationally respond by committing less fraud.

As a first step to provide evidence for the indirect channel, we estimate a bivariate probit model employed by Wang (2011). This is a latent variable model that aims to exploit the timing differences in detection and commission (with commission being prior to detection). The model solves two simultaneous probit specifications and achieves identification through exclusion restrictions: namely, that some variables are only associated with detection while others are only associated with commission. Following Wang (2011) we include Relative Performance Evaluation, ROA, Equity Finance Needed, Book Leverage and Institutional Ownership only in the commission regression and Abnormal Industry Litigation, Abnormal Stock Return Volatility, Abnormal Turnover, and a Disastrous Return Dummy only in the detection regression. All other controls are included in both regressions. In Table A.3 of the Internet Appendix, we report coefficient estimates from the partially observable bivariate probit model, P(Z=1) = P(F=1)P(D=1—F=1). This table provides evidence that Average Similarity Score (as well as top 5, top 10, and top 15 Average Similarity Score) are strongly associated with a decline in fraud commission and weakly related to an increase in fraud detection. These findings are consistent with the indirect channel, that managers understand they are more likely to get caught if they have close benchmarks and respond by committing less fraud.

As an additional test on this front, we exploit the granularity of the data at the competitor-pair level (pairwise observations) and examine the incidence of fraud, conditional on a similar product market rival getting caught. We report results from this analysis in Table A.6 of the Internet Appendix. The findings indicate that a firm is more likely to be accused of fraud if a rival was recently charged with fraudulent reporting practices. This finding is also consistent with prior work indicating that fraud occurs in industry waves (Povel et al., 2007; Wang et al., 2010) or that there is contagion in financial misconduct Dimmock et al. (2018).

IV.2 Product Differentiation and Discipline Channels

Thus far, we have documented a strong mitigating effect of product market similarity (lack of differentiation) on corporate fraud. In this section, we explore two primary channels through which this discipline can originate: managerial slack and benchmarking. We then discuss potential alternatives explanations.

Economists have long argued that product market competition can impose discipline by reducing managerial slack (Machlup, 1967). For example, competition can cause managers to exert more effort by diminishing the benefits of shirking (Hart, 1983) and can reduce resources available for rent extraction (e.g. Ades and Di Tella, 1999; Nickell, 1996). Extending this concept to corporate fraud, competition potentially reduces the economic profits that may be extracted through financial reporting manipulation.[16]. Furthermore, the availability of product market substitutes offered by rivals may exacerbate lost market share due to the reputational costs of fraud. We refer to this channel of product market discipline as the managerial slack effect.

Second, information conveyed by close product market rivals can yield insight about common shocks to production costs and demand, enabling more precise signals of firm-specific performance (Holmstrom, 1982; Nalebuff and Stiglitz, 1983). Along these lines, evidence suggests that closer benchmarks inform boards regarding CEO ability (Murphy, 1986) as well as market-and industry-wide conditions when determining CEO pay (Oyer, 2004). Further, public firms provide a large amount of information through disclosures, which reduces uncertainty (Badertscher et al., 2013). Through a similar process, rivals with significant product market overlap can facilitate monitoring for investors, regulators, and auditors by providing contexts to interpret financial statement (e.g. Hart, 1983; Dyck et al., 2010). For instance, a survey of CFOs by Dichev et al. (2013) indicates that comparability between rival firms is an important means for identifying financial reporting abnormalities.

IV.2.1 Tariffs

To explore whether the discipline effect of product market similarity is driven by variation in managerial slack, we exploit large tariff reductions at the industry level. Tariff reductions have been shown to increase the intensity of foreign competition (Fresard, 2010), which can ultimately decrease managerial slack (Hart, 1983). While firms could respond to changes in foreign competition in the long run by adjusting their product mix, changes in tariff rates directly affect the short- and intermediate-term ability of foreign rivals to offer competitive prices. Changes in tariffs, however, do not directly affect the quality or quantity of readily available information through financial disclosures in 10-Ks. Thus, tariffs provide a good setting to analyze the short-term and intermediate effects of changes in competition that are likely independent of the benchmarking channels.

Following the literature, industry tariff rates are calculated as duties collected by U.S. Customs divided by the value of U.S. imports for consumption. The duties and customs value are collected from the U.S. International Trade Commission. We then aggregate the values from ten-digit U.S. Harmonized System codes to each three-digit SIC, using the concordance table provided by Pierce and Schott (2012). *Tariff shock* is an indica-

[16] For instance, competition can mitigate the benefits of earnings manipulations in order to maintain higher valuations during acquisition activity or capital raising (Shleifer, 2004)

tor variable that takes value of 1 if the 4-year percentage change in tariff rate is the bottom (tercile/quartile/quintile), 0 otherwise.[17]

Table 5 shows the tariff results. Each tariff reduction specification with and without average score. For the quartile specification (column 3), there is some evidence that an increase in foreign competition pressures managers to commit more fraud. For the tercile and quintile specifications (column 1 and 5) the coefficient is also positive but not significant. This is likely a power issue as this industry level measure trades-off having stronger shocks and not having enough industries. Importantly, when we add Average Score, the coefficient is still negative and significant in all three specifications (columns 2,4,6). This tells us two things. First, Average Score is operating through a different channel than tariff reductions and works in the opposite direction. Thus, even controlling for changes to foreign competition, and thus the levels of managerial slack, managers still face discipline from an alternate channel related to product market differentiation.

IV.2.2 Firm Complexity

To explore the benchmarking channel, we study the disciplining effects of product market similarity and a measure of firm complexity. Cohen and Lou (2012) argue complicated firms require more complicated analysis to impound the same piece of information into the price of a firm with multiple operating segments. It stands to reason that regulators, media, and employees can more easily disseminate information for firms with a simple organizational structure, and are therefore, more likely to detect abnormal performance or financial reporting. Thus, for firms with a very simple organizational structure and product mix, the information provided by having similar rivals (benchmarks) would have a lower marginal effect on outsiders monitoring ability. In contrast, complex firms can be very difficult to understand and detect abnormal behavior without a clear benchmark. Thus, having close rivals for complex firms should intuitively provide a larger marginal effect on the ability to detect earnings manipulations.

All else equal, a firm that operates in several product markets has greater scope to conceal financial information. Operating across a multitude of product markets reduces substantive analytic procedures that auditors can perform and will require more subjective and detailed testing. This notion is reflected in the higher audit fees for firms with many segments (Brinn et al., 1994). For example, a firm that competes in pharmaceuticals, manufacturing, and consumer durables, could hide information by shifting resources across segments or using

[17] We thank Chotibhak Jotikasthira for kindly sharing the methodology to calculate tariff shocks.

complex transactions. Furthermore, monitors would need to understand all three industries to confidently detect reporting abnormalities.

As such, we define complexity as the unique number of industries (three-digit SIC codes) in which a firm operates each year. To calculate this value, we sum the number of distinct industries spanned by a firms TNIC-based competitor set. For example, if a firm has three rivals that each operate in a different three-digit SIC code, then we consider that firm to be operating in three distinct markets. A higher score on complexity indicates that a firm operates in an environment where rivals are from many different industries, and thus the firm is likely more diversified and has a complex basket of products that compete across several markets. Our measure of complexity builds on the intuition provided by Cohen and Lou (2012) who measure complexity as whether a firm operates in multiple markets.

We split our sample into quartiles according complexity rankings. Then, we estimate our main specification for the relationship between corporate fraud and product market similarity separately for each quartile. The results are presented in Table 6. In Panel A, we report the average number of unique SIC codes and the number of competitors in each firms TNIC. Each specification is estimated using our full set of control variables, described in Section II and in our analysis of Table 2. We estimate regressions separately for each complexity quartile in Panel B.

Consistent with the benchmarking channel, we find that the disciplining effect of product similarity increases monotonically across complexity quartiles for Panel B. The partial effect for the top quartile is more than four times as large as that for the lowest quartile. To put this finding into perspective, a one standard deviation increase in Average Similarity Score for the least complex firms leads to a decrease in propensity of fraud from 1.9% to approximately 1.6%, or a 0.3 percentage point decline. By comparison, a one standard deviation change in Average Similarity Score for the most complex firms decreases the propensity of fraud from 1.9% to 0.65%, or a 1.25 percentage point decline. This result is robust to several variations of controls for firm size. While large firms are more complex than small firms, on average, our sorts capture product market complexity beyond firm size.

We run the regressions separately for each quartile because it does not constrain coefficients to be the same for all the control variables across quartiles, ensuring the greatest degree of flexibility. Furthermore, the presence of a monotonic relationship across quartiles is strongly supportive of the benchmarking channel. Nonetheless, we test the robustness of our results with an interaction, instead of running separate regressions, and find consistent results. In untabulated results we also perform a Seemingly Unrelated Estimation and confirm that the 4th quartile partial effect is significantly different from the partial effects of the other 3 quartiles (tested jointly) and the 3rd quartile partial effect tested independently. *Additionally, we find qualitatively similar results in untabulated probit and logit specifications.*

IV.2.3 Institutional Ownership and Analyst Coverage

Next, we examine the role that product differentiation plays for two particularly important corporate governance actors with varying degrees of reliance on public financial statements: institutional investors and sell-side analysts. Specifically, we explore whether the marginal disciplining effect of product market similarity is a function of the intensity of institutional ownership and analyst coverage.

The negative effects of corporate fraud on shareholder value can be particularly costly to investors with large ownership stakes, such as institutional investors, thus creating strong monitoring incentives. Furthermore, higher ownership stakes by institutional investors can facilitate access to managers and internal information relevant to fraud (Shleifer and Vishny, 1986; Dyck et al., 2010). This access to inside information can potentially reduce the reliance on public signals of firm performance, such as the information contained in 10-Ks. Thus, if the disciplining effect of product market similarity stems from its informational aspect (i.e. benchmarking), then we should expect to see a muted effect for firms with highly concentrated institutional ownership.

Following Hartzell and Starks (2003), we proxy for institutional ownership influence and monitoring intensity using a Herfindahl Index (HHI) of institutional ownership. Hartzell and Starks (2003) show the concentration of institutional ownership, rather than the level, determines the intensity of institutional monitoring. To explore the differential effect of product similarity, we estimate our primary specifications for subsets of our sample split into quartiles based on Institutional Ownership HHI. As shown in Panel A of Table 7, the partial effect of average score on fraud is much higher for firms with disperse ownership than it is for firms with concentrated ownership. This finding is consistent with the disciplining effect of product market similarity playing a less significant role for firms with highly concentrated institutional owners who have greater access to fraud-relevant information.

Sell-side analysts are also important corporate governance actors that can help detect fraud. For instance, Dyck et al. (2010) find that analysts and auditors uncover roughly 24% of fraud. However, while institutional investors potentially have access to fraud-relevant inside information, "analysts ... [have] to delve through details of financial reports and industry trends to uncover misrepresentations" (Dyck et al., 2010). Thus, the availability of similar product market rivals can provide context through which analysts can better detect abnormal reporting behavior. In addition, similar rivals lower the marginal cost of gathering information for analysts(Engelberg et al., 2018), and can reduce uncertainty (Badertscher et al., 2013). Thus, if analysts rely on publicly available information provided by product market rivals, the marginal disciplining effect of product market similarity would be higher for firms with more analyst coverage.

In Table 7, we present results on fraud for the sample split into quartiles based on analyst coverage. In Panel B, we show that the partial effect of product market differentiation on fraud is significantly larger for firms with greater analyst coverage. Next, we explore differences in results around Reg FD. The notion is that, prior to Reg FD, analysts could have had greater access to non-public information about firms, similar to a concentrated shareholder, but afterwards would rely more on public information, including financial statements. Consistent with this possibility, the results in Panels B and C show that there is a stronger differential impact across quartiles after Reg FD, which is consistent with the complement channel of analysts and product market similarity.

As previously noted, one possible explanation for the difference in the impact of institutional investors and analyst on fraud is the difference in external versus internal access to information. Anecdotally, we use data from (Dyck et al., 2010) on who blows the whistle on corporate fraud and discern whether each case was caught internally (e.g. firm employees or auditors) or externally (e.g. regulators or the media). In Table A.12 of the Internet Appendix, we show that the average score for firms that commit fraud, but are caught by external whistleblowers, have significantly higher average scores than firms caught by internal whistleblowers. This finding is indicative of external monitors relying more on benchmarks. We further conduct a binary split on a measure of financial statement comparability (De Franco et al., 2011) and find that more financial statement comparability appears to complement product market similarity in that the partial effect of product market differentiation on fraud is significantly smaller for firms with more comparable accounting statements (these results are included in Table A.5).

IV.2.4 Shocks to Fraud Detection

To further investigate the benchmarking channel, we exploit potential shocks to fraud detection. If detection becomes unexpectedly more likely, previously less disciplined firms are more likely to be caught for fraud, whereas a detection shock for already disciplined firms should not lead to significantly more fraud discovery.

Our first examination utilizes two events. We again turn to using IPO, as well as M&A,

activity by a firms rivals. In particular, both events plausibly lead to a shock to the firms information environment. We first study the event of IPOs by a firms rivals. These events increase the publicly-available financial information of previously existing, private competitors, which in turn, enhances the ability to assess, compare, and scrutinize a firms own financial statements. Consistent with this view, Bauguess et al. (2018) provide evidence that IPOs lead to in an increase EDGAR traffic for rival firms that are already publicly traded. Next, we study acquisitions by a firms rivals. Acquisitions are material events that can draw considerable scrutiny from investors, analysts, regulators and the media, thus increasing the saliency of existing information in the industry. For instance, acquisitions often occur in waves, suggesting an increase in attention for other firms that could potentially be involved in a deal (Song and Walkling, 2000).

For the IPO tests, we take each pairwise observation of competitors, i and j, and flag whether firm j underwent an IPO in year t. We then aggregate the data to the firm-year level for firm i, counting the number of rivals that underwent an IPO in year t. For robustness we also define a dummy variable (Competitor IPO) equal to 1 if any of a firms rivals underwent an IPO in year t, and 0 otherwise. There are 3.2 rival IPOs per firm-year in our sample, with a median of 0 (43% of firms have at least one rival IPO). Contingent on having at least one IPO rival, each firm has an average 7.6 rivals undertaking IPOs, which is consistent with the documented evidence that IPOs occur in waves (e.g. Lowry and Schwert, 2002).

A competitors IPO is a shock to competition via two channels. First, as discussed, more information about economic conditions becomes publicly available for the rival, as well as increased attention, which enhances monitoring abilities for a firms own investors, regulators, and auditors. Second, the IPO provides a capital injection for a firms competitor, thus enhancing the intensity of competition with that rival. We control for the amount of funds raised by the competitor during the IPO to help separate the shock to a firms information environment from the influence of changes in competitors capital structure and size. While this solution is imperfect, it assists in isolating the effect of rival IPOs due to information rather than the intensity of competition.

We report OLS estimates for the relationship between fraud and rival firm IPOs in Column 1 and 2 of Table A.7. More specifically, we estimate the interaction between the natural log of the number of rival firms undergoing IPOs and a firms product market similarity (Average Similarity Score) prior to the rivals IPOs. A firms pre-existing product market similarity is indicative of the level of market discipline provided by competitors prior to the rivals IPO. The coefficients on the level terms suggest that rival IPOs are positively related to fraud detection and Average Similarity Score continues to exhibit a negative relation to fraud propensity. The positive effect of rival firm IPOs on fraud suggests a shock to detection. In particular, IPOs by rivals change the available information for comparison rather abruptly, before a firm has time to fully wind down financial misconduct.

The coefficient estimate for the interaction term is negative (lower for firms with greater pre-existing product market similarity). This finding suggests that the increased detection resulting from rival IPOs is significantly more pronounced for firms with more product differentiation prior to the rivals IPO (i.e., the effect is more pronounced for ex ante undisciplined firms).[18] We also split the sample between pre-IPO year high and low Average Similarity Score firms to verify that the IPO-detection effect is greater for firms that had lower ex ante discipline due to fewer related firms before the IPO year. Due to a lack of power in the high pre-IPO regression these coefficients are not significantly different from each other. We find that these effects hold after conditioning on the amount of capital raised by rivals during the IPOs. Overall, the findings in this section are consistent with rival IPOs having a greater detection effect for firms with lower pre-existing product market discipline.

Next, for the M&A tests, for each pairwise observation of competitors, i and j, we flag whether firm j was acquired in year t. We then aggregate the data back to the firm-year level for firm i, and take the log of the number of acquired firms competing with firm i that were acquired in year t. For the average firm in our sample, there are 0.059 competitor acquisitions of rival firms each year. Conditional on at least one competitor being acquired, the average increases to 1.09 competitor acquisitions per year.

Column 3 and 4 of Table A.7 reports the results for rival firms being acquired. Much like the IPO results, we find a negative coefficient on average similarity and a positive coefficient for the number of competitors being acquired. The interaction term is also negative, indicating that the partial effect of firms that had higher score is lower when rivals are being acquired. Thus, firms that are ex ante less disciplined exhibit the greatest response to the information/interest generated around takeovers. The split by firms pre-existing similarity scores, highlights that the effect of the acquisitions exists predominantly in the subsample with less similar product market rivals (i.e. those less disciplined ex ante).

IV.2.5 Alternate Measures of Product Similarity

Throughout our analysis, we use firm-level product similarity scores that aggregate a firm's similarity with each of its rivals in a given year using an equally weighted average. As noted

[18]We repeat this test excluding the rival undergoing the IPO from a firms Average Similarity Score calculation to ensure that the similarity with the IPO firm is not driving the results.

in Section III, we implement alternative constructions of using product market similarity to ensure our results are not driven by our main construct.

In Panel A of Table 9, we re-estimate our main specification with each of the alternative aggregation schemes. Given the high correlation between these measures (shown in Table A.2 of the Internet Appendix), it is unsurprising that all variations yield a highly significant negative relationship with fraud. These results mitigate concerns that the equal weighted average potentially obscures the association between fraud and product differentiation. To facilitate comparison the economic magnitudes of the various measures, we estimate a specification in which we standardize all variables and report the results Panel B. The coefficient estimates exhibit monotonicity based on the number of competitors used to compute each firms average product similarity score (i.e., Average Similarity Score < top 15 average < top 10 average < top 5 average). This result is reassuring for our benchmarking hypothesis, as it suggests that the closest rivals provide the largest marginal effect. The effect of the precision score is also similar in magnitude and significance to the overall Average Similarity Score.

IV.2.6 Financial Restatements (Falsification Tests)

Our primary dependent variable is motivated from recent literature that suggests using settled AAER misstatements and fraud-driven class action lawsuits as a proxy for fraud (e.g. Karpoff et al., 2017; Donelson et al., 2017). A large number of accounting restatements are not intentional, material, or damaging to investors. Indeed, in their studies on fraud, Griffin et al. (2018) and McGuire et al. (2011) purge all accounting-related restatements from their data because accounting-related restatements are frequently due to new interpretations or guidance on accounting rules as opposed to firm-level actions. Based on these arguments, we do not expect product differentiation to have significant power in predicting such restatements, unless our main results are driven by omitted firm- or industry-level characteristics.

We perform falsification tests of our main analysis using all (unfiltered) restatements as the dependent variable from the full sample of pre-screened restatements and report the results in Table 10.[19] Our final sample includes 3,623 restatements from 1996-2012 collected from Audit Analytics Database. In general, we do not observe a pattern between financial restatements and product market similarity. This finding suggests product market similarity has no discernable relationship with indirect proxies of fraud that are largely obscured by accidental clerical errors and immaterial accounting differences. While we cannot rule out that the lack of an association between restatements and product differentiation is driven by

[19]We find a weak correlation between restatements and our measure of fraud (0.10).

the particular model specifications that we have chosen, we have explored many variations and failed to find an association in any of the variations that we tried.[20] Overall, this analysis increases our confidence that we have identified an economically meaningful link between fraud and product market similarity in our primary analysis.

V Conclusion

Our paper examines the relationship between an individual firm's competitive landscape and the incidence of corporate financial fraud. Empirically examining such associations is challenging due the multifaceted nature of product market competition, which leads to difficulty in both capturing firm-level, rather than industry-level, values and picking up the aspects of competition that likely play the largest role in the detection and commission of fraud. As the degree of competition is predicated on the substitutability, we use pairwise product market similarity scores developed by Hoberg and Phillips (2010, 2016). This firm-specific measure is based on 10-K descriptions of public rivals, and thus captures publicly-available information about a firm's differentiation.

We find that firms with lower product market differentiation exhibit significantly lower incidences of fraud. The economic magnitude of product market similarity is large compared to many other factors that have previously been explored in the fraud literature. We corroborate these findings using an instrumental variables analysis using rival firm IPOs, which helps mitigate concerns about selection effects and industry heterogeneity.

To ascertain whether the publicly available information about product market rivals captures unique informational content of that is related to fraud, we explore a battery of other measures of competition. These include both traditional measures that have a strong industry component and newer measures that encapsulate intensity of competition, but less about the information gleaned from rivals. Individually, these measures have little predictive power on the incidence of corporate fraud, and using such measures as control variables does not materially affect the coefficient estimates for product market similarity on fraud.

Given this evidence, we argue that product market differentiation plays a unique role in determining fraudulent behavior relative to other characteristics of competition. More specifically, greater product market similarity with rivals could provide helpful information that external constituents use to obtain better signals about firm value or it could impose

[20] In particular, we experimented with various levels of winsorization, with non-linear specifications (e.g. probit and conditional logit), with our set of control variables, and with different sample periods (i.e., before and after the financial crisis) These tests are reported in the Internet Appendix.

managerial discipline by reducing managerial slack. To differentiate these channels, we explore a variety of settings and conduct cross-sectional tests. Using tariff reductions as a change to competition, but no effect on the degree of near term information from financial disclosures, we do not find evidence that this form of competition mitigates the incidence of fraud.

We next turn to cross-sectional tests that aim to capture when having more information about rivals has higher marginal impact reveal that product market similarity is more important when firms are more complex, a firm has more dispersed institutional ownership, and for firms with higher analyst coverage. These tests collectively provide more support for product market similarity resulting in a lower incidence of fraud to to better information that facilitates benchmarking. To corroborate the benchmarking channel, we show that events that could affect the information environment of firms, like M&As and IPOs, are associated with a greater detection effect for firms with lower pre-existing market discipline.

Collectively, our paper provides new insight on how a particular aspect of competition, product market differentiation, influences the incentives to commit fraud by enhancing the ability to benchmark a firm against similar peers. Thus, our paper highlights the role of one market-based mechanism that an affect commission and detection of corporate fraud. Our results suggest that external parties could focus efforts on examining firms with fewer comparable rivals when looking for potential fraudulent reporting.

References

Ades, A., and R. Di Tella. 1999. Rents, Competition, and Corruption. *The American Economic Review* 89:982–993.

Andergassen, R. 2016. Managerial compensation, product market competition and fraud. *International Review of Economics & Finance* 45:1–15.

Badertscher, B., N. Shroff, and H. D. White. 2013. Externalities of Public Firm Presence: Evidence from Private Firms Investment Decisions. *The Journal of Financial Economics* 109:682–706.

Bain, J. S. 1951. Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936-1940. *The Quarterly Journal of Economics* 65:293–324.

Balakrishnan, K., and D. A. Cohen. 2013. Competition and Financial Accounting Misreporting. *Working Paper* .

Bauguess, S. W., J. Cooney, and K. W. Hanley. 2018. Investor Demand for Information in Newly Issued Securities. *Working Paper* .

Beasley, M. S. 1996. An Empirical Analysis of the Relation between the Board of Director Composition and Financial Statement Fraud. *The Accounting Review* 71:443–465.

Bergstresser, D., and T. Philippon. 2006. CEO incentives and earnings management. *Journal of Financial Economics* 80:511–529.

Brinn, T., M. J. Peel, and R. Roberts. 1994. Audit Fee Determinants Of Independent & Subsidiary Unquoted Companies In The Uk – An Exploratory Study. *The British Accounting Review* 26:101–121.

Buzby, S. L. 1975. Company Size, Listed Versus Unlisted Stocks, and the Extent of Financial Disclosure. *Journal of Accounting Research* 13:16–37.

Cheng, I.-H. 2011. Corporate Governance Spillovers. SSRN Scholarly Paper, Social Science Research Network, Rochester, NY.

Chhaochharia, V., Y. Grinstein, G. Grullon, and R. Michaely. 2016. Product Market Competition and Internal Governance: Evidence from the Sarbanes – Oxley Act. *Management Science* 63:1405–1424.

Choi, S. J., K. K. Nelson, and A. C. Pritchard. 2009. The Screening Effect of the Private Securities Litigation Reform Act. *Journal of Empirical Legal Studies* 6:35–68.

Cohen, L., and D. Lou. 2012. Complicated firms. *Journal of Financial Economics* 104:383–400.

De Franco, G., S. P. Kothari, and R. S. Verdi. 2011. The Benefits of Financial Statement Comparability. *Journal of Accounting Research* 49:895–931.

Dechow, P. M., W. Ge, C. R. Larson, and R. G. Sloan. 2011. Predicting Material Accounting Misstatements. *Contemporary Accounting Research* 28:17–82.

Demirg-Kunt, A., and V. Maksimovic. 1998. Law, Finance, and Firm Growth. *The Journal of Finance* 53:2107–2137.

Dichev, I. D., J. R. Graham, C. R. Harvey, and S. Rajgopal. 2013. Earnings quality: Evidence from the field. *Journal of Accounting and Economics* 56:1–33.

Dimmock, S. G., and W. C. Gerken. 2012. Predicting fraud by investment managers. *Journal of Financial Economics* 105:153–173.

Dimmock, S. G., W. C. Gerken, and N. P. Graham. 2018. Is Fraud Contagious? Coworker Influence on Misconduct by Financial Advisors. *The Journal of Finance* 73:1417–1450.

Donelson, D. C., A. Kartapanis, J. McInnis, and C. G. Yust. 2017. Measuring Financial Reporting Fraud Using Public and Private Enforcement. *Working Paper* .

Dyck, A., A. Morse, and L. Zingales. 2010. Who Blows the Whistle on Corporate Fraud? *The Journal of Finance* 65:2213–2253.

Dyck, I. J. A., A. Morse, and L. Zingales. 2013. How Pervasive is Corporate Fraud? *Working Paper* .

Engelberg, J., R. D. McLean, and J. Pontiff. 2018. Analysts and Anomalies. SSRN Scholarly Paper ID 2939174, Social Science Research Network, Rochester, NY.

Faccio, M., and L. Zingales. 2017. Political Determinants of Competition in the Mobile Telecommunication Industry. *Working Paper* .

Farber, D. B. 2005. Restoring Trust after Fraud: Does Corporate Governance Matter? *The Accounting Review* 80:539–561.

Fee, C. E., and S. Thomas. 2004. Sources of gains in horizontal mergers: evidence from customer, supplier, and rival firms. *Journal of Financial Economics* 74:423–460.

Fresard, L. 2010. Financial Strength and Product Market Behavior: The Real Effects of Corporate Cash Holdings. *The Journal of Finance* 65:1097–1122.

Giannetti, M., and T. Y. Wang. 2016. Corporate Scandals and Household Stock Market Participation. *The Journal of Finance* 71:2591–2636.

Gillan, S. L., J. C. Hartzell, and L. T. Starks. 2011. Tradeoffs in Corporate Governance: Evidence From Board Structures and Charter Provisions. *Quarterly Journal of Finance* 01:667–705.

Giroud, X., and H. M. Mueller. 2010. Does corporate governance matter in competitive industries? *Journal of Financial Economics* 95:312–331.

Graham, J. R., C. R. Harvey, and S. Rajgopal. 2005. The economic Implications of Corporate Financial Reporting. *Journal of Accounting and Economics* 40:3–73.

Greenspan, A. 2008. *The Age of Turbulence: Adventures in a New World*. Penguin.

Griffin, J. M., S. Kruger, and G. Maturana. 2018. Do labor markets discipline? Evidence from RMBS bankers. *Journal of Financial Economics* .

Griffin, P. A., J. A. Grundfest, and M. A. Perino. 2004. Stock Price Response to News of Securities Fraud Litigation: An Analysis of Sequential and Conditional Information. *Abacus* 40:21–48.

Grossman, S. J., and O. D. Hart. 1982. Corporate Financial Structure and Managerial Incentives. *University of Chicago Press* pp. 107–14.

Guiso, L., P. Sapienza, and L. Zingales. 2008. Trusting the Stock Market. *The Journal of Finance* 63:2557–2600.

Gurun, U. G., N. Stoffman, and S. E. Yonker. 2018. Trust Busting: The Effect of Fraud on Investor Behavior. *The Review of Financial Studies* 31:1341–1376.

Hart, O. D. 1983. The Market Mechanism as an Incentive Scheme. *The Bell Journal of Economics* 14:366–382.

Hartzell, J. C., and L. T. Starks. 2003. Institutional Investors and Executive Compensation. *The Journal of Finance* 58:2351–2374.

Healy, P. M., and J. M. Wahlen. 1999. A Review of the Earnings Management Literature and Its Implications for Standard Setting. *Accounting Horizons* 13:365–383.

Heflebower, R. B. 1957. Barriers to New Competition. *The American Economic Review* 47:363–371.

Herfindahl, O. 1950. *Concentration in the steel industry.* Columbia Univ., Ph. D. thesis. Columbia University.

Hirschman, A. O. 1945. *National Power and the Structure of Foreign Trade.* University of California Press. Google-Books-ID: BezqxPq50dwC.

Hoberg, G., and G. Phillips. 2010. Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. *The Review of Financial Studies* 23:3773–3811.

Hoberg, G., and G. Phillips. 2016. Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy* 124:1423–1465.

Hoberg, G., G. Phillips, and N. Prabhala. 2014. Product Market Threats, Payouts, and Financial Flexibility. *The Journal of Finance* 69:293–324.

Holmstrom, B. 1982. Moral Hazard in Teams. *The Bell Journal of Economics* 13:324–340.

Hsu, C., X. Li, Z. Ma, and G. M. Phillips. 2017. Does Product Market Competition Influence Analyst Coverage and Analyst Career Success? *Working Paper* .

Jayaraman, S., and T. Milbourn. 2009. Does equity-based CEO compensation really increase litigation risk. *Working Paper* .

Karaoglu, E., T. Sandino, and R. P. Beatty. 2006. Benchmarking Against the Performance of High Profile 'Scandal' Firms. *Working Paper* .

Karpoff, J. M., A. Koester, D. S. Lee, and G. S. Martin. 2017. Proxies and Databases in Financial Misconduct Research. *The Accounting Review* 92:129–163.

Karpoff, J. M., D. S. Lee, and G. S. Martin. 2008. The Cost to Firms of Cooking the Books. *The Journal of Financial and Quantitative Analysis* 43:581–611.

Kedia, S., and T. Philippon. 2009. The Economics of Fraudulent Accounting. *The Review of Financial Studies* 22:2169–2199.

Khanna, V., E. H. Kim, and Y. Lu. 2015. CEO Connectedness and Corporate Fraud. *The Journal of Finance* 70:1203–1252.

Koh, P.-S., and D. M. Reeb. 2015. Missing R&D. *Journal of Accounting and Economics* 60:73–94.

Li, F., R. Lundholm, and M. Minnis. 2013. A Measure of Competition Based on 10-K Filings. *Journal of Accounting Research* 51:399–436.

Lowry, M., and G. W. Schwert. 2002. IPO Market Cycles: Bubbles or Sequential Learning? *The Journal of Finance* 57:1171–1200.

Machlup, F. 1967. Theories of the Firm: Marginalist, Behavioral, Managerial. *The American Economic Review* 57:1–33.

McGuire, S. T., T. C. Omer, and N. Y. Sharp. 2011. The Impact of Religion on Financial Reporting Irregularities. *The Accounting Review* 87:645–673.

Murphy, K. J. 1986. Incentives, Learning, and Compensation: A Theoretical and Empirical Investigation of Managerial Labor Contracts. *The RAND Journal of Economics* 17:59–76.

Nalebuff, B. J., and J. E. Stiglitz. 1983. Prizes and Incentives: Towards a General Theory of Compensation and Competition. *The Bell Journal of Economics* 14:21–43.

Nickell, S. J. 1996. Competition and Corporate Performance. *Journal of Political Economy* 104:724–746.

Oyer, P. 2004. Why Do Firms Use Incentives That Have No Incentive Effects? *The Journal of Finance* 59:1619–1650.

Pierce, J. R., and P. K. Schott. 2012. A concordance between ten-digit US Harmonized System Codes and SIC/NAICS product classes and industries. *Journal of Economic and Social Measurement* 37:61–96.

Povel, P., R. Singh, and A. Winton. 2007. Booms, Busts, and Fraud. *The Review of Financial Studies* 20:1219–1254.

Rangan, S. 1998. Earnings management and the performance of seasoned equity offerings. *Journal of Financial Economics* 50:101–122.

Reynolds, J. K., and J. R. Francis. 2000. Does size matter? The influence of large clients on office-level auditor reporting decisions. *Journal of Accounting and Economics* 30:375–400.

Richardson, S. A., R. G. Sloan, M. T. Soliman, and . Tuna. 2005. Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics* 39:437–485.

Schmidt, K. M. 1997. Managerial Incentives and Product Market Competition. *The Review of Economic Studies* 64:191–213.

Shleifer, A. 2004. Does Competition Destroy Ethical Behavior? *Working Paper* .

Shleifer, A., and R. W. Vishny. 1986. Large Shareholders and Corporate Control. *Journal of Political Economy* 94:461–488.

Shleifer, A., and R. W. Vishny. 1997. A survey of Corporate Governance. *The Journal of Finance* LII:737–783.

Shroff, N., R. S. Verdi, and B. P. Yost. 2017. When does peer information matter. *Journal of Accounting and Economics* .

Smith, A. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. William Strahan, Thomas Cadell.

Song, M. H., and R. A. Walkling. 2000. Abnormal returns to rivals of acquisition targets: A test of the 'acquisition probability hypothesis'. *Journal of Financial Economics* 55:143–171.

Teoh, S. H., I. Welch, and T. Wong. 1998. Earnings Management and the Long-Run Market Performance of Initial Public Offerings. *The Journal of Finance* 53:1935–1974.

Thompson, R. B., and H. A. Sale. 2003. Securities Fraud as Corporate Governance: Reflections upon Federalism. *Vanderbilt Law Review* 56:859–910.

Tirole, J. 1988. *The Theory of Industrial Organization*. MIT Press.

Tirole, J. 2010. *The Theory of Corporate Finance*. Princeton University Press.

Wang, T. Y. 2011. Corporate Securities Fraud: Insights from a New Empirical Framework. *The Journal of Law, Economics, and Organization* 29:535–568.

Wang, T. Y., and A. Winton. 2014. Product Market Interactions and Corporate Fraud. *Working Paper* .

Wang, T. Y., A. Winton, and X. Yu. 2010. Corporate Fraud and Business Conditions: Evidence from IPOs. *The Journal of Finance* 65:2255–2292.

Whited, T. M., and G. Wu. 2006. Financial Constraints Risk. *The Review of Financial Studies* 19:531–559.

Table 1: Summary Statistics

This table reports summary statistics of firm characteristics at the firm-year level. Variable definitions are provided in the Appendix. Our sample spans 1996 through 2011.

Variable	No. Obs	Mean	Std. Dev.	10th Percentile	90th Percentile
AAER Misstatement	55,381	0.014	0.119	0.000	0.000
SCAC	55,381	0.006	0.074	0.000	0.000
Fraud	55,381	0.019	0.135	0.000	0.000
Avg Similarity Score	55,381	0.030	0.023	0.012	0.055
Avg Top5 Similarity	55,381	0.080	0.058	0.017	0.156
Avg Top10 Similarity	55,381	0.066	0.050	0.014	0.135
Avg Top15 Similarity	55,381	0.059	0.047	0.013	0.123
Avg Score Precision	55,381	1.002	0.103	0.924	1.053
Sum Similarity	55,381	2.847	4.999	0.074	7.659
Product Market Fluidity	50,402	7.182	3.292	3.292	11.685
SIC3 HHI	55,381	0.176	0.145	0.062	0.332
SIC3 Profit Margin	55,381	-0.039	0.272	-0.346	0.156
TNIC HHI	55,381	0.235	0.197	0.064	0.518
NCOMP TNIC	55,381	74.204	90.520	5.000	204.000
NCOMP SIC3	55,381	121.607	170.694	6.000	351.000
RSST accruals	51,487	0.024	0.240	-0.182	0.220
Change AR	55,381	0.010	0.065	-0.045	0.070
Change Inventory	55,060	0.006	0.049	-0.028	0.050
Pct Soft Assets	55,377	0.541	0.245	0.175	0.852
Change in Cash Sales	51,888	0.195	0.710	-0.214	0.574
ROA	51,497	-0.005	0.195	-0.205	0.141
Change in ROA	54,671	-0.007	0.175	-0.149	0.120
Change in employee	54,053	-0.080	0.469	-0.365	0.241
Dummy Security Issue	55,381	0.920	0.272	1.000	1.000
Whited-Wu Index	54,954	-0.196	0.198	-0.389	0.012
Book Leverage	55,237	0.299	0.294	0.000	0.733
Capex	55,381	0.060	0.093	0.000	0.140
R&D	55,381	0.069	0.117	0.000	0.184
R&D dummy	55,381	0.627	0.484	0.000	1.000
Age	53,295	15.353	11.825	4.0000	35.000
Inst Ownership	43,018	0.516	0.315	0.068	0.922
Number of Analysts	55,381	5.837	7.008	0.000	15.000
Stock Industry Return R2	53,238	0.342	0.173	0.121	0.580
Relative Perf Eval Flag	55,179	0.677	0.467	0.000	1.000
Ln Asset	55,381	5.618	1.937	3.155	8.181

Table 2: Product Market Differentiation and Corporate Fraud

This table reports OLS estimates for the incidence of fraud on the average similarity of each firms rivals. Our proxy for corporate fraud includes a combination of AAER misstatements from the AAER dataset and Securities Class Actions from the Stanford University Lawsuit Database. The specification in Column 1 does not include control variables. The specification in Column 2 includes controls used in Dechow et al. (2011). In Columns 3-5 we include our full set of controls as described in Section II and Column 3 also includes Institutional Ownership. In Column 6 we report the standardized regression. All specifications are run at the firm-year level, include year fixed effects, and include explanatory variables are lagged by one year. Column 4 also includes three-digit SIC code (SIC3) fixed effects, Column 5 adds year SIC3 fixed effects. In Column 6, we run the specification from Column 4 but with standardized regressors. The t-statistics, calculated from standard errors clustered at SIC3 level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
						Std Reg
	Fraud	Fraud	Fraud	Fraud	Fraud	Fraud
Avg Similarity Score	-0.113**	-0.180***	-0.220***	-0.171***	-0.163***	-0.004***
	(-2.089)	(-3.517)	(-3.915)	(-3.946)	(-3.243)	(-3.946)
Ln Asset		0.006***	0.005***	0.006***	0.007***	0.011***
		(4.617)	(3.245)	(2.913)	(3.546)	(2.913)
R&D			0.009	-0.011	-0.012	-0.001
			(0.740)	(-1.008)	(-1.143)	(-1.008)
R&D dummy			-0.000	-0.003	-0.004	-0.003
			(-0.011)	(-0.914)	(-1.135)	(-0.914)
Ln number analysts			0.001	0.001	0.001	0.001
			(0.227)	(0.267)	(0.392)	(0.267)
Inst Ownership			0.007			
			(0.792)			
Whited-Wu Index			0.005	-0.000	0.033*	-0.000
			(0.859)	(-0.057)	(1.892)	(-0.057)
RSST accruals		0.002	-0.003	0.001	0.002	0.000
		(0.538)	(-0.636)	(0.342)	(0.572)	(0.342)
Change AR		0.022*	0.023	0.016	0.023*	0.001
		(1.791)	(1.460)	(1.296)	(1.808)	(1.296)
Change Inventory		0.016	0.026	0.020	0.027	0.001
		(0.756)	(1.258)	(0.937)	(1.225)	(0.937)
Pct Soft Assets		0.019***	0.022***	0.019***	0.019***	0.005***
		(4.241)	(4.186)	(3.825)	(3.733)	(3.825)
Change in Cash Sales		0.005**	0.004**	0.005**	0.005**	0.003**
		(2.250)	(2.057)	(2.227)	(2.367)	(2.227)
Change in ROA		-0.023***	-0.017***	-0.021***	-0.018***	-0.004***
		(-6.132)	(-3.336)	(-5.887)	(-5.212)	(-5.887)
Change in employee		-0.004**	-0.004*	-0.003	-0.003	-0.001
		(-2.101)	(-1.868)	(-1.456)	(-1.414)	(-1.456)
Ln Age		-0.010***	-0.009***	-0.008***	-0.007***	-0.006***
		(-3.413)	(-3.040)	(-3.513)	(-2.974)	(-3.513)
Dummy Security Issue		0.003	-0.001	0.002	0.001	0.002
		(1.327)	(-0.258)	(0.737)	(0.275)	(0.737)
Stock Industry Return R2			-0.009	0.017		0.003
			(-0.822)	(1.515)		(1.515)
Relative Perf Eval Flag			0.007**			
			(2.124)			
TNIC NCOMP			0.001	0.002	0.001	0.002
			(1.205)	(0.945)	(0.756)	(0.945)
Constant	0.015***	-0.004	-0.001	-0.009	-0.007	0.020***
	(6.859)	(-0.708)	(-0.129)	(-0.707)	(-0.635)	(7.128)
Observations	50,526	39,519	28,912	37,144	38,916	37,144
R-squared	0.005	0.015	0.014	0.034	0.079	0.034
FE	Year	Year	Year	Year+Sic3	Year#Sic3	Year+Sic3
Cluster	sic3	sic3	sic3	sic3	sic3	sic3

Table 3: Shock to Product Market Differentiation from Rival IPO

This table reports 2SLS estimates for the relationship between product market similarity and corporate fraud. In columns 1-3, we report the first stage result for 2SLS regression namely the relationship between a firm's IPO-rival's lagged similarity score on the firm's overall average score. In columns 4-6, we use similarity scores with competitors undergoing an IPO as an instrument for the firms Average Similarity Score on fraud. Our proxy for corporate fraud includes a combination of AAER misstatements from the AAER dataset and Securities Class Actions from the Stanford University Lawsuit Database. T-statistics, calculated from standard errors clustered at the three-digit SIC code (SIC3) level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	Avg Similarity Score			Fraud		
IPO Avg Score	0.342***	0.291***	0.216***			
	(8.424)	(10.554)	(9.382)			
Avg Similarity Score (Inst)				-0.618***	-0.530***	-0.495**
				(-4.498)	(-3.181)	(-2.361)
Ln Num Competitor IPO	-0.003***	-0.004***	-0.002***	0.006***	0.005**	0.004**
	(-3.002)	(-6.682)	(-4.426)	(3.106)	(2.477)	(2.122)
IPO Size ($)	0.001**	-0.001*	-0.001***	0.000	-0.000	-0.000
	(2.304)	(-1.741)	(-2.654)	(0.037)	(-0.403)	(-0.722)
Ln Asset	0.001***	0.002***	0.001***	0.007***	0.006***	0.006***
	(3.988)	(6.760)	(7.959)	(4.543)	(3.952)	(3.893)
Ln Age	-0.002***	-0.002***	-0.002***	-0.010***	-0.009***	-0.009***
	(-2.741)	(-3.873)	(-5.814)	(-4.135)	(-4.025)	(-3.994)
Observations	37,335	37,335	37,335	37,335	37,335	37,335
R-squared	0.168	0.229	0.352	0.010	0.014	0.033
Controls	No	Full	Full	No	Full	Full
FE	Year	Year	Year+Sic3	Year	Year	Year+Sic3
Cluster	sic3	sic3	sic3	sic3	sic3	sic3

Table 4: Product Market Differentiation and Corporate Fraud
(Controlling for Alternative Measures of Competition)

This table reports OLS estimates for the incidence of fraud on Average Similarity Score, while controlling for alternative measures of competition. Our proxy for corporate fraud includes a combination of misstatements from the AAER dataset and Securities Class Actions from the Stanford University Lawsuit Database. Column 1 includes sales based Herfindahl-Hirschman Index (HHI) according to three digit SIC code (SIC3). Column 2 also includes the number of competitors (logged) in the same SIC3. Column 3 also includes the profit margin and an industry concentration measure. In Column 4 we include the sales based HHI according to the firm's TNIC. Column 5 also includes the number of competitors within a firm's TNIC. Column 6 also includes the sum similarity score. In Column 7 we included a 10-k based competition measure from Li et al. (2013) which is only available for a subset of our sample. The specifications include the full set of controls as described in Section II. All specifications are run at the firm-year level, include year and SIC3 fixed effects, and explanatory variables lagged by one year. The t-statistics, calculated from standard errors clustered at SIC3 level, are reported in parentheses below coefficient estimates Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

	(1) Fraud	(2) Fraud	(3) Fraud	(4) Fraud	(5) Fraud	(6) Fraud	(7) Fraud
Avg Similarity Score	-0.159***	-0.161***	-0.159***	-0.161***	-0.171***	-0.183***	-0.172**
	(-3.317)	(-3.305)	(-3.248)	(-3.518)	(-3.936)	(-4.209)	(-2.264)
SIC3 HHI	0.018	0.035	0.009				
	(0.842)	(1.435)	(0.383)				
SIC3 NCOMP		0.016**	0.016***				
		(2.155)	(2.680)				
SIC3 PM sale			-0.007				
			(-0.700)				
SIC3 Top 4 Concentration			0.042**				
			(2.121)				
TNIC HHI				-0.002	0.006	0.006	
				(-0.286)	(0.968)	(0.944)	
TNIC NCOMP					0.002	0.002	
					(1.107)	(0.912)	
Product Market Fluidity						0.001	
						(1.373)	
Competition 10K							0.004
							(1.074)
Controls	Full	Full	Full	Full	Full	Full	Full
Observations	37,144	37,144	37,144	37,144	37,144	35,999	18,696
R-squared	0.034	0.034	0.034	0.034	0.034	0.035	0.048
FE	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3
Cluster	sic3	sic3	sic3	sic3	sic3	sic3	sic3

Table 5: Product Market Differentiation, Tariff Reductions and and Corporate Fraud

This table reports OLS estimates for the incidence of fraud on the average similarity of each firms rivals combined with large industry level tariff reductions. Our proxy for corporate fraud includes a combination of AAER misstatements from the AAER dataset and Securities Class Actions from the Stanford University Lawsuit Database. We aggregate ten-digit U.S. Harmonized System codes to find year-SIC3 level tariff levels and identify large year over year tariff reductions. Columns 1 and 2 use the top tercile of reductions, 3 and 4 the top quartile and 5 and 6 the top quintile. Regressions 1,3, and 5 are run without average similarity score while 2, 4, and 6 include it. All specifications include our full set of controls as described in Section II. Regressions include Year and SIC3 fixed effects. The t-statistics, calculated from standard errors clustered at SIC3 level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	Fraud	Fraud	Fraud	Fraud	Fraud	Fraud
Avg Similarity Score		-0.172**		-0.173**		-0.172**
		(-2.453)		(-2.462)		(-2.446)
Big Tariff Reduction (Tercile)	0.008	0.008				
	(1.582)	(1.612)				
Big Tariff Reduction (Quartile)			0.009**	0.009**		
			(2.072)	(2.117)		
Big Tariff Reduction (Quintile)					0.006	0.006
					(1.410)	(1.418)
Observations	22,636	22,636	22,636	22,636	22,636	22,636
R-squared	0.036	0.037	0.036	0.037	0.036	0.036
FE	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3
Cluster	sic3	sic3	sic3	sic3	sic3	sic3

Table 6: Product Differentiation and Fraud by Complexity Quartiles

This table reports OLS estimates for the incidence of fraud on the average similarity of each firms rivals split into complexity quartiles. Our proxy for corporate fraud includes a combination of AAER misstatements from the AAER dataset and Securities Class Actions from the Stanford University Lawsuit Database. We define complexity as the number of unique SIC codes spanned by a firms set of competitors according to the TNIC developed by Hoberg and Phillips, 2016. Panel A reports competitor and fraud classifications for each quartile. Panel B reports OLS estimates for each quartile including our full set of control variables described in Section II. The t-statistics, calculated from standard errors clustered at the three-digit SIC code (SIC3) level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

Complexity	Low			High
	Q1	Q2	Q3	Q4
Panel A				
Unique SICs in TNIC	3.3	8.2	13	22.5
Competitors in TNIC	13	49	117	150
% Fraud	1.6	1.7	1.9	2.1
Avg Similarity Score	2.7	2.8	3.3	3.5
Panel B				
Avg Similarity Score	-0.168***	-0.197**	-0.201	-0.683***
	(-3.418)	(-2.029)	(-1.508)	(-5.053)
Observations	9,995	9,628	9,018	8,503
R-squared	0.016	0.017	0.024	0.026
FE	Year	Year	Year	Year
Controls	Full	Full	Full	Full

Table 7: Product Differentiation and Fraud by Institutional Ownership and Analyst Coverage Splits

This table reports OLS estimates for the incidence of fraud on the average similarity of each firms rivals split into groups based on Institutional Ownership HHI and Analyst Coverage. Our proxy for corporate fraud includes a combination of AAER misstatements from the AAER dataset and Securities Class Actions from the Stanford University Lawsuit Database. Panel A reports OLS estimates for each Institutional ownership HHI quartile including our full set of control variables described in Section II. Panel B contains the same but splits by quartiles based on the number of sell-side analysts covering a firm in a given year and whether that firm is covered by at least one star analyst or not. Panel's C and D repeat the analysis in Panel B but look only at observations either before or after REG FD in 2000. The t-statistics, calculated from standard errors clustered at the three-digit SIC code (SIC3) level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

Panel A Institutional Ownership HHI						
	Low Q1	Q2	Q3	High Q4		
Avg Similarity Score	-0.425*** (-3.373)	-0.298*** (-3.145)	-0.107** (-2.175)	-0.066 (-1.217)		
Observations	7,943	8,225	8,085	7,153		
R-squared	0.023	0.022	0.012	0.007		
Panel B Analyst Coverage						
	Low Q1	Q2	Q3	High Q4	Non Star	Star
Avg Similarity Score	-0.001 (-0.014)	-0.204** (-2.417)	-0.456*** (-3.940)	-0.487*** (-4.378)	-0.133* (-1.737)	-0.388*** (-3.985)
Observations	7,742	7,179	6,420	6,450	12,944	7,221
R-squared	0.006	0.014	0.038	0.028	0.011	0.028
Panel C Analyst Coverage - Pre 2000						
Analyst Coverage	Low Q1	Q2	Q3	High Q4	Non Star	Star
Avg Similarity Score	0.058 (1.037)	-0.267** (-2.170)	-0.786*** (-3.296)	-0.581** (-2.487)	-0.336** (-2.442)	-0.508* (-1.774)
Observations	2,704	2,580	2,263	2,201	3,275	2,167
R-squared	0.012	0.020	0.072	0.039	0.024	0.057
Panel D Analyst Coverage - Post 2000						
Analyst Coverage	Low Q1	Q2	Q3	High Q4	Non Star	Star
Avg Similarity Score	-0.041 (-0.730)	-0.164 (-1.533)	-0.218* (-1.778)	-0.426*** (-2.939)	-0.068 (-0.830)	-0.312*** (-2.880)
Observations	5,038	4,599	4,157	4,249	9,669	5,054
R-squared	0.008	0.013	0.025	0.029	0.011	0.022
FE	Year	Year	Year	Year	Year	Year
Controls	Full	Full	Full	Full	Full	Full

Table 8: IPOs and Acquisitions of Rivals as Change to Information Environment

This table reports OLS estimates for the association between fraud and rival IPOs (M&A) activity including our full set of control variables described in Section II.The specifications include rival firm IPO activity in columns 1 and 2 and M&A activity in panels 3 and 4. We split the data by high and low non-IPO (non-acquired) similarity scores in year t-1. All specifications include year fixed effects and three-digit SIC code (SIC3) fixed effects, and all control variables are lagged one year. The t-statistics, calculated from standard errors clustered at the SIC3 level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

	(1)	(2)	(3)	(4)
	Low Non-IPO Rival Score Fraud	High Non-IPO Rival Score Fraud	Low Non-M&A Rival Score Fraud	High Non-M&A Rival Score Fraud
Ln Num Competitor IPO	0.005***	0.004		
	(2.820)	(1.440)		
IPO Size ($)	0.000	0.000		
	(-0.02)	(-0.72)		
Ln Num Competitor Target			0.103***	0.000
			(3.141)	(-0.024)
Ln Target MarketCap			-0.008***	0.002
			(-2.829)	(1.380)
Observations	18858	18279	18672	18449
R-squared	0.05	0.037	0.052	0.039
Controls	Full	Full	Full	Full
FE	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3

Table 9: Product Market Differentiation and Corporate Fraud
(Alternate Constructions of Independent Variable)

This table reports OLS estimates for the incidence of fraud on alternative constructions of our primary independent variable. Our proxy for corporate fraud includes a combination of AAER misstatements from the AAER dataset and Securities Class Actions from the Stanford University Lawsuit Database. Panel A reports OLS regressions and Panel B reports the regressions with standardized regressors. Column 1 presents results for the main dependent variable used throughout our analysis. Columns 2-4 replace Average Similarity Score with a firms product market similarity score averaged across its closest 15, 10, and 5 competitors, respectively. In Column 5, we replace Average Similarity Score with the Precision measure outlined in section III. The unit of observation in this analysis is the firm-year. All specifications include the full set of controls as described in Section II. They include year and SIC3 fixed effects, and the explanatory variables lagged by one year. The t-statistics, calculated from standard errors clustered at three digit SIC code (SIC3) level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

	(1) Fraud	(2) Fraud	(3) Fraud	(4) Fraud	(5) Fraud
Panel A OLS Regression					
Avg Similarity Score	-0.171*** (-3.946)				
Avg Top 15 Similarity		-0.089** (-2.020)			
Avg Top 10 Similarity			-0.091** (-2.244)		
Avg Top 5 Similarity				-0.086*** (-2.730)	
Avg Score Precision					-0.037*** (-3.893)
Observations	37,144	37,144	37,144	37,144	37,144
R-squared	0.034	0.034	0.034	0.034	0.034
Panel B Standardized Regression					
Avg Similarity Score	-0.004*** (-3.946)				
Avg Top 15 Similarity		-0.004** (-2.020)			
Avg Top 10 Similarity			-0.005** (-2.244)		
Avg Top 5 Similarity				-0.005*** (-2.730)	
Avg Score Precision					-0.004*** (-3.893)
Observations	37,144	37,144	37,144	37,144	37,144
R-squared	0.034	0.034	0.034	0.034	0.034
FE	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3
Cluster	sic3	sic3	sic3	sic3	sic3

Table 10: Product Similarity and Restatements
(Falsification Test)

This table reports OLS estimates for the incidence of accounting restatement on the average similarity of each firms rivals. Our primary dependent variable accounting restatement (RE AA) is obtained from Audit Analytics database from 1996-2012. We do not perform any screens for our restatement variable. The specification in Column 1 does not include control variables. The specification in Column 2 includes controls used in Dechow et al. (2011). In Columns 3-5 we include our full set of controls as described in Section II and Column 3 also includes Institutional Ownership. In Column 6 we report the standardized regression. All specifications are run at the firm-year level, include year fixed effects, and include explanatory variables are lagged by one year. Column 4 also includes three-digit SIC code (SIC3) fixed effects, Column 5 adds year SIC3 fixed effects. In Column 6, we run the specification from Column 4 but with standardized regressors. The t-statistics, calculated from standard errors clustered at SIC3 level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

	(1) RE_AA	(2) RE_AA	(3) RE_AA	(4) RE_AA	(5) RE_AA	(6) RE_AA
Avg Similarity Score	-0.065	-0.117	-0.053	-0.098	-0.103	-0.002
	(-0.722)	(-1.387)	(-0.498)	(-0.908)	(-0.984)	(-0.908)
R&D			-0.019	-0.009	-0.008	-0.001
			(-1.108)	(-0.593)	(-0.562)	(-0.593)
R&D Dummy			0.012**	0.010	0.014*	0.010
			(2.136)	(1.418)	(1.954)	(1.418)
Ln number analysts			0.003	0.002	0.002	0.002
			(0.800)	(0.650)	(0.628)	(0.650)
Inst Ownership			0.021*			
			(1.714)			
Whited-Wu Index			-0.015	-0.008	-0.038	-0.002
			(-1.623)	(-1.202)	(-0.866)	(-1.202)
RSST accruals		-0.001	-0.007	-0.004	-0.002	-0.001
		(-0.187)	(-1.350)	(-0.937)	(-0.343)	(-0.937)
Change AR		-0.032**	-0.028*	-0.028*	-0.020	-0.002*
		(-2.200)	(-1.735)	(-1.964)	(-1.289)	(-1.964)
Change Inventory		0.029	0.016	0.025	0.034	0.001
		(1.268)	(0.681)	(1.163)	(1.575)	(1.163)
Pct Soft Assets		-0.001	-0.014	-0.012	-0.011	-0.003
		(-0.146)	(-1.168)	(-1.141)	(-1.013)	(-1.141)
Change in Cash Sales		-0.001	0.000	-0.000	-0.000	-0.000
		(-0.393)	(0.190)	(-0.352)	(-0.015)	(-0.352)
Change in ROA		-0.002	-0.009*	-0.003	-0.005	-0.000
		(-0.530)	(-1.896)	(-0.580)	(-1.241)	(-0.580)
Change in employee		-0.001	-0.005**	-0.003	-0.002	-0.001
		(-0.819)	(-2.008)	(-1.463)	(-1.084)	(-1.463)
Ln Age		0.002	0.004	0.004	0.005	0.003
		(0.580)	(0.951)	(0.985)	(1.151)	(0.985)
Dummy Security Issue		0.001	0.001	0.001	0.002	0.001
		(0.159)	(0.147)	(0.220)	(0.319)	(0.220)
Stock Ind Return R2			0.001	0.003		0.000
			(0.030)	(0.155)		(0.155)
RPE Flag			0.003			
			(0.466)			
TNIC NCOMP			-0.005**	-0.002	-0.001	-0.002
			(-2.167)	(-0.565)	(-0.287)	(-0.565)
Ln Asset		0.011***	0.007**	0.009***	0.008***	0.017***
		(6.440)	(2.381)	(3.745)	(3.484)	(3.745)
Constant	0.012***	-0.042***	0.012	0.004	-0.003	0.055***
	(4.165)	(-3.392)	(0.618)	(0.183)	(-0.124)	(8.703)
Observations	50,526	39,519	28,912	37,144	38,916	37,144
R-squared	0.017	0.021	0.022	0.039	0.103	0.039
FE	Year	Year	Year	Year+Sic3	Year#Sic3	Year+Sic3
Cluster	sic3	sic3	sic3	sic3	sic3	sic3

Table 11: Alternative Measures of Competition and Corporate Fraud

This table reports OLS estimates for the incidence of fraud on commonly used industry-level proxies for competition. Our proxy for corporate fraud includes a combination of AAER misstatements from the AAER dataset and Securities Class Actions from the Stanford University Lawsuit Database. Our measures of competition include: sales based HHI, average profit margin, top-4 sales concentration and number of competitors constructed using three-digit SIC code, sales based HHI and number of competitors according to TNIC3, product market fluidity and the 10-K-based competition word measure. Columns 1-8 include the full set of controls as described in Section II. The firm-year is the unit of observation in this analysis. All specifications include year and SIC3 fixed effects, and control variables lagged by one year. The t-statistics, calculated from standard errors clustered at the SIC3 level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

	(1) Fraud	(2) Fraud	(3) Fraud	(4) Fraud	(5) Fraud	(6) Fraud	(7) Fraud	(8) Fraud
SIC3 HHI	0.019							
	(0.903)							
SIC3 PM sale		-0.021						
		(-1.453)						
SIC3 Top 4 Concentration			0.035*					
			(1.770)					
SIC3 NCOMP				0.013*				
				(1.785)				
TNIC HHI					0.000			
					(0.008)			
TNIC NCOMP						0.001		
						(0.633)		
prodmktfluid							0.001	
							(0.951)	
Competition 10K								0.004
								(1.065)
Observations	37,144	37,144	37,144	37,144	37,144	37,144	35,999	18,696
R-squared	0.033	0.034	0.034	0.034	0.033	0.033	0.034	0.048
Controls	Full	Full	Full	Full	Full	Full	Full	Full
FE	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3
Cluster	sic3	sic3	sic3	sic3	sic3	sic3	sic3	sic3

Internet Appendix for: Product Differentiation, Information, and Fraud

Table A.1: Variable Description

Variable	Definitions
AAER Misstatement	Equal to 1 for firm-years for which firms have settled with the SEC for corporate Fraud. Note: This is not the actual settlement year, which is usually several years after the alleged fraud, but the year in which the fraud allegedly occurred.
SCAC	Securities and Class Action Equal to 1 for all firm-years for which firms settle a securities class action lawsuit for an alleged 10B-5 fraud allegation.
Fraud	Equal to 1 for all firm-years with an AAER or SCA.
SIC3 HHI	Herfindahl-Hirschman index based on firm sales and three-digit SIC code industry classifications
TNIC HHI	Herfindahl-Hirschman index based on firm sales Text-based Network Industry classifications (TNIC) from Hoberg and Phillips.
Avg Similarity Score	Mean Hoberg and Phillips Similarity Score for all rivals within each firm-years TNIC
Precision	Defined as $(\frac{1}{NCOMP_{TNIC}} \times \sum \frac{1}{(1-score)^2})^{\frac{1}{2}}$
Profit Margin	Average EBITDA/sales ratio for firms within each three-digit SIC code
Top 4 Concentration	Proportion of sales within a three-digit SIC code attributable to the four largest firms within an industry
Age	Number of years the firm has been in Compustat
Analyst Num	Number of analysts covering the firm in each year from IBES (0 if missing).
Inst Ownership	Percentage of shares outstanding held by 13-F institutions
Assets	Total Assets
Capex	Capital Expenditures / log Assets
Book Leverage	(Total Long-Term Debt +Debt in Current Liabilities)/ log Assets
ROA	Net Income / Assets
EFN	Equity Finance Needed defined as ROA/(1 - ROA).
RSST Accruals	$(NOA_t - NOA_t - 1)/NOA_t - 1$. NOA (Net Operating Assets) = OA- OL where OA (Operating Assets) = sum of COA (current operating assets) and NCOA (non-current operating assets) and OL = sum of COL (current operating liabilities) and NCOL (noncurrent liabilities). COA = Current Assets - Cash and Short-Term Equivalents. NCOA = Total Assets - Current Assets - Investments and Advances. COL = Current Liabilities - Debt in Current Liabilities. NCOL = Total Liabilities - Current Liabilities - Long-Term Debt
Dummy Security Issue	An indicator variable equal to 1 if the firm issued securities during year
Change AR	Change in Accounts Receivable/Total Assets
Change Inventory	Change in Inventory/Total Assets
Pct Soft Assets	(Total Assets - PP&E- Cash and Cash Equivalent)/Total Assets
Change in Cash Sales	Percentage change in Cash Sales - Change in Accounts Receivable
Change in ROA	Change in Return on Assets
Change in Employee	Percentage change in the number of employees - percentage change in assets
R&D	Research and Development scaled by assets. Missing observations are filled with either the firm average, if a time series exists, or the industry average if not.
R&D (dummy)	Equal to 1 if R&D is missing and 0 otherwise. We follow Koh and Reeb (2015) when using R&D.
NCOM SIC3	Number of competitors within the three-digit SIC Code.
Ind R2	Following Wang and Winton (2014), we first regress each firms daily stock returns on the weighted-average daily market return and the weighted-average daily industry return. Then, we take the average r-squared for each firm in each three-digit SIC code.
RPE Flag	See Page 8
NCOMP TNIC	Number of competitors according to Text-based Network Industry classifications (TNIC) from Hoberg and Phillips.

Table A.2: Correlations

Correlation coefficients are reported for various measures of product market similarity and competition. Our sample covers 1996 through 2011.

	Avg Sim. Score	Top 15 Similarity	Top 10 Similarity	Top 5 Similarity	Sim. Precision	Sum Similarity	SIC3 HHI	SIC3 PM	SIC3 NCOMP	TNIC HHI	TNIC NCOMP	Product Fluidity	10-K Competition
Avg Similarity Score	1												
Top 15 Similarity	0.781	1											
Top 10 Similarity	0.76	0.993	1										
Top 5 Similarity	0.725	0.949	0.975	1									
Sim. Precision	0.656	0.593	0.611	0.656	1								
Sum Similarity	0.331	0.718	0.693	0.628	0.267	1							
SIC3 HHI	-0.059	-0.171	-0.172	-0.161	-0.106	-0.189	1						
SIC3 Profit Margin	0.073	-0.001	0.009	0.024	-0.018	-0.244	0.226	1					
SIC3 NCOMP	-0.056	0.093	0.082	0.056	0.078	0.284	-0.38	-0.521	1				
TNIC HHI	-0.154	-0.435	-0.461	-0.466	-0.4	-0.385	0.142	0.025	-0.092	1			
TNIC NCOMP	0.161	0.606	0.586	0.525	0.245	0.897	-0.227	-0.315	0.419	-0.443	1		
Product Fluidity	0.201	0.47	0.468	0.437	0.206	0.515	-0.222	-0.186	0.246	-0.308	0.518	1	
10-K Competition	-0.049	0.039	0.036	0.027	0.017	0.096	-0.126	-0.156	0.202	-0.059	0.197	0.118	1

Table A.3: Product Market Differentiation and Corporate Fraud - Bivariate Probit

This table reports coefficient estimates from the partially observable bivariate probit model, $P(Z=1) = P(F=1)P(D=1|F=1)$, used in Wang and Winton (2014) . In specifications 2-4, we replace Average Similarity Score with the average from the firms most similar 5, 10, and 15 peers respectively.

	(1)		(2)		(3)		(4)	
	P(F)	$P(D\|F)$	P(F)	$P(D\|F)$	P(F)	$P(D\|F)$	P(F)	$P(D\|F)$
Avg Similarity Score	-7.378*	4.221						
	(-1.938)	(0.497)						
Avg Top 15 Sim.			-3.209**	4.603				
			(-2.385)	(1.582)				
Avg Top 10 Sim.					-2.970**	4.111		
					(-2.444)	(1.580)		
Avg Top 5 Sim.							-2.856***	3.812*
							(-2.695)	(1.665)
SIC3 NCOMP	-0.070**	0.303***	-0.076**	0.300***	-0.076**	0.304***	-0.079**	0.312***
	(-1.987)	(3.875)	(-2.088)	(3.393)	(-2.095)	(3.570)	(-2.148)	(3.870)
MA	0.367**	-0.514	0.371**	-0.518*	0.371**	-0.515*	0.371**	-0.505
	(2.055)	(-1.619)	(2.032)	(-1.672)	(2.034)	(-1.650)	(2.034)	(-1.592)
Stock Ind Return R2	0.642**	-2.082***	0.706**	-2.165***	0.707**	-2.176***	0.716**	-2.208***
	(1.974)	(-3.444)	(2.177)	(-3.598)	(2.173)	(-3.699)	(2.188)	(-3.881)
R&D	-0.123	-0.338	0.144	-0.931	0.125	-0.887	0.138	-0.924
	(-0.209)	(-0.199)	(0.215)	(-0.502)	(0.188)	(-0.482)	(0.210)	(-0.503)
R&D dummy	0.257***	-0.664***	0.315***	-0.710***	0.317***	-0.718***	0.324***	-0.740***
	(3.052)	(-3.879)	(3.509)	(-3.872)	(3.512)	(-3.962)	(3.554)	(-4.125)
Capx	0.212	-0.606	0.265	-0.724	0.256	-0.706	0.260	-0.716
	(0.468)	(-0.653)	(0.580)	(-0.781)	(0.560)	(-0.763)	(0.570)	(-0.774)
Ln number analysts	-0.034	0.113	-0.023	0.089	-0.024	0.092	-0.027	0.097
	(-0.744)	(1.347)	(-0.504)	(1.027)	(-0.523)	(1.066)	(-0.578)	(1.137)
Inst Ownership	0.522***	-0.521	0.517***	-0.551*	0.516***	-0.541	0.520***	-0.532
	(3.438)	(-1.473)	(3.374)	(-1.691)	(3.367)	(-1.639)	(3.396)	(-1.555)
Ln Asset	0.228***	-0.239***	0.229***	-0.249***	0.230***	-0.247***	0.233***	-0.244***
	(7.600)	(-2.705)	(7.514)	(-3.360)	(7.536)	(-3.222)	(7.588)	(-2.899)
Ln Age	-0.563***	0.834***	-0.576***	0.863***	-0.577***	0.858***	-0.576***	0.841***
	(-8.257)	(4.597)	(-8.317)	(5.667)	(-8.326)	(5.402)	(-8.355)	(4.802)
Abnormal ROA		-1.079		-1.001		-1.033		-1.081
		(-1.629)		(-1.444)		(-1.521)		(-1.614)
Abnormal AAER		0.130***		0.116***		0.118***		0.121***
		(3.373)		(2.694)		(2.896)		(3.151)
Abnormal Volatility		-1.015		-0.816		-0.835		-0.862
		(-1.333)		(-1.146)		(-1.167)		(-1.187)
Abnormal Turnover		2.359		2.233		2.300		2.419
		(1.062)		(1.048)		(1.070)		(1.102)
Disastrous Return		0.489**		0.433**		0.440**		0.449**
		(2.358)		(2.101)		(2.186)		(2.276)
Dummy Security Issue	-0.249	0.683*	-0.229	0.621	-0.227	0.619	-0.215	0.596
	(-1.310)	(1.839)	(-1.176)	(1.644)	(-1.166)	(1.641)	(-1.104)	(1.571)
EFN	-0.111		-0.112		-0.113		-0.116	
	(-0.489)		(-0.524)		(-0.525)		(-0.528)	
ROA	0.064		0.061		0.068		0.077	
	(0.285)		(0.280)		(0.309)		(0.345)	
Book Leverage	-0.087		-0.103		-0.104		-0.106	
	(-1.140)		(-1.391)		(-1.393)		(-1.389)	
Pct Soft Assets	0.548***		0.537***		0.544***		0.554***	
	(3.577)		(3.433)		(3.532)		(3.557)	
Constant	-1.683***	-0.737	-1.733***	-0.479	-1.734***	-0.551	-1.736***	-0.661
	(-4.896)	(-0.454)	(-4.996)	(-0.307)	(-5.015)	(-0.357)	(-5.040)	(-0.422)
Observations	30,117	30,117	30,117	30,117	30,117	30,117	30,117	30,117
FE	Year	Year	Year	Year	Year	Year	Year	Year

Table A.4: Product Differentiation and Corporate Fraud
(Alternate Constructions of Independent Variable)

This table reports OLS estimates for the incidence of fraud on alternative constructions of our primary independent variable. Our proxy for corporate fraud includes a combination of AAER misstatements from the AAER dataset and Securities Class Actions from the Stanford University Lawsuit Database. In Column 1, we report results for the main dependent variable used throughout our analysis. In Columns 2-4, Average Similarity Score is replaced with the (natural log of) number of a firms rivals in the top 75th, 90th and 95th percentile of similarity scores, respectively in the full cross section of firms in year t. Panel B reports the standardized regression. The unit of observation in this analysis is the firm-year. All specifications include the full set of controls as described in Section II, and they include year fixed effects, and explanatory variables lagged by one year. The t-statistics, calculated from standard errors clustered at three digit SIC code (SIC3) level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

	(1) Fraud	(2) Fraud	(3) Fraud	(4) Fraud	(5) Fraud	(6) Fraud	(7) Fraud	(8) Fraud
Panel A OLS Regression								
Avg Similarity Score	-0.203***				-0.171***			
	(-4.248)				(-3.946)			
Ln NCOMP TNIC 75th		-0.003*				-0.001		
		(-1.778)				(-0.561)		
Ln NCOMP TNIC 90th			-0.003**				-0.002	
			(-2.265)				(-0.988)	
Ln NCOMP TNIC 95th				-0.004**				-0.002
				(-2.314)				(-1.123)
Observations	37,144	37,144	37,144	37,144	37,144	37,144	37,144	37,144
R-squared	0.016	0.015	0.015	0.016	0.034	0.034	0.034	0.034
Panel B Standardized Regression								
Avg Similarity Score	-0.005***				-0.004***			
	(-4.248)				(-3.946)			
Ln NCOMP TNIC 75th		-0.004*				-0.002		
		(-1.778)				(-0.561)		
Ln NCOMP TNIC 90th			-0.004**				-0.002	
			(-2.265)				(-0.988)	
Ln NCOMP TNIC 95th				-0.004**				-0.002
				(-2.314)				(-1.123)
Observations	37,144	37,144	37,144	37,144	37,144	37,144	37,144	37,144
R-squared	0.016	0.015	0.015	0.016	0.034	0.034	0.034	0.034
FE	Year	Year	Year	Year	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3
Cluster	sic3	sic3	sic3	sic3	sic3	sic3	sic3	sic3

Table A.5: Product Market Differentiation and Financial Statement Comparability

This table reports estimates for the incidence of fraud on the average similarity of each firms rivals using ordinary least squares (OLS) regressions adding in the output-based measure of accounting comparability from De Franco et al. (2011). Our proxy for corporate fraud includes a combination of AAER misstatements from the AAER dataset and Securities Class Actions from the Stanford University Lawsuit Database. Panel A presents the mean Average Similarity Score and % Fraud for firm-years with below and above median Accounting Comparability and the correlation between the two measures. Panel B includes regressions using the full specification from Table 2, including Year#Sic3 fixed effect transformations. The specification in Column 1 contains observations where Accounting Comparability is not missing. The specification in Column 2 includes Accounting Comparability as a control. In Columns 3-4 we split the data between low (below median) and high (above median) accounting statement comparability. All specifications are run at the firm-year level, include Year#Sic3 fixed effects, and explanatory variables are lagged by one year. The t-statistics, calculated from standard errors clustered at SIC3 level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

Panel A			
Accounting Comparability	Low	High	Missing
Average Similarity Score	3.1	2.9	3.1
% Fraud	1.3	2.0	2.0
Observations	16,898	16,558	33,400
Corr(Avg Score:Comp)		-1.5%	

Panel B	(1)	(2)	(3) Low	(4) High
Accounting Comp		-.001		
		(-1.24)		
Avg Similarity Score	-0.147**	-0.150**	0.039	-0.265***
	(-2.32)	(-2.34)	(0.04)	(-4.01)
Observations	20,429	20,429	9,728	9,972
R-squared	0.093	0.093	0.178	0.121
FE	Year#Sic3	Year#Sic3	Year#Sic3	Year#Sic3
Controls	Full	Full	Full	Full

Table A.6: Rival Fraud, Product Differentiation, and Fraud Detection

This table reports estimates for the incidence of fraud on the average similarity of each firms rivals and rival firm fraud activity using ordinary least squares (OLS) regressions. Our proxy for corporate fraud includes a combination of AAER misstatements from the AAER dataset and Securities Class Actions from the Stanford University Lawsuit Database. The variable Rival Fraud is a dummy variable equal to 1 if a rival firm has been charged with committing fraud in year t. The variable Similarity Score is the similarity between the competitor pair in year t and Rival Fraud $\times$ Similarity Score is the interaction term. The unit of observation in these analyses is the competitor-pair-year. The specification in Column 1 does not include control variables. Column 2 includes the Dechow et al. (2011) set of controls, and Columns 3 includes our full set of controls defined in Section II. Column 4 includes the full set of controls but removes the industry level relative performance flag measure. All specifications include year fixed effects explanatory variables are lagged by one year. The t-statistics, calculated from standard errors clustered at the three-digit SIC code (SIC3) level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

	(1) Fraud	(2) Fraud	(3) Fraud	(4) Fraud
Rival Fraud	0.011***	0.007**	0.003	0.003
	(2.916)	(2.536)	(1.093)	(1.190)
Similarity Score	-0.054	-0.050**	-0.055***	-0.074***
	(-1.305)	(-2.442)	(-2.784)	(-3.654)
Rival Fraud Similarity Score	0.088**	0.078*	0.053	0.070*
	(1.990)	(1.846)	(1.439)	(1.864)
Observations	4,790,666	2,582,151	2,201,709	2,201,709
R-squared	0.010	0.017	0.020	0.019
Controls	None	Dechow	Full	Full (no RPE)
FE	Year	Year	Year	Year

Table A.7: IPOs and Acquisitions of Rivals as Change to Information Environment

This table reports OLS estimates for the association between fraud and rival IPOs (M&A) activity. The specifications are the same as model (4) of Table 2, but also include rival firm IPO (M&A) activity. For each firm-year, include the natural log of the number of firms that compete with firm i and that underwent an IPO or were acquired in year t, and an interaction term Ln Num Competitor IPO Avg Similarity Score or Ln Num Competitor Target Avg Similarity Score. In Column 2 and 5, we control for IPO (M&A) Size ($) which is the sum of all-capital raised by IPO rivals (total market capitalization of Target rivals). All specifications include year fixed effects and all control variables are lagged one year. Columns 3 and 6 also include three-digit SIC code (SIC3) fixed effects. The t-statistics, calculated from standard errors clustered at the SIC3 level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively..

	(1) Fraud	(2) Fraud	(3) Fraud	(4) Fraud	(5) Fraud	(6) Fraud
Avg Similarity Score	-0.149***	-0.149***	-0.134***	-0.185***	-0.185***	-0.162***
	(-3.488)	(-3.473)	(-3.382)	(-4.060)	(-4.061)	(-3.927)
Ln Num Competitor IPO	0.009***	0.009***	0.008***			
	-4.767	-4.909	-4.565			
Av Score× Ln Num Comp IP	-0.119***	-0.123***	-0.109***			
	(-3.370)	(-3.460)	(-2.660)			
IPO Size ($)		0.000				
		(-0.498)				
Ln Num Competitor Target				0.052***	0.061**	0.048***
				-3.373	-2.436	-3.38
Avg Score ×Ln Num Comp Target				-0.844***	-0.874**	-0.685**
				(-2.693)	(-2.593)	(-2.430)
Ln Target MarketCap					-0.001	
					(-0.588)	
Observations	37144	37144	37144	37144	37144	37144
R-squared	0.017	0.017	0.034	0.017	0.017	0.035
Controls	Full	Full	Full	Full	Full	Full
FE	Year	Year	Year+Sic3	Year	Year	Year+Sic3

Table A.8: Product Market Differentiation and Corporate Fraud - Firm FE

This table reports estimates for the incidence of fraud on the average similarity of each firms rivals using ordinary least squares (OLS) regressions. Our proxy for corporate fraud includes a combination of AAER misstatements from the AAER dataset and Securities Class Actions from the Stanford University Lawsuit Database. All specifications include the full set of controls as described in Section II. All specifications are run at the firm-year level, include year and firm fixed effects, and explanatory variables are lagged by one year. The t-statistics, calculated from standard errors clustered at SIC3 level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

	(1)	(2)	(3)
	Fraud	Fraud	Fraud
Avg Similarity Score	-0.060	-0.092*	-0.093*
	(-1.455)	(-1.713)	(-1.691)
R&D			-0.009
			(-0.906)
R&D Dummy			0.008
			(1.306)
Ln number analysts			0.009***
			(3.996)
Whited-Wu Index			0.002
			(0.554)
RSST accruals		0.001	0.002
		(0.493)	(0.537)
Change AR		-0.001	-0.003
		(-0.051)	(-0.275)
Change Inventory		0.001	0.004
		(0.036)	(0.219)
Pct Soft Assets		0.012**	0.013**
		(2.136)	(2.214)
Change in Cash Sales		0.003*	0.003*
		(1.783)	(1.776)
Change in ROA		-0.018***	-0.017***
		(-4.379)	(-3.775)
Change in employee		-0.000	-0.001
		(-0.158)	(-0.343)
Ln Age		-0.014*	-0.013*
		(-1.823)	(-1.859)
Dummy Security Issue		0.003	0.003
		(1.147)	(0.855)
Stock Ind Return R2			0.022*
			(1.801)
TNIC NCOMP			0.001
			(0.740)
Ln Asset		0.014***	0.009***
		(8.870)	(4.653)
Constant	0.022***	-0.029	-0.034*
	(17.625)	(-1.423)	(-1.683)
Observations	49,794	38,747	36,380
R-squared	0.417	0.434	0.437
FE	Year+Firm	Year+Firm	Year+Firm
Cluster	sic3	sic3	sic3

Table A.9: IPO Rival Similarity as a shock to product market differentiation

This table reports 2SLS estimates (second stage) for the relationship between product market similarity and corporate fraud. In columns 1-4, we report the first stage result for 2SLS regression. In columns 5-8, we use similarity scores with competitors undergoing an IPO (being acquired) as an instrument for the firms Average Similarity Score (Avg Top 15/10/5 Similarity). Column 1 and 5 only includes subsample with number of competitor IPO > 0. Our proxy for corporate fraud includes a combination of AAER misstatements from the AAER dataset and Securities Class Actions from the Stanford University Lawsuit Database. The t-statistics, calculated from standard errors clustered at the three-digit SIC code (SIC3) level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Avg Sim. Score	Avg Top 15 Sim.	Avg Top 10 Sim.	Avg Top 5 Sim.	Fraud			
IPO Avg Score	0.258***	0.397***	0.430***	0.476***				
	(8.859)	(8.501)	(10.065)	(12.796)				
Ln Num Comp IPO	-0.000	-0.001	-0.002***	-0.004***	0.003	0.003**	0.003*	0.002*
	(-1.133)	(-1.098)	(-2.820)	(-5.414)	(1.092)	(2.042)	(1.951)	(1.763)
IPO Size ($)	0.016***	0.088***	0.071***	0.045**	0.127***	0.105***	0.100***	0.094**
	(3.055)	(5.368)	(4.050)	(2.291)	(3.025)	(2.674)	(2.579)	(2.432)
$\widehat{AvgSim.Score}$					-0.314*			
					(-1.673)			
$\widehat{AvgTop15Sim.}$						-0.219*		
						(-1.683)		
$\widehat{AvgTop10Sim.}$							-0.203*	
							(-1.657)	
$\widehat{AvgTop5Sim.}$								-0.183
								(-1.619)
Constant	0.003	-0.021***	-0.026***	-0.029***	-0.021	0.009	0.008	0.008
	(1.446)	(-6.495)	(-7.902)	(-7.999)	(-1.079)	(0.666)	(0.614)	(0.589)
Observations	14,823	28,786	28,786	28,786	14,823	28,786	28,786	28,786
R-squared	0.679	0.682	0.681	0.630	0.046	0.040	0.040	0.040
FE	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3	Year+Sic3
Cluster	sic3	sic3	sic3	sic3	sic3	sic3	sic3	sic3

Table A.10: Product Differentiation and Fraud by Size Quartiles

This table reports estimates for the incidence of fraud on various competition measures using ordinary least squares (OLS) regressions. Our proxy for corporate fraud includes a combination of AAER misstatements, from the AAER dataset and Securities Class Actions from the Stanford University Lawsuit Database. We split the data into four size-based quartiles with 1 being the smallest and 4 the largest. All specifications are run at the firm-year level, include year fixed effects, and explanatory variables are lagged by one year. The t-statistics, calculated from standard errors clustered at SIC3 level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

	(1)	(2)	(3)	(4)
Size	Q1(Small)	Q2	Q3	Q4 (Large)
	Fraud	Fraud	Fraud	Fraud
Avg Similarity Score	-0.072**	-0.146*	-0.221***	-0.511***
	(-2.354)	(-1.850)	(-2.625)	(-3.564)
Num Competitor IPO	0.000**	0.001**	0.001***	0.001**
	(2.295)	(2.239)	(3.075)	(2.034)
Ln NCOMP TNIC	-0.000	0.001	-0.001	0.002
	(-0.042)	(0.497)	(-0.441)	(0.884)
Observations	9,949	10,150	9,711	7,332
R-squared	0.008	0.019	0.023	0.035
Controls	Full	Full	Full	Full
FE	Year	Year	Year	Year

Table A.11: Traditional Competition Measures and Restatements

This table reports estimates for accounting restatements using ordinary least squares (OLS) regressions. We include sales based HHIs (FF48 industry classifications) and a scaled version of HHI (scaled by number of competitors). Following Balakrishnan and Cohen, we include the G-index and E-index as governance controls. All specifications include year fixed effects. The t-statistics, calculated from standard errors clustered at SIC3 level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	pct AA	pct AA	pct AA	pct AA	pct fraud	pct fraud	pct fraud	pct fraud	pct fraud	pct fraud
FF HHI	0.094**		0.110**		-0.017		-0.022		-0.017*	
	(2.207)		(2.332)		(-1.351)		(-1.592)		(-1.953)	
N FF HHI		0.545		0.599		-0.088		-0.103		-0.092
		(1.533)		(1.452)		(-0.987)		(-0.969)		(-1.432)
GIndex	0.006*	0.005			0.001	0.001				
	(1.739)	(1.643)			(0.645)	(0.739)				
EIndex			0.013	0.012			0.001	0.001		
			(1.451)	(1.398)			(0.568)	(0.573)		
Book Leverage	-0.074	-0.081*	-0.107*	-0.111**	-0.029	-0.028	-0.034	-0.033	-0.020	-0.018
	(-1.525)	(-1.702)	(-1.938)	(-2.041)	(-1.354)	(-1.303)	(-1.473)	(-1.439)	(-1.388)	(-1.286)
Inst Ownership	-0.115	-0.137*	-0.105	-0.139	-0.015	-0.010	-0.013	-0.004	-0.021	-0.016
	(-1.625)	(-1.887)	(-1.276)	(-1.634)	(-0.465)	(-0.318)	(-0.367)	(-0.122)	(-1.124)	(-0.825)
Size	0.027***	0.026***	0.030***	0.030***	-0.002	-0.002	-0.001	-0.001	-0.001	-0.001
	(3.488)	(3.402)	(3.055)	(2.987)	(-0.363)	(-0.342)	(-0.183)	(-0.222)	(-0.278)	(-0.276)
ROA	-0.214***	-0.230**	-0.252***	-0.273***	-0.030	-0.028	-0.034	-0.030	-0.036	-0.035
	(-2.828)	(-2.619)	(-3.125)	(-2.987)	(-0.670)	(-0.588)	(-0.707)	(-0.595)	(-1.027)	(-0.989)
BtoM	0.035**	0.035**	0.034*	0.036*	-0.011	-0.011	-0.010	-0.010	-0.008	-0.008
	(2.570)	(2.578)	(1.982)	(1.962)	(-1.267)	(-1.291)	(-0.996)	(-1.050)	(-1.284)	(-1.340)
ZScore	-0.000	-0.000	-0.001	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000
	(-0.339)	(-0.121)	(-0.681)	(-0.275)	(-0.291)	(-0.393)	(-0.082)	(-0.312)	(0.782)	(0.683)
Equity Raised	-0.094	-0.136	-0.119	-0.173	-0.022	-0.014	-0.026	-0.013	-0.035	-0.031
	(-1.016)	(-1.363)	(-1.133)	(-1.541)	(-0.486)	(-0.305)	(-0.456)	(-0.226)	(-1.029)	(-0.912)
Constant	-0.133**	-0.102**	-0.108*	-0.085	0.041	0.036	0.047*	0.044	0.040**	0.036*
	(-2.625)	(-2.114)	(-1.917)	(-1.383)	(1.674)	(1.358)	(1.784)	(1.505)	(2.340)	(1.965)
Observations	484	484	392	392	484	484	392	392	748	748
R-squared	0.478	0.465	0.424	0.405	0.124	0.120	0.124	0.114	0.187	0.182
FE	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year

Table A.12: Arthur Andersen Auditor Tests

For our sample of firms, we identify if that company used Arthur Andersen as its auditor and whether such a firm was detected for fraud after it switched auditors (2002-2004 time period) based on conduct that occurred while it was an Arthur Anderson client (2000 to 2002 time period). High and low Average Similarity Score splits is based on their 1999 values.

	Obs	Detected Fraud	t-stat
Low Avg Score	39	0.231	
High Avg Score	43	0.558	
Diff		-0.327***	-3.126

Table A.13: Whistle Blowers

We use the data from Dyck et al. (2010) on the whistleblower types. The whistleblowers for Internal Crime come from people within the firm. The whistleblowers for External Crime are analysts, auditors, clients or competitors, equity holders, industry regulators, law firms, newspapers, the SEC, and the short-sellers. We compare the Average Similarity Score of year firm first has SCAC case between the two groups.

	Obs	Average Score	t-stat
External Crime	86	0.033	
Internal Crime	31	0.025	
Diff		-0.007**	2.047

Table A.14: Product Market Differentiation and Corporate Fraud - Non-Linear Specifications

This table reports estimates for the incidence of fraud on the average similarity of each firms rivals using non-linear regressions. Our proxy for corporate fraud includes a combination of AAER misstatements from the AAER dataset and Securities Class Actions from the Stanford University Lawsuit Database. Columns 1, 2, 4 and 5 are logit specifications and 3 and 6 are probit specifications. Columns 1-3 only include size as a control variable, while Columns 4-6 include the full set of controls as described in Section II. In specifications 2 and 4, we standardize the variable of interest for the purposes of understanding economic magnitude. The unit of observation in this analysis is the firm-year. All specifications include year fixed effects, and control variables are lagged by one year. The t-statistics, calculated from standard errors clustered at SIC3 level, are reported in parentheses below coefficient estimates. Statistical significance (two-sided) at the 10% 5%, and 1% level is denoted by *, **, and ***, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	Fraud	Fraud	Fraud	Fraud	Fraud	Fraud
Avg Similarity Score	-9.709*			-21.955***		-8.782***
	(-1.926)			(-3.847)		(-3.901)
Standardized Avg Simi. Score		-0.236*	-0.098**		-0.533***	
		(-1.926)	(-2.124)		(-3.847)	
Ln Asset	0.269***			0.287***	0.287***	0.118***
	(11.378)			(3.887)	(3.887)	(3.602)
Standardized Ln Asset		0.534***	0.223***			
		(11.378)	(9.400)			
Constant	-5.627***	-4.409***	-2.246***	-6.458***	-7.129***	-3.041***
	(-24.338)	(-26.371)	(-34.920)	(-9.594)	(-11.118)	(-11.273)
Observations	54,852	54,852	54,852	37,144	37,144	37,144
Specification	Logit	Logit	Probit	Logit	Logit	Probit
Controls	None	None	None	Full	Full	Full
FE	Year	Year	Year	Year	Year	Year