Product Differentiation, Benchmarking, and Corporate Fraud

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

July 16, 2018

Abstract:

We find that firms with lower product market differentiation exhibit significantly lower rates of fraud. This relationship is more pronounced for complex firms and is robust to controlling for various measures of competition, predictors of fraud, and industry heterogeneity. To help establish causality, we show this relationship holds when we exploit changes in product differentiation stemming from rivals' IPO and acquisition activity. Finally, we find that IPOs (and acquisitions) of rivals facilitate the detection of fraud for firms with *ex ante* greater product differentiation. Overall, our findings suggest that lower differentiation disciplines firms by facilitating fraud detection through a benchmarking channel.

We thank Donald Bowen, Umit Gurun, Gerard Hoberg, Zack Liu, Tanakorn Makaew, Gonzalo Maturana, Joerg Picard, and Vesa Pursiainen, participants at the 2018 FMA Asia Annual Conference as well as seminar participants at Clemson University, Southern Methodist University, Universidad de los Andes, University of Nevada Las Vegas, and the U.S. Securities and Exchange Commission for helpful comments.

[†]Texas Christian University, [‡]Michigan State University, [§]U.S. Securities and Exchange Commission. The Securities and Exchange Commission 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

Trust is a central factor in the proper functioning of financial markets (Greenspan, 2008). In particular, distrust in the accuracy of financial statements stemming from corporate fraud can erode firm value (Karpoff et al., 2008; Dyck et al., 2010), impose negative externalities on other firms (Guiso et al., 2008; Kedia and Philippon, 2009), and influence investors' allocation decisions (Giannetti and Wang, 2016; Gurun et al., 2017). These consequences have obliged auditors, regulators, and researchers to improve their understanding of managers' incentives to commit fraud and the ability of various parties to detect financial reporting manipulations.

A survey of CFOs by Dichev et al. (2013) indicates that comparability between rival firms is an important means for identifying financial reporting abnormalities. Building on their insight, we posit that greater product market overlap can enrich financial statement comparability, thus facilitating monitoring and improving fraud detection. This benchmarking effect could discipline managers' reporting practices, leading them to commit less fraud. Alternatively, intense product market competition stemming from less differentiated products can also pressure managers to distort perceived relative performance through reporting manipulations (see Shleifer, 2004; Tirole, 2010). Until recently, only coarse industry-level measures of competition have been available to researchers, which creates challenges in identifying the potential effect of these opposing forces. As a result, the relationship between corporate fraud and product market interactions remains underexplored.

In this paper, we move beyond traditional measures of competition to shed light on whether product market differentiation imposes discipline or increases pressure, on average, as it pertains to corporate fraud. To measure corporate fraud, we use the combination of settled Accounting and Auditing Enforcement Releases (AAER) and settled Securities Class Action Clearinghouse (SCAC) financial lawsuits.[1] To measure product differentiation, we use pairwise product market similarity scores developed by Hoberg and Phillips (2010, 2016) that are based on the product descriptions in firms' annual financial reports. Whereas most industry classifications are binary, product similarity scores capture the degree of product market overlap between firms. Furthermore, these similarity scores improve on the accuracy of other competitor classification schemes and allow for firm-specific and time-varying competitor networks. These features enable us to move beyond identifying associations between industry-level characteristics and fraud by conducting more powerful tests using firm-level measures of product market differentiation.

Our analyses reveal a strong relationship between product market differentiation and corporate fraud. We find that the incidence of fraud is significantly lower for firms with a less differentiated product mix. Specifically, a one standard deviation higher average product similarity score for a firm is associated with a 14.8-23.7% decrease in the rate of SEC enforcement actions, relative to the unconditional sample average. This result is robust to the inclusion of control variables that have a documented relation to corporate fraud, such as measures of firm

[1] 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.

size, accounting quality, internal and external corporate governance, the number of distinct markets in which a firm operates, industry and industry-year fixed effects, and various levels of clustered standard errors. Our findings suggest that the effect of product market differentiation on the incidence of fraud has a relatively large economic effect when compared to virtually any predictor of corporate fraud previously explored in the literature.[2]

It is well documented that there are pervasive differences in the rate of observed fraud across industries (Povel et al., 2007; Wang et al., 2010). While our initial regression specifications include industry and industry-year fixed effects, we cannot perfectly control for time-varying industry characteristics that influence fraudulent activity. However, our firm-level measures of competition allow us to exploit within-industry heterogeneity that is plausibly outside of a firm's control. To this end, we exploit initial public offerings (IPOs) and mergers and acquisitions (M&As) of a firm's rivals as potential shocks to the firm's product differentiation. While selection into industries is endogenous, the average firm likely has little control over the timing of a competitor's IPO or M&A decisions. In separate analyses, we use a firm's similarity score with competitors undergoing an IPO or being acquired as an instrumental variable for the firm's total product differentiation score. Using two-stage least squares regressions, we find strong corroborating evidence that the incidence of corporate fraud is significantly lower for firms with a less differentiated product mix. Importantly, these findings persist with the inclusion of industry fixed effects. While we cannot entirely rule out the potential for omitted variables to jointly determine a firm's fraudulent reporting and rival IPO and M&A activity, these findings suggest a causal link between product differentiation and corporate fraud.

One concern with empirical studies on 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 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).[3] Indeed, it is difficult to ascertain a plausible explanation for why the presence

[2] We run a regression with a large set of standardized predictors of fraud found in the literature and find that firm size is the only variable with a stronger economic effect on fraud. We report results in the Internet Appendix.

[3] 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.

of similar rivals would decrease outsiders' ability to detect reporting manipulations. Thus, our findings suggest that managers rationally respond to enhanced detection rates by committing less fraud.

To explore whether the apparent disciplining effect of low product differentiation stems from a benchmarking channel, 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 firm's 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.

To further explore the benchmarking channel, we exploit an additional feature of IPO and M&A activity. Specifically, we utilize IPOs and M&As of a firm's rivals as a shock to the firm's 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., 2013). 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 (Song and Walking, 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. In particular, IPOs or M&As by rivals increase the saliency of publicly available information in a firm's industry before a firm has time to fully wind down misreporting. A potential concern is that IPOs and M&A activity likely impact the competitive nature of the industry by injecting capital into a newly public firm or by consolidating market power in existing competitors. While we cannot entirely rule out the competition effect, we condition on the amount of capital raised in an IPO (size of M&A deal) to help isolate the effect of information saliency on fraud detection.

Product market differentiation likely reflects aspects of the competitive landscape, in addition to the amount of comparable information provided by rivals. To isolate the benchmarking aspect of product market similarity, we control for a variety of measures meant to capture various aspects of competitive intensity and market power. These measures include industry-wide profit margin, sales concentration, top-4 firm market share, the number of competitors a firm has, and product market fluidity. We find that including these measures of

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

competition as control variables does not significantly influence the coefficient estimates for product market differentiation. Furthermore, we fail to find a robust statistical relationship between these alternative competition measures and fraud when examined separately. These results suggest that product market differentiation plays an important and unique role in determining fraudulent behavior that is not influenced by controlling for other characteristics of competition.

To further mitigate concerns that confounding variables largely drive our results, we conduct falsification tests using restatements, rather than misstatements and class action lawsuits, as our left-hand side variable. We should not expect product differentiation to have much power in predicting restatements if our findings are truly due to a disciplining channel, because recent work suggests that most restatements are not intentional, material, or damaging to investors (e.g., McGuire et al., 2012; Karpoff et al., 2017; Donelson et al., 2017; and Griffin et al., 2018). We explore several model specifications and, in general, we do not observe a discernable pattern between financial restatements and product market similarity. Overall, this test increases our confidence that we have identified an economically meaningful link between fraud and product market similarity in our primary analysis.

We also attempt to allay concerns that our results are driven by the particular model specifications that we have chosen. Our results remain qualitatively similar when we estimate non-linear (i.e., probit or conditional logit) model specifications. 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 firm's top 5, 10 or 15 closest rivals, rather than averaging across the firm's 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 (75^{th}, 90^{th} and 95^{th}). Additionally, we have explored several variations in our set of control variables and the time period used in our estimation. Thus, our evidence appears to be quite robust to a variety of model specifications. Lastly, the results are robust to a) variations in the level of winsorization, b) to removing outliers, rather than winsorizing, and c) eliminating any winsorization of the data.

Finally, while the direction of our findings is consistent with a disciplining effect (i.e. less commission), we seek to further address the confounding effects of detection versus commitment by exploiting the granularity of our data at the competitor-pair level. To this end, we study a set of firms where the latent incentives to commit fraud outweigh the deterrence effect that our findings suggest holds on average. This test allows us to examine whether the detection of fraud in one firm facilitates the detection for rivals with a similar product mix. Consistent with waves of industry fraud (Povel et al., 2007; Wang et al., 2010) or contagion (Dimmock et al., 2017), we find that a firm is more likely to be charged if a rival was recently accused of financial misconduct. Importantly for the benchmarking channel, this effect is decreasing with the degree of differentiation from the convicted rival.

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 play an important role in both the incentive to commit fraud and the ability of external parties to uncover fraudulent activities.

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 (Andes 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 (Shliefer, 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. 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 such as Hsu et al. (2017) indicates that a firm's competitive landscape is an important determinant of analyst coverage and forecast accuracy. 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 and Winton (2014) 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 and Winton (2014). 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; Faber, 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.

II. 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, 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.[5]

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 literatures related to corporate fraud. 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 ($\Delta Inventory$), the percentage of soft assets (% *Soft Assets*), change in cash sales ($\Delta Cash\ Sales$), change in ROA (ΔROA), change in employees ($\Delta Employees$), 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 firm's 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 firm's short-term financial performance by immediately lowering expenses. Finally, we include 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 (Rangon, 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

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

Institutional Ownership, the natural log of the number of analysts covering a firm's stock (*Ln Num Analysts*), research and development expenses (*R&D*), and industry stock return r-squared (*Ind R2).* [6] To construct the industry r-squared, we follow Wang and Winton (2014) and first regress each firm's 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-Wu Index* (2006) for financial constraints.[7] 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 may have greater probabilities of financial distress, which has been shown to predict fraud (e.g. Healy and Wahlen, 1999). Alternatively, debt 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 (e.g., 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:

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

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 *i*'s 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.

[6] To handle observations with missing R&D, we follow the method outlined in Koh et al. (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).

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

III. Product Differentiation and Alternative Measures of Competition

A. Product Market Differentiation

For our measure of product differentiation, we use the product similarity score developed by Hoberg and Phillips (2010, 2016), who use textual analytics to capture the relatedness of a firm's product market with all other firms in that file annual reports with the U.S. Securities and Exchange Commission (SEC). The process involves vectorizing the product market vocabulary from the business description from each firm's annual 10-Ks, 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 contend that product market overlap measures relatedness that can be used to assess both financial statements and understand the common factors that affect the performance of related firms.

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. Thus, for any two firms i and j that exceed a given threshold, c, we have a real number in the interval [c,1] describing the similarity in the two firms vectorized product market description. 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. For example, if two firms were taken at random from the CRSP/Compustat universe, the likelihood of them being in the same three-digit SIC code is 2.05%. The TNIC-3 cutoff is specified such that the likelihood of two randomly drawn firms being deemed related according to TNIC-3 is also 2.05%. Hence, TNIC-3 is constructed to be "as coarse" as three-digit SIC codes.

Importantly, TNIC allows the flexibility for each firm to have its own distinct set of competitors. For example, Nike competes with Callaway in golf, and competes with Head in tennis, but Callaway and Head are not direct competitors with each other. 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. In particular, we can isolate firm specific features of competition 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, these 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 Exon 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. These features allow us to conduct more powerful tests than transitive and time-invariant industry classifications would allow.

More generally, standard measures of competition historically rely on static and transitive industry classifications (e.g., SIC and NAICS codes). These classification schemes presume that all firms have an equal relationship within industries and are unrelated across industries. 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. Therefore, because most product markets are not perfectly homogenous and are complex and dynamic, we deploy less restrictive industry classifications to properly assess the effect of competition on fraud. Indeed, Hoberg and Phillips (2016) find that these measures explain differences in financial metrics such as sales growth and profitability across industries better than traditional industry measures.

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 firm's 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).

There are potential issues when using the *Average Similarity Score* based on the TNIC classification. First, the TNIC only includes pairs of 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 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 have stronger discipline effect through benchmarking.

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 benchmarking 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.

In addition to the simple *Average Similarity Score* and rival counts, we develop a more nuanced measure that helps account for 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 firm's 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:

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

where N_i is the number of competitors in firm i's TNIC, and $score_{i,j,t}$ is the product similarity score between firm i and competitor j in year t.[8] Higher *precision* is indicative of a greater signal provided by a firm's product market rivals.

Table 2 reports the correlations for our main measure, Average Similarity Score, and the alternative measures using the similarity scores noted above. All three of these measures are highly correlated (over 95%) with each other and with the main measure that averages across all competitors (around 75%), which mitigates concerns regarding the distribution of similarity scores.

B. *Alternative Measures of Competition*

The similarity measures based on product market differentiation should capture the extent to which a firm's rivals provide suitable benchmarking of performance, thus facilitating the detection of fraud. However, the degree of product market differentiation also reflects the notion that competition is an endogenous outcome of market forces and that firms choose to differentiate to the greatest degree possible (e.g., Tirole, 1988). Therefore, less differentiation suggests more intense competition, all else equal. To isolate the benchmarking effect, we control for commonly used measures in the literature that are designed to capture other aspects of competition, including 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:

$$HHI_j = \sum_{i=1}^{N_j}(MS_i)^2$$

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

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 homogenous 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. According to this approach, each firm has a unique TNIC-based HHI, which differs from traditional SIC-based HHI measures, which are aggregated to some grouping of SIC codes (e.g. two- or three-digit). Furthermore, when constructing each firm's TNIC-HHI we can weight the sales of each rival by the firm's product market similarity with that rival. Thus, a firm's more similar rivals receive greater weight in its TNIC-HHI measure.

As additional measures of competition, we include the number of competitors according to a firm's TNIC or three-digit SIC codes. Classic models of competition suggest that the more firms there are offering similar products, the competition would be more intense (Tirole, 1988). We also include profit margin (Bain, 1951) and the sales concentration ratio of the top four firms in an industry (Heflebower, 1957) as an additional measure of market power at the industry level. Lastly, we include product market fluidity, which captures competitive pressure from potential entrants that captures each firm's ex ante competitive threats (Hoberg, et al., 2014). This measure also uses textual analytics and compares the use of unique words in each firm's product descriptions to changes in the overall use of a given word by other firms in their product descriptions. This measure lies between zero and one and is higher the more the words used a firm's product description overlap with the changes in the word changes by competitor firms. The intent is to capture threats based on the actions by rival firms, rather than changes of the firm itself.

Table 2 also contains the correlation of these traditional measures of competition with the product market similarity measures. It reveals that the traditional measures of competition are not strongly correlated with product similarity scores. Thus, product market similarity appears to capture different aspects of competition not explained by traditional measures. We explore these issues further in Section IV.

IV. Empirical Results

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 then move on to discuss empirical tests that are aimed to highlight the particular channel that could explain the findings of our exploratory regressions.

A. Product Market Similarity

We report OLS estimates for the association between average product similarity score (*Average Similarity Score)* and corporate fraud in Table 3.[9] 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 Dechow set of controls (i.e. *accruals*, change in accounts receivable (*ΔAR)*, change in Inventory ($\Delta Inventory$), the percentage of soft assets (% *Soft Assets*), change in cash sales ($\Delta Cash\ Sales$), change in ROA (ΔROA), change in employees ($\Delta Employees$), and a dummy for security issuance (*d_Security Issue*)).

In Column 3, we also include the natural log of total assets (*Ln Assets*), *Book Leverage*, *Industry Stock Return R-squared*, the *Whited-Wu Index*, a flag for relative performance evaluation (*RPE flag*), *R&D*, an R&D flag, and $Institutional\ \ Ownership$. 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 and use the specification from Column 4 as 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 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.[10]

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

[9] In untabulated analysis, we estimate this relationship with probit and logit specifications and find similar results.
[10] Our results are robust to clustering at broader industry classifications (e.g., two-digit SIC) and at the firm level.

to have a larger economic relation to fraud than *Average Similarity Score*.[11] 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.[12]

B. *Rival IPOs and M&As and Product Differentiation*

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). For instance, CEOs that are more likely to commit fraud might select into certain industries based on characteristics that are potentially associated with product market differentiation. While our inclusion of industry and industry-year fixed effects should at least partially mitigate this concern, we cannot perfectly control for differences in industry characteristics that are related to fraud. However, according to the TNIC developed by Hoberg and Phillips (2016), each firm has a unique set of competitors, as well as varying degrees of overlap with each competitor. This granularity allows us to exploit the effect of within-industry changes in product similarity scores.

To this end, we exploit IPOs and M&As of a firm's rivals as a shock to firm-level product differentiation that is plausibly outside of a firm's control. In particular, we implement two-stage least squares regressions using the product similarity scores between firm *i* and its rivals undergoing an IPO or being acquired as an instrument for firm *i*'s total product market differentiation.

We collect our 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 *i*'s Average Similarity Score with all rival firms undergoing an IPO in year *t*. Rivals that do not undergo an IPO in year *t* are excluded from this calculation. Next, we create a variable which counts the 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 define a dummy variable (Competitor IPO) equal to 1 if any of a firm's 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, which is consistent with the documented evidence that IPOs occur in waves (e.g. Lowry and Schwert, 2002).

We collect our sample of acquisitions from Thomson Reuters SDC Platinum database of U.S. publicly-traded targets from 1996-2012. Consistent with prior work, we examine deals where at least 50% of the target was

---

[11] We report results in the Internet Appendix for a specification with all variables standardized for expositional convenience and ease of comparison.
[12] We also control for financial statement comparability developed by De Franco et al. (2011) and we obtain similar results.

being bought and focus on completed deals that likely lead to more attention than withdrawn deals. For each pairwise observation of competitors, *i* and *j*, we flag whether firm *j* was acquired in year *t*. We then aggregate this data to the firm-year level to create our instrumental variable, *Rival-Acquisition Similarity*, which is the average similarity of firm *i* with each of its rivals that underwent an acquisition in year *t*. Rivals that were not acquired do not enter into this calculation. Each firm has an average of 0.06 rival firm acquisitions each year. Conditional on at least one competitor being acquired, the average increases to 1.09 competitor acquisitions per firm-year.

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 ex ante propensity to commit fraud. The first criterion requires that our instrumental variable should exhibit a strong relation to a firm's product differentiation. Competitor IPOs provides capital injections for a firm's existing (previously private) competitors, thus potentially altering the intensity of competition with those rivals. Similarly, acquisition activity involving rivals can consolidate market power and change within-industry relationships. In turn, both IPOs and M&A activity by rivals can influence a firm's total degree of product market differentiation. In our first stage results, reported in Table A2 of the Internet Appendix, we find a strong positive relationship between both Rival-IPO and *Rival-Acquisition Similarity* and *Average Similarity Score*. The positive sign indicates that when rivals undergoing an IPO are more similar with firm *i*, this will increase firm *i*'s 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).

The second criterion, often referred to as the exclusion restriction, requires that our instrumental variable be uncorrelated with fraudulent behavior, other than through its effect on product differentiation. While the joint decision of selection into industries and fraudulent behavior is the endogenous effect that we are trying to circumvent, the average firm likely has little control over the timing of a competitor's IPO or M&A decisions. One potential argument against our exclusion restriction is that acquisitions and IPOs both occur in waves that correspond to industry evolutions, which could also be associated with pressure to commit fraud. We control for industry fixed effects, as well as the number of rival IPOs (acquisitions) to mitigate this concern. Furthermore, the direction of this effect would work against the discipline effect that we observe. Finally, the variation in our setting comes from the degree of similarity with the rival undergoing an IPO (acquisition) rather than the presence of the IPO itself. Thus, the exclusion restriction in our setting assumes that firms have no control over the similarity of rivals undergoing an IPO (being acquired) and that similarity with those rivals, in conjunction with the timing of the IPO (acquisition), is unrelated to the firm's ex ante incentives to commit fraud.

In Table 4, we report the second stage results of two-stage least squares regressions using *Rival-IPO Similarity* (*Rival-Acquisition similarity*) as an instrument for *Average Similarity Score* in Panel A (B). 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 Columns 3 and 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 and M&A activity. However, the estimates are in line with those of the top quartile in our complexity analysis presented in Table 5. While we cannot entirely rule out the potential for omitted variables to jointly determine a firm's fraudulent reporting and rival IPO and M&A activity, the IV results are suggestive of a causal relationship between product differentiation and corporate fraud.

C. Firm Complexity and Product Differentiation

Thus far, we have documented a strong mitigating effect of product market similarity (lack of differentiation) on corporate fraud. In this section, we aim to further establish that the disciplining effect of product market similarity on earnings manipulation acts through a benchmarking channel. Tirole (2010) claims that relative performance evaluation (benchmarking) plays an important role in corporate governance because the performance of rival firms is partly governed by common shocks to production cost and demand. Along these lines, evidence suggests that benchmarking informs boards regarding CEO ability (Murphy, 1986) as well as market-and industry-wide conditions when determining CEO pay (Oyer, 2004). Similarly, Hsu et al. (2017) find that analysts produce more accurate for forecasts for firms that face more intense competition. Additionally, researchers have argued that rivals provide information regarding firm performance, which facilitates monitoring by investors, auditors, and regulators (e.g. Hart, 1983; Dyck et al., 2010). Consistent with this view, Dichev et al. (2013) provide survey evidence that deviations from industry rivals serve as an important feature in detecting earnings manipulations. However, there is little empirical evidence on the relationship between competition, benchmarking, and empirical proxies for corporate fraud.

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 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 firm's 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 5. In Panel A, we report the average number of unique SIC codes and the number of competitors in each firm's TNIC. Each specification is estimated using our full set of control variables, described in Section II and in our analysis of Table 3. We estimate regressions separately for each complexity quartile without *Institutional Ownership* (Panel B) and with *Institutional Ownership* (Panel C).

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.

In Panel C, we include a control for *Institutional Ownership* because institutions can serve as monitors (Hartzell and Starks, 2003). We split this specification into a separate panel because the number of observations is reduced substantially. The coefficient estimates continue to exhibit a strong monotonic relationship across complexity quartiles, and therefore, we omit institutional ownership form other specifications to maintain the larger sample size. Again, the effect of product market similarity in the top quartile is almost four times as large as the effect in the bottom quartile.

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 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.

D. Rival IPOs and M&As and Information

To further investigate the benchmarking channel, we exploit an additional feature of IPO and M&A activity by a firm's rivals. In particular, both events plausibly lead to a shock to the firm's information environment. We first study the event of IPOs by a firm's 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 firm's own financial statements. Consistent with this view, Bauguess et al. (2013) 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 firm's 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 each pairwise observation of competitors, *i* and *j*, we 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 firm's 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 competitor's 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 firm's own investors, regulators, and auditors. Second, the IPO provides a capital injection for a firm's 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 firm's 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 Panel A of Table 6. More specifically, we estimate the interaction between the natural log of the number of rival firms undergoing IPOs and a firm's product market similarity (Average Similarity Score) prior to the rivals' IPOs. A firm's pre-existing product market similarity is indicative of the level of market discipline provided by competitors prior to the rival's 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 rival's IPO (i.e., the effect is more pronounced for *ex ante* undisciplined firms).[13] 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 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.

Panel B of Table 6 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.

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

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).

E. *Alternative Measures of Competition, Product Similarity, and Fraud*

In this section, we show that product differentiation captures a particular dimension of competition not explained by traditionally used measures. Traditional competition measures include the Herfindahl-Hirschman Index (HHI) developed in Hirschman, (1945) and Herfindahl, (1950), profit margin (Bain, 1951), the sales concentration ratio of the top four firms in an industry (Heflebower, 1957), and the number of competitors. We also use the newer measure, product market fluidity, which captures the intensity of product market changes for a given firm each year (see Hoberg et al., 2014). All measures of competition are discussed in detail in Section III.

In each Column of Table 7, we estimate a specification that includes our full set of control variables, described in Section II. In Column 1, we control for the sales based HHI according to a firm's primary three-digit SIC code. This measure of competition is among the most widely used in academic research. In Column 2, we also control for the number of competitors that each firm has according to its primary three-digit SIC code. Classic models of competition, in which more firms offering the same product results in more competition, motivate our inclusion of the number of competitors. Additionally, in Column 3 we include profit margin (Bain, 1951) and the sales concentration ratio of the top four firms in an industry (Heflebower, 1957) to account for market power at the industry level. The effect of *Average Similarity Score* on fraud remains consistent in both significance and magnitude across Columns 1-3 of Table 3.

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 firm's 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. Finally, in Column 6 we include product market fluidity and sum similarity from Hoberg et al. (2014). Again, the relationship between *Average Similarity Score* on fraud remains consistent in both significance and magnitude in Columns 4-6.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 8, we report estimates for the relationship between corporate fraud and traditional 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 3 (described in section II).[14] We estimate the specifications in Table 8 without the inclusion of industry fixed effects to provide the best chance at highlighting a statistical relationship. The number of rivals in a firm's 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 8. 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 in Table 8 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.

At a minimum, the results reported in Tables 7 and 8 suggest that product similarity captures a dimension of competition unrelated to these alternative empirical measures. As these 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 and particularly important about the relationship between fraud and the degree of similarity with rivals. While we cannot perfectly rule out that product similarity is merely capturing competition more accurately than these alternative measures, these results add confidence to the benchmarking channel highlighted throughout our analyses.

F. Product Differentiation and Financial Restatements (Falsification Tests)

Recent work argues that misstatements and class action lawsuits are better measures of fraud than financial restatements, because most restatements are not intentional, material, or damaging to investors (e.g., Karpoff et al., 2017; Donelson et al., 2017). Indeed, in their studies on fraud, Griffin et al. (2018) and McGuire, et al. (2012) 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." This screening narrows the sample from 2,170 unique firms with restatements to 123 unique firms (see Griffin et al., 2018). Thus, we should not expect product differentiation to have much power in predicting restatements in the pre-screened sample, unless our main results are driven by omitted firm- or industry-level characteristics.

[14] 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; and Graham et al., 2005).

We perform falsification tests of our main analysis using restatements as the dependent variable from the full sample of pre-screened restatements and report the results in Table 10.[15] 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 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.[16] Overall, this analysis increases our confidence that we have identified an economically meaningful link between fraud and product market similarity in our primary analysis.

G. *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 competitors in a given year using an equally weighted average. As noted in Section III, we construct alternative measures of product market similarity to ensure our results are not overly sensitive to our particular measure of product market similarity.

In Table 10, we re-estimate our main specification with each of the alternative aggregation schemes discussed in Section III. Given the high correlation between these measures documented in Table 2, it is unsurprising that all variations yield a highly significant negative relationship with detected fraud. These results mitigate concerns regarding the obscuring effects of aggregation. Each measure has a different standard deviation, and so directly comparing the economic magnitude of each coefficient is not straightforward. To facilitate comparison the economic magnitudes, we estimate a specification in which we standardize all variables and report the results in Table A7 of the Internet Appendix. The coefficient estimates exhibit monotonicity based on the number of competitors used to compute each firm's 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 similar in magnitude and significance to the overall Average Similarity Score.

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

[16] 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).

H. Fraud Detection vs. Commission

A challenge for empirical studies on corporate fraud is that proxies for corporate fraud capture *detected* fraud and we do not observe all *committed* fraud (Dimmock and Gerken, 2012; Dyck et al., 2013; Wang and Winton, 2014). Being accused of committing fraud is a function of the probability of detection and a firm's actual incidence of fraud. Our estimates for the partial effect of product market similarity on *detected* fraud are negative and economically meaningful. These results suggest that either detection rates are lower for firms with less product differentiation or firms with less product differentiation commit less fraud. Our results thus far suggest that managers with high quality benchmarks *commit* less fraud.

To provide empirical evidence that product market similarity increases the ability to detect fraud, we exploit the granularity of the data at the competitor-pair level (pairwise observations). Thus, in this set up we use pairwise product similarity scores as our measure of product market similarity. We study a set of firms where the latent incentives to commit fraud outweigh the deterrence effect that our findings suggest holds on average.[17] That is, we examine incidence of fraud, conditional on a competitor getting caught. Further, we explore how the incidence is related to the similarity with the competitor that was caught. Specifically, we estimate the following equation:

$$Fraud_{i,j,t} = \alpha + Fraud_{j,t-1} + similarity_{i,j,t-1} + Fraud_{j,t-1} \times similarity_{i,j,t-1} + controls_{i,t-1} + \epsilon_{i,j,t}$$

where *i,j* represents a rival pair with *i* being the reference firm and *j* being the rival. That is, for each competitor-pair *i-j*, we examine whether incidence of fraud for firm *i* is a function of whether firm *j* commits fraud in year *t-1* as well as their similarity with firm *j* in year *t-1*.

We report estimates from this exercise in Table A3 of the Internet Appendix. The specification in Column 1 only includes year fixed-effects and we progressively add additional controls in subsequent specifications. The results suggest that product market similarity continues to have a large disciplining effect, which supports the main findings in the paper. The results also 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 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., 2017). Importantly, the coefficient estimate for the interaction term is positive for all specifications and statistically significant in Columns 2-4. These results suggest that the likelihood of getting caught after a rival has been accused of fraud is increasing in the product market similarity with that rival. We believe this evidence is consistent with product market similarity enhancing the ability to detect fraud. These

[17] For 96.17% of the competitor-pairs in our sample, neither firm has been detected for fraudulent behavior. In this setting, we are focused on the remaining 3.83% for which at least one firm has been charged.

results also increase our confidence that the negative association between product market similarity and observed fraud is due to less commission by managers. In turn, this evidence supports the notion that product market similarity has a strong disciplining effect on average.

As an additional step to mitigating concerns regarding *committed* and *detected* fraud, 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 A4 of the Appendix, we report coefficient estimates from the partially observable bivariate probit model, $P(Z=1) = P(F=1)P(D=1|F=1)$, which show that Average Similarity Score (as well as top 5, top 10, and top 15 Average Similarity Score) are strongly associated with a decline in commission as opposed to a decline in detection. These findings are consistent with managers rationally updating and committing less fraud.

V. Conclusion

Our paper empirically examines the relationship between product market differentiation and the incidence of corporate fraud. Having rivals with significant product market overlap can have two potential effects on firms' incentives to commit fraud. On one hand, less product market differentiation could facilitate the ability to evaluate common shocks faced by firms, enhancing monitoring by external parties such as regulators, auditors, and investors. In turn, enhanced monitoring should increase the likelihood that committed fraud would be detected. Whether this effect would result in more or fewer cases of fraud being observed depends on the extent to which managers respond to enhanced detection rates by committing less fraud. On the other hand, less differentiation could foster more competition, leading firms to commit more fraud to boost their own perceived relative performance.

We first establish that firms with less product market differentiation exhibit a significantly lower incidence of fraud. We corroborate these findings using an instrumental variables analysis of rival firm IPOs and M&A. This relationship holds even after controlling for traditional measures of competition, and further, these traditional measures do not have explanatory power in predicting fraud. We also find that having more similar rivals appears to have a stronger disciplining effect when the firm is more complex than when it is simple.

We also show that events that could affect the information environment of firms are associated with a greater detection effect for firms with lower pre-existing market discipline. We contend that these results are largely due to the ability to benchmark firm performance when there are more similar rivals with publicly available information. These results suggest that aspects of competition faced by a firm have a disciplining effect on the incentive to commit fraud.

Collectively, our paper provides new insight on how a particular aspect of competition, product market differentiation, influences the incentives to commit fraud via the ability to benchmark a firm against similar peers. Thus, our paper highlights the role of one market-based mechanism that can 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 fraudulent reporting.

References

Ades, A., & Di Tella, R. (1999). Rents, Competition, and Corruption. *The American Economic Review*, *89*(4), 982–993.

Ali, A., Klasa, S., & Yeung, E. (2009). The Limitations of Industry Concentration Measures Constructed with Compustat Data: Implications for Finance Research. The Review of Financial Studies, 22(10), 3839–3871.

Andergassen, R. (2016). Managerial Compensation, Product Market Competition and Fraud. International Review of Economics & Finance, 45, 1–15.

Bain, J. S. (1951). Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936-1940. The Quarterly Journal of Economics, 65(3), 293–324.

Balakrishnan, K., & Cohen, D. A. (2013). Competition and Financial Accounting Misreporting, Rochester, NY: Social Science Research Network.

Bauguess, S., Cooney, J., & Hanley, K. (2013). Investor demand for information in newly issued securities.

Beasley, M. S. (1996). An Empirical Analysis of the Relation between the Board of Director Composition and Financial Statement Fraud. The Accounting Review, 71(4), 443–465.

Bergstresser, D., & Philippon, T. (2006). CEO incentives and earnings management. Journal of Financial Economics, 80(3), 511–529.

Bertrand, J. (1883) "Book Review of Theorie Mathematique de la Richesse Sociale and of Recherches sur les Principles Mathematiques de la Theorie des Richesses", Journal de Savants 67: 499–508.

Brinn, T., Peel, M. J., & Roberts, R. (1994). Audit fee determinants of independent & subsidiary unquoted companies in the UK—an exploratory study. The British Accounting Review, 26(2), 101-121.

Buzby, S. L. (1975). Company Size, Listed Versus Unlisted Stocks, and the Extent of Financial Disclosure. Journal of Accounting Research, 13(1), 16–37.

Cheng, I.-H. (2011). Corporate Governance Spillovers. Rochester, NY: Social Science Research Network.

Chhaochharia, V., Grinstein, Y., Grullon, G., & Michaely, R. (2016). Product market competition and internal governance: Evidence from the Sarbanes–Oxley Act. Management Science, 63(5), 1405-1424.

Choi, S. J., Nelson, K. K., & Pritchard, A. C. (2009). The Screening Effect of the Private Securities Litigation Reform Act. Journal of Empirical Legal Studies, 6(1), 35–68.

Cohen, L., & Lou, D. (2012). Complicated firms. Journal of financial economics, 104(2), 383-400.

De Franco, G., Kothari, S. P., & Verdi, R. S. (2011). The benefits of financial statement comparability. Journal of Accounting Research, 49(4), 895-931.

Dechow, P. M., Ge, W., Larson, C. R., & Sloan, R. G. (2011). Predicting Material Accounting Misstatements*. Contemporary Accounting Research, 28(1), 17–82.

Demirgüç-Kunt, A., & Maksimovic, V. (1998). Law, finance, and firm growth. The Journal of Finance, 53(6), 2107-2137.

Dichev, I. D., Graham, J. R., Harvey, C. R., & Rajgopal, S. (2013). Earnings quality: Evidence from the field. Journal of Accounting and Economics, 56(2, Supplement 1), 1–33.

Dimmock, S. G. & Gerken, W. C., (2012). Predicting Fraud by Investment Managers, Journal of Financial Economics 105, 153-173

Dimmock, S. G., Gerken, W. C., & Graham, N. P. (2017). Is Fraud Contagious? Co-Worker Influence on Misconduct by Financial Advisor, forthcoming Journal of Finance.

Donelson, D. C., Kartapanis, A., McInnis, J., & Yust, C. G. (2017). Measuring Financial Reporting Fraud Using Public and Private Enforcement, working paper.

Dyck, A., Morse, A., & Zingales, L. (2010). Who Blows the Whistle on Corporate Fraud? The Journal of Finance, 65(6), 2213–2253.

Dyck, I. J. A., Morse, A., & Zingales, L. (2013). How Pervasive is Corporate Fraud? Rochester, NY: Social Science Research Network. Retrieved

Fee, C. E., & Thomas, S. (2004). Sources of gains in horizontal mergers: evidence from customer, supplier, and rival firms. Journal of Financial Economics, 74(3), 423-460.

Faccio, M.,& Zingales, L. (2017). Political Determinants of Competition in the Mobile Telecommunication Industry. National Bureau of Economic Research.

Farber, D. B. (2005). Restoring Trust after Fraud: Does Corporate Governance Matter? The Accounting Review, 80(2), 539–561.

Giannetti, M., & Wang, T. Y. (2016). Corporate Scandals and Household Stock Market Participation. The Journal of Finance, 71(6), 2591–2636.

Gillan, S. L., Hartzell, J. C., & Starks, L. T. (2011). Tradeoffs in corporate governance: Evidence from board structures and charter provisions. The Quarterly Journal of Finance, 1(04), 667-705.

Giroud, X., & Mueller, H. M. (2010). Does Corporate Governance Matter in Competitive Industries? Journal of Financial Economics, 95(3), 312–331.

Graham, J. R., Harvey, C. R., & Rajgopal, S. (2005). The economic implications of corporate financial reporting. Journal of Accounting and Economics, 40(1), 3–73.

Graham, J. R., Li, S., & Qiu, J. (2008). Corporate Misreporting and Bank Loan Contracting. Journal of Financial Economics, 89(1), 44–61.

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

Griffin, P. A., Grundfest, J. A., & Perino, M. A. (2004). Stock Price Response to News of Securities Fraud Litigation: An Analysis of Sequential and Conditional Information. Abacus 40(1),21-48.

Grossman, S. J., & Hart, O. D. (1982). Corporate financial structure and managerial incentives. In The economics of information and uncertainty (pp. 107-140). University of Chicago Press.

Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the Stock Market. The Journal of Finance, 63(6), 2557–2600.

Gurun, U. G., Stoffman, N., & Yonker, S. E. (2017). Trust busting: The effect of fraud on investor behavior. The Review of Financial Studies, 31(4), 1341-1376.

Hart, O. D. (1983). The Market Mechanism as an Incentive Scheme. The Bell Journal of Economics, 14(2), 366–382.

Hartzell, J. C., & Starks, L. T. (2003). Institutional Investors and Executive Compensation. The Journal of Finance, 58(6), 2351–2374.

Healy, P. M., & Wahlen, J. M. (1999). A Review of the Earnings Management Literature and Its Implications for Standard Setting. Accounting Horizons, 13(4), 365–383.

Heflebower, R. B. (1957). Barriers to New Competition. The American Economic Review, 47(3), 363–371.

Herfindahl, O.C. (1950). Concentration in the steel industry. Diss. Columbia University.

Hirschman, A.O. (1945), National Power and the Structure of Foreign Trade. University of California Press: Berkeley, CA. 155-62.

Hoberg, G., & Phillips, G. (2010). Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. The Review of Financial Studies, 23(10), 3773–3811.

Hoberg, G., & Phillips, G. (2016). Text-Based Network Industries and Endogenous Product Differentiation. Journal of Political Economy 124 (5), 1423-1465.

Hoberg, G., Phillips, G., & Prabhala, N. (2014). Product Market Threats, Payouts, and Financial Flexibility. The Journal of Finance, 69(1), 293–324.

Holmstrom, B. (1982). Moral Hazard in Teams. The Bell Journal of Economics, 13(2), 324–340.

Hsu, C., Li, X., Ma, Z., & Phillips, G. M. (2017). Does Product Market Competition Influence Analyst Coverage and Analyst Career Success?, working paper.

Jayaraman, S., & Milbourn, T. (2009). Does equity-based CEO compensation really increase litigation risk. Unpublished Working Paper, Washington University.

Karaoglu, E., Sandino, T., & Beatty, R. P. (2006). Benchmarking Against the Performance of High Profile 'Scandal' Firms, Columbia University working paper.

Karpoff, J. M., Lee, D. S., & Martin, G. S. (2008). The Cost to Firms of Cooking the Books. The Journal of Financial and Quantitative Analysis, 43(3), 581–611.

Karpoff, J. M., Koester, A., Lee, D. S., & Martin, G. S. (2017). Proxies and databases in financial misconduct research. The Accounting Review, 92(6), 129-163.

Kedia, S., & Philippon, T. (2009). The Economics of Fraudulent Accounting. The Review of Financial Studies, 22(6), 2169–2199.

Khanna, V., Kim, E. H., & Lu, Y. (2015). CEO Connectedness and Corporate Fraud. The Journal of Finance, 70(3), 1203–1252.

Koh, P. S., & Reeb, D. M. (2015). Missing r&d. Journal of Accounting and Economics, 60(1), 73-94.

Li, F., Lundholm, R., & Minnis, M. (2013). A Measure of Competition Based on 10-K Filings. Journal of Accounting Research, 51(2), 399–436.

Lowry, M., & Schwert, G. W. (2002). IPO market cycles: Bubbles or sequential learning?. The Journal of Finance, 57(3), 1171-1200.

McGuire, Sean T., Thomas C. Omer, and Nathan Y. Sharp, 2012, The impact of religion on financial reporting irregularities, The Accounting Review 87, 645.

Murphy, K. J. (1986). Incentives, learning, and compensation: A theoretical and empirical investigation of managerial labor contracts. The Rand Journal of Economics, 59-76.

Nalebuff, B. J., & Stiglitz, J. E. (1983). Prizes and Incentives: Towards a General Theory of Compensation and Competition. The Bell Journal of Economics, 14(1), 21–43.

Nickell, S. J. (1996). Competition and Corporate Performance. Journal of Political Economy, 104(4), 724–746.

Oyer, P. (2004). Why do firms use incentives that have no incentive effects?. The Journal of Finance, 59(4), 1619-1650.

Povel, P., Singh, R., & Winton, A. (2007). Booms, Busts, and Fraud. The Review of Financial Studies, 20(4), 1219–1254.

Reynolds, J. K., & Francis, J. R. (2000). Does size matter? The influence of large clients on office-level auditor reporting decisions. Journal of Accounting and Economics, 30(3), 375–400.

Richardson, S. A., Sloan, R. G., Soliman, M. T., & Tuna, I. (2005). Accrual reliability, earnings persistence and stock prices. Journal of accounting and economics, 39(3), 437-485.

Schmidt, K. M. (1997). Managerial Incentives and Product Market Competition. The Review of Economic Studies, 64(2), 191–213.

Shleifer, A. (2004). Does Competition Destroy Ethical Behavior? (Working Paper No. 10269). National Bureau of Economic Research.

Song, M. H., & Walkling, R. A. (2000). Abnormal returns to rivals of acquisition targets: A test of the acquisition probability hypothesis'. Journal of Financial Economics, 55(2), 143-171.

Teoh, S. H., Welch, I., & Wong, T. J. (1998). Earnings management and the long-run market performance of initial public offerings. The Journal of Finance, 53(6), 1935-1974.

Thompson, R. B., & Sale, H. A. (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, & Organization, 29(3), 535-568.

Wang, T. Y., & Winton, A. (2014). Product Market Interactions and Corporate Fraud. Rochester, NY: Social Science Research Network.

Wang, T. Y., Winton, A., & Yu, X. (2010). Corporate Fraud and Business Conditions: Evidence from IPOs. The Journal of Finance, 65(6), 2255–2292.

Whited, T. M., & Wu, G. (2006). Financial Constraints Risk. The Review of Financial Studies, 19(2), 531–559.

Appendix: Variable Definitions

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-year's TNIC
Precision	Defined as $\left(\frac{1}{NCOMP_{TNIC}} \times \sum \frac{1}{(1-score)^2}\right)^{\frac{1}{2}}$
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 (noncurrent 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 firm's 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 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.

VARIABLES	No. Obs	Mean	Std. Dev.	10^{th} Percentile	90^{th} 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: Correlations

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

	Avg Sim. Score	HHI	SIC3 Profit	NCOMP SIC3	TNIC HHI	NCOMP TNIC	Top5 Similarity	Top10 Similarity	Top15 Similarity	Avg Precision	Sum Similarity	Product Fluidity
Avg Similarity Score	1.000											
SIC3 HHI	-0.086	1.000										
SIC3 Profit Margin	-0.116	0.254	1.000									
NCOMP SIC3	0.010	-0.401	-0.540	1.000								
TNIC HHI	-0.173	0.163	0.092	-0.125	1.000							
NCOMP TNIC	0.293	-0.255	-0.528	0.471	-0.436	1.000						
Top5 Similarity	0.745	-0.182	-0.166	0.125	-0.467	0.586	1.000					
Top10 Similarity	0.778	-0.197	-0.214	0.161	-0.459	0.654	0.977	1.000				
Top15 Similarity	0.796	-0.198	-0.240	0.177	-0.435	0.679	0.953	0.994	1.000			
Avg Score Precision	0.671	-0.119	-0.097	0.099	-0.400	0.271	0.636	0.601	0.587	1.000		
Sum Similarity	0.447	-0.202	-0.533	0.337	-0.333	0.887	0.635	0.707	0.738	0.272	1.000	
Product Mkt Fluidity	0.279	-0.245	-0.357	0.316	-0.309	0.587	0.483	0.519	0.527	0.229	0.574	1.000

Table 3: Product Market Similarity and Corporate Fraud

This table reports OLS estimates for the incidence of fraud on the average similarity of each firm's 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. 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. 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
Avg. Similarity Score	-0.113**	-0.180***	-0.220***	-0.171***	-0.163***
	(-2.089)	(-3.517)	(-3.915)	(-3.946)	(-3.244)
R&D			0.009	-0.011	-0.012
			(0.740)	(-1.008)	(-1.143)
R&D dummy			-0.000	-0.003	-0.004
			(-0.011)	(-0.914)	(-1.135)
Ln number analysts			0.001	0.001	0.001
			(0.227)	(0.267)	(0.392)
Inst Ownership			0.007		
			(0.792)		
Whited-Wu Index			0.005	-0.000	0.033*
			(0.859)	(-0.057)	(1.892)
RSST accruals		0.002	-0.003	0.001	0.002
		(0.538)	(-0.636)	(0.342)	(0.572)
Change AR		0.022*	0.023	0.016	0.023*
		(1.791)	(1.460)	(1.296)	(1.808)
Change Inventory		0.016	0.026	0.020	0.027
		(0.756)	(1.258)	(0.937)	(1.225)
Pct. Soft Assets		0.019***	0.022***	0.019***	0.019***
		(4.241)	(4.186)	(3.825)	(3.733)
Change in Cash Sales		0.005**	0.004**	0.005**	0.005**
		(2.250)	(2.057)	(2.227)	(2.367)
Change in ROA		-0.023***	-0.017***	-0.021***	-0.018***
		(-6.132)	(-3.336)	(-5.887)	(-5.212)
Change in employee		-0.004**	-0.004*	-0.003	-0.003
		(-2.101)	(-1.868)	(-1.456)	(-1.414)
Ln Age		-0.010***	-0.009***	-0.008***	-0.007***
		(-3.413)	(-3.040)	(-3.513)	(-2.974)
Dummy Security Issue		0.003	-0.001	0.002	0.001
		(1.327)	(-0.258)	(0.737)	(0.275)
Stock Industry Return R2			-0.009	0.017	
			(-0.822)	(1.515)	
Relative Perf Eval Flag			0.007**		
			(2.124)		
Ln NCOMP TNIC			0.001	0.002	0.001
			(1.205)	(0.945)	(0.756)
Ln Asset		0.006***	0.005***	0.006***	0.007***
		(4.617)	(3.245)	(2.913)	(3.546)
Observations	50,526	39,519	28,912	38,916	37,144
R-squared	0.005	0.015	0.014	0.034	0.079
FE	Year	Year	Year	Year+SIC3	Year×SIC3

Table 4: Product Market Similarity, Rival IPOs and M&A, and Corporate Fraud

This table reports 2SLS estimates (second stage) for the relationship between product market similarity and corporate fraud. In columns 1-3 (4-6), we use similarity scores with competitors undergoing an IPO (being acquired) as an instrument for the firm's *Average Similarity Score*. 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.

	IV Rival IPOs			IV Rivals Acquired		
	(1)	(2)	(3)	(4)	(5)	(6)
	Fraud	Fraud	Fraud	Fraud	Fraud	Fraud
Avg Similarity Score	-0.618***	-0.530***	-0.495**	-0.563***	-0.670**	-0.623*
	(-4.498)	(-3.181)	(-2.365)	(-2.929)	(-2.431)	(-1.708)
Ln Asset	0.007***	0.006***	0.006***	0.006***	0.006***	0.007***
	(4.543)	(3.952)	(3.885)	(4.103)	(3.661)	(3.515)
Ln Age	-0.010***	-0.009***	-0.009***	-0.012***	-0.010***	-0.009***
	(-4.135)	(-4.025)	(-4.000)	(-3.860)	(-3.934)	(-3.826)
Ln Num Competitor IPO	0.006***	0.005**	0.004**			
	(3.106)	(2.477)	(2.094)			
IPO Size ($)	0.000	-0.000	-0.000			
	(0.037)	(-0.403)	(-0.727)			
Ln Num Competitor Acquired				0.030*	0.029*	0.025
				(1.768)	(1.772)	(1.642)
Ln Target Size				-0.001	-0.001	-0.000
				(-0.360)	(-0.503)	(-0.052)
Ln NCOMP_TNIC		0.002	0.002		0.003***	0.002*
		(1.093)	(0.837)		(2.901)	(1.831)
Observations	37,335	37,335	37,335	37,335	37,335	37,335
R-squared	0.010	0.014	0.033	0.010	0.011	0.031
FE	Year	Year	Year + SIC3	Year	Year	Year + SIC3
Controls	No	Full	Full	No	Full	Full

Table 5: Product Similarity and Fraud by Complexity Quartiles

This table reports OLS estimates for the incidence of fraud on the average similarity of each firm's 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 firm's 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, and Panel C includes Institutional Ownership. 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
		Panel C		
Avg Similarity Score	-0.191***	-0.169	-0.166	-0.677***
	(-3.217)	(-1.631)	(-1.029)	(-4.081)
Inst Ownership	0.010	0.004	0.009	0.011
	(0.769)	(0.225)	(0.729)	(1.012)
Observations	7,707	7,469	7,048	6,688
R-squared	0.015	0.021	0.023	0.023
FE	Year	Year	Year	Year
Controls	Full	Full	Full	Full

Table 6: Product Market Similarity, Rival IPOs and M&A, and Corporate Fraud

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 3, but also include rival firm IPO (M&A) activity in Panel A (Panel B). 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 of Panel A (B), we control for IPO (M&A) Size ($) which is the sum of all-capital raised by IPO rivals (total market capitalization of Target rivals). In Columns 4 and 5 of Panel A (Panel B), 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 all control variables are lagged one year. Columns 3-5 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.

Panel A: Rival IPO

	(1)	(2)	(3)	(4)	(5)
				Low Non-IPO Rival Score	High Non-IPO Rival Score
	Fraud	Fraud	Fraud	Fraud	Fraud
Avg Similarity Score	-0.149***	-0.149***	-0.134***		
	(-3.488)	(-3.473)	(-3.382)		
Ln Num Competitor IPO	0.009***	0.009***	0.008***	0.005***	0.004
	(4.767)	(4.909)	(4.565)	(2.82)	(1.44)
Avg Score×Ln Num Comp IPO	-0.119***	-0.123***	-0.109***		
	(-3.370)	(-3.460)	(-2.660)		
IPO Size ($)		-0.000		-0.000	-0.000
		(-0.498)		(-0.02)	(-0.72)
Observations	37,144	37,144	37,144	18,858	18,279
R-squared	0.017	0.017	0.034	0.050	0.037

Panel B: Rival M&A

	(1)	(2)	(3)	(4)	(5)
				Low Non-M&A Rival Score	High Non-M&A Rival Score
	Fraud	Fraud	Fraud	Fraud	Fraud
Avg Similarity Score	-0.185***	-0.185***	-0.162***		
	(-4.060)	(-4.061)	(-3.927)		
Ln Num Competitor Target	0.052***	0.061**	0.048***	0.103***	-0.000
	(3.373)	(2.436)	(3.380)	(3.141)	(-0.024)
Avg Score×Ln Num Comp Target	-0.844***	-0.874**	-0.685**		
	(-2.693)	(-2.593)	(-2.430)		
Ln Target MarketCap		-0.001		-0.008***	0.002
		(-0.588)		(-2.829)	(1.380)
Observations	37,144	37,144	37,144	18672	18449
R-squared	0.017	0.017	0.035	0.052	0.039
Controls	Full	Full	Full	Full	Full
SIC3 FE	No	No	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

Table 7: Product Market Similarity 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. The specifications include the full set of controls as described in Section II. All specifications are run at the firm-year level, include year 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
Avg Similarity Score	-0.158***	-0.160***	-0.158***	-0.160***	-0.169***	-0.173***
	(-3.318)	(-3.305)	(-3.250)	(-3.519)	(-3.938)	(-3.712)
SIC3 HHI	0.017	0.033	0.009			
	(0.831)	(1.458)	(0.393)			
NCOMP_SIC3		0.016**	0.016***			
		(2.164)	(2.699)			
SIC3 PM sale			-0.007			
			(-0.719)			
Top 4 Concentration			0.041**			
			(2.140)			
TNIC HHI				-0.002	0.006	0.006
				(-0.276)	(0.985)	(1.091)
NCOMP_TNIC					0.002	0.002
					(1.113)	(1.299)
Product Market Fluidity						0.000
						(1.062)
Sum Similarity						0.000
						(0.123)
Observations	37,335	37,335	37,335	37,335	37,335	37,335
R-squared	0.034	0.034	0.034	0.034	0.034	0.034
SIC3 FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 8: Fraud and Alternative Measures of Competition

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 according to TNIC, sales based HHI according to SIC3, sum similarity, product market fluidity, average (SIC3) profit margin, top-4 sales concentration and number of competitors constructed using three-digit SIC code. Columns 1-7 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 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
TNIC HHI	0.002						
	(0.397)						
SIC3 HHI		-0.001					
		(-0.108)					
Sum Similarity			-0.000				
			(-0.770)				
Product Market Fluidity				0.000			
				(0.800)			
SIC3 Profit Margin					-0.008		
					(-1.272)		
Top 4 Concentration						0.006	
						(0.719)	
Ln NCOMP SIC3							0.003*
							(1.871)
Observations	37,335	37,335	37,335	36,180	37,335	37,335	37,335
R-squared	0.014	0.014	0.014	0.014	0.014	0.014	0.015
Controls	Full	Full	Full	Full	Full	Full	Full
FE	Year	Year	Year	Year	Year	Year	Year

Table 9: Restatements and Product Similarity (Falsification Test)

This table reports OLS estimates for the relation between accounting restatements and *Average Similarity Score*. We replace our primary dependent variable with restatements (RE AA), obtained from Audit Analytics database from 1996-2012. We do not perform any screens for our restatement variable. Details are discussed in Section IV, Section F. All specifications include year fixed effects and explanatory variables lagged 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) RE AA	(2) RE AA	(3) RE AA	(4) RE AA	(5) RE AA	(6) RE AA	(7) RE AA
Avg Similarity Score	-0.055	-0.117	-0.058	-0.053	-0.098	-0.103	0.066
	(-0.705)	(-1.387)	(-0.669)	(-0.498)	(-0.908)	(-0.984)	(0.509)
R&D			-0.018	-0.019	-0.009	-0.008	0.014
			(-1.318)	(-1.108)	(-0.593)	(-0.562)	(0.639)
R&D dummy			0.010*	0.012**	0.010	0.014*	0.027**
			(1.844)	(2.136)	(1.418)	(1.954)	(2.081)
Ln number analysts			0.004	0.003	0.002	0.002	-0.001
			(1.256)	(0.800)	(0.650)	(0.628)	(-0.272)
Inst Ownership				0.021*			
				(1.714)			
Whited-Wu Index			-0.018**	-0.015	-0.008	-0.038	0.003
			(-2.300)	(-1.623)	(-1.202)	(-0.866)	(0.487)
RSST accruals		-0.001	-0.005	-0.007	-0.004	-0.002	-0.004
		(-0.187)	(-1.033)	(-1.350)	(-0.937)	(-0.343)	(-1.114)
Change AR		-0.032**	-0.028*	-0.028*	-0.028*	-0.020	-0.021*
		(-2.200)	(-1.929)	(-1.735)	(-1.964)	(-1.289)	(-1.680)
Change Inventory		0.029	0.021	0.016	0.025	0.034	0.003
		(1.268)	(0.988)	(0.681)	(1.163)	(1.575)	(0.137)
Pct Soft Assets		-0.001	-0.008	-0.014	-0.012	-0.011	0.019
		(-0.146)	(-0.791)	(-1.168)	(-1.141)	(-1.013)	(1.591)
Change in Cash Sales		-0.001	-0.001	0.000	-0.000	-0.000	0.000
		(-0.393)	(-0.744)	(0.190)	(-0.352)	(-0.015)	(0.094)
Change in ROA		-0.002	-0.004	-0.009*	-0.003	-0.005	-0.007
		(-0.530)	(-0.856)	(-1.896)	(-0.580)	(-1.241)	(-1.371)
Change in employee		-0.001	-0.003	-0.005**	-0.003	-0.002	-0.003
		(-0.819)	(-1.633)	(-2.008)	(-1.463)	(-1.084)	(-1.270)
Ln Age		0.002	0.003	0.004	0.004	0.005	-0.004
		(0.580)	(0.789)	(0.951)	(0.985)	(1.151)	(-0.332)
Dummy Security Issue		0.001	0.001	0.001	0.001	0.002	0.006
		(0.159)	(0.269)	(0.147)	(0.220)	(0.319)	(1.207)
Stock Industry Return R2			0.004	0.001	0.003		0.008
			(0.229)	(0.030)	(0.155)		(0.357)
Relative Perf Eval Flag			0.004	0.003			
			(0.925)	(0.466)			
Ln NCOMP TNIC			-0.002	-0.005**	-0.002	-0.001	0.000
			(-0.941)	(-2.167)	(-0.565)	(-0.287)	(0.033)
Ln Asset		0.011***	0.008***	0.007**	0.009***	0.008***	0.010**
		(6.440)	(3.382)	(2.381)	(3.745)	(3.484)	(2.117)
Observations	61,395	39,519	37,144	28,912	37,144	38,916	36,380
R-squared	0.018	0.021	0.020	0.022	0.039	0.103	0.429
FE	Year	Year	Year	Year	Year + SIC3	Year + SIC3	Year + Firm

Table 10: 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, we replace *Average Similarity Score* with a firm's product market similarity score averaged across its closest 5, 10, and 15 competitors, respectively. In Column 5, we replace *Average Similarity Score* with the *Precision* measure outlined in section III. In Columns 6-8, *Average Similarity Score* is replaced with the (natural log of) number of a firm's rivals in the top 75th, 90th and 95th percentile of similarity scores, respectively in the full cross section of firms in year *t*. 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
Avg Similarity Score	-0.203***							
	(-4.248)							
Avg Top5 Similarity		-0.118***						
		(-3.753)						
Avg Top10 Similarity			-0.129***					
			(-3.383)					
Avg Top15 Similarity				-0.128***				
				(-3.202)				
Avg Score Precision					-0.041***			
					(-4.231)			
Ln NCOMP TNIC 75^{th}						-0.003*		
						(-1.778)		
Ln NCOMP TNIC 90^{th}							-0.003**	
							(-2.265)	
Ln NCOMP TNIC 95^{th}								-0.004**
								(-2.314)
Observations	37,144	37,144	37,144	37,144	37,144	37,144	37,144	37,144
R-squared	0.016	0.016	0.016	0.016	0.016	0.015	0.015	0.016
Controls	Full	Full	Full	Full	Full	Full	Full	Full
FE	Year	Year	Year	Year	Year	Year	Year	Year

Internet Appendix for:

Product Differentiation, Benchmarking, and Corporate Fraud

Audra Boone William Grieser Rachel Li Parth Venkat

Table A1: Product Market Similarity and Corporate Fraud – Standardized Variables

This table reports OLS estimates for the incidence of fraud on standardized RHS variables. Our proxy for corporate fraud includes a combination of AAER misstatements 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 control variables used in Dechow et al. (2011). In Columns 3-6 we include our full set of controls as described in Section II and Column 3 also includes Institutional Ownership. 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. Column 4 includes three-digit SIC code (SIC3) fixed effects, Column 5 year x SIC3 fixed effects, and Column 6 firm fixed effects. All continuous RHS variables are standardized. 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
Avg. Similarity Score	-0.003**	-0.004***	-0.005***	-0.004***	-0.004***	-0.002*
	(-2.089)	(-3.502)	(-3.915)	(-3.946)	(-3.230)	(-1.691)
R&D			0.001	-0.001	-0.001	-0.001
			(0.740)	(-1.008)	(-1.134)	(-0.906)
R&D dummy			-0.000	-0.003	-0.004	0.008
			(-0.011)	(-0.914)	(-1.171)	(1.306)
Ln number analysts			0.001	0.001	0.001	0.009***
			(0.227)	(0.267)	(0.378)	(3.995)
Inst Ownership			0.002			
			(0.792)			
Whited-Wu Index			0.001	-0.000	0.007*	0.000
			(0.859)	(-0.057)	(1.889)	(0.554)
RSST accruals		0.000	-0.001	0.000	0.000	0.000
		(0.550)	(-0.636)	(0.342)	(0.567)	(0.537)
Change AR		0.001*	0.001	0.001	0.002*	-0.000
		(1.805)	(1.460)	(1.296)	(1.817)	(-0.275)
Change Inventory		0.001	0.001	0.001	0.001	0.000
		(0.762)	(1.258)	(0.937)	(1.228)	(0.219)
Pct. Soft Assets		0.005***	0.005***	0.005***	0.005***	0.003**
		(4.227)	(4.186)	(3.825)	(3.735)	(2.214)
Change in Cash Sales		0.003**	0.003**	0.003**	0.004**	0.002*
		(2.236)	(2.057)	(2.227)	(2.359)	(1.776)
Change in ROA		-0.004***	-0.003***	-0.004***	-0.003***	-0.003***
		(-6.131)	(-3.336)	(-5.887)	(-5.182)	(-3.775)
Change in employee		-0.002**	-0.002*	-0.001	-0.001	-0.000
		(-2.105)	(-1.868)	(-1.456)	(-1.421)	(-0.343)
Ln Age		-0.007***	-0.007***	-0.006***	-0.005***	-0.010*
		(-3.402)	(-3.040)	(-3.513)	(-2.967)	(-1.859)
Dummy Security Issue		0.003	-0.001	0.002	0.001	0.003
		(1.479)	(-0.258)	(0.737)	(0.284)	(0.855)
Stock Industry Return R2			-0.002	0.003		0.004*
			(-0.822)	(1.515)		(1.800)
Relative Perf Eval Flag			0.003**			-106.330
			(2.124)			(-0.000)
Ln NCOMP TNIC			0.002	0.002	0.002	0.002
			(1.205)	(0.945)	(0.756)	(0.740)
		0.011***	0.009***	0.011***	0.014***	0.018***
Ln Asset		(4.607)	(3.245)	(2.913)	(3.547)	(4.653)
Observations	50,526	39,465	28,912	37,144	38,910	36,380
R-squared	0.005	0.015	0.014	0.034	0.078	0.437
FE	Year	Year	Year	Year Sic3	Year#Sic3	Year Gvkey

Table A2: Product Market Similarity, Rival IPOs and M&A, and Corporate Fraud (First Stage of 2SLS)

This table reports 2SLS estimates (first stage) for the relationship between product market similarity and corporate fraud. In columns 1-3 (4-6), we use similarity scores with competitors undergoing and IPO (being acquired) as an instrument for the firm's Average Similarity Score. 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) Avg Similarity Score	(2) Avg Similarity Score	(3) Avg Similarity Score	(4) Avg Similarity Score	(5) Avg Similarity Score	(6) Avg Similarity Score
IPO Avg Score	0.342***	0.291***	0.216***			
	(8.424)	(10.554)	(9.382)			
Ln Num Competitor IPO	-0.003***	-0.004***	-0.002***			
	(-3.002)	(-6.682)	(-4.426)			
IPO Size ($)	0.001**	-0.001*	-0.001***			
	(2.304)	(-1.741)	(-2.654)			
Target Avg Score				0.239***	0.163***	0.115***
				(5.906)	(11.243)	(8.431)
Ln Num Competitor Target				0.005	0.000	0.000
				(1.390)	(0.210)	(0.334)
Ln Target MarketCap				-0.001**	-0.001**	-0.000
				(-2.049)	(-2.274)	(-0.483)
Observations	37,335	37,335	37,335	37,335	37,335	37,335
R-squared	0.168	0.229	0.352	0.091	0.179	0.328
Controls		Full	Full		Full	Full
FE	Year	Year	Year SIC3	Year	Year	Year SIC3
Instrument F-stat	70.96	111.38	88.02	34.88	126.39	71.09

Table A3: Rival Fraud, Product Similarity, and Fraud Detection

This table reports estimates for the incidence of fraud on the average similarity of each firm's 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* × *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)
Year FE	Yes	Yes	Yes	Yes

Table A4: Product Market Similarity 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 firm's 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	-4.710***							
	(-3.634)							
Avg Top5 Similarity			-1.807***					
			(-3.423)					
Avg Top10 Similarity					-2.004***			
					(-3.390)			
Avg Top15 Similarity							-2.077***	
							(-3.343)	
Stock Industry Return R2	0.485	-1.248**	0.479	-1.248**	0.477	-1.208**	0.486	-1.216**
	(1.444)	(-2.253)	(1.403)	(-2.122)	(1.393)	(-2.102)	(1.412)	(-2.112)
Ln NCOMP TNIC	0.027	-0.032	0.045	-0.009	0.045	-0.010	0.043	-0.012
	(0.809)	(-0.632)	(1.265)	(-0.171)	(1.279)	(-0.207)	(1.221)	(-0.249)
Relative Perf Eval Flag	0.104***		0.107***		0.104***		0.103***	
	(3.403)		(3.247)		(3.274)		(3.265)	
ROA	-0.220***		-0.231***		-0.227***		-0.226***	
	(-2.902)		(-2.980)		(-2.968)		(-2.956)	
Whited-Wu Index	0.030		0.006		0.012		0.015	
	(0.407)		(0.083)		(0.171)		(0.217)	
Book Leverage	-0.020		-0.021		-0.022		-0.023	
	(-0.444)		(-0.470)		(-0.496)		(-0.532)	
Inst Ownership	0.079		0.083		0.078		0.077	
	(1.512)		(1.547)		(1.490)		(1.481)	
R&D	-1.578***	2.956***	-1.665***	3.192***	-1.663***	3.163***	-1.682***	3.218***
	(-2.898)	(2.906)	(-3.109)	(3.086)	(-3.092)	(3.151)	(-3.109)	(3.213)
R&D dummy	0.120	-0.239*	0.121	-0.236*	0.138	-0.255*	0.145	-0.261*
	(1.241)	(-1.750)	(1.220)	(-1.690)	(1.388)	(-1.859)	(1.452)	(-1.910)
Capex	-0.283	0.073	-0.262	0.025	-0.289	0.085	-0.310	0.125
	(-0.589)	(0.110)	(-0.543)	(0.038)	(-0.599)	(0.129)	(-0.640)	(0.189)
Analyst Num	-0.000	0.005	0.000	0.003	0.001	0.003	0.000	0.004
	(-0.026)	(0.475)	(0.067)	(0.334)	(0.080)	(0.346)	(0.046)	(0.391)
Ln Asset	0.253***	-0.295***	0.248***	-0.286***	0.250***	-0.288***	0.250***	-0.289***
	(6.900)	(-6.479)	(6.757)	(-6.267)	(6.770)	(-6.418)	(6.745)	(-6.444)
Ln Age	-0.427***	0.524***	-0.431***	0.528***	-0.434***	0.529***	-0.438***	0.535***
	(-5.822)	(4.744)	(-5.791)	(4.798)	(-5.841)	(4.903)	(-5.882)	(4.983)
abnormal AAER		0.046***		0.047***		0.045***		0.045***
		(2.817)		(2.616)		(2.628)		(2.610)
abnormal Volatility		-0.579*		-0.608*		-0.585*		-0.577*
		(-1.660)		(-1.672)		(-1.686)		(-1.680)
abnormal Turnover		1.770*		1.746		1.669		1.644
		(1.680)		(1.585)		(1.599)		(1.597)
Disastrous Return		0.181*		0.193*		0.180*		0.177*
		(1.836)		(1.856)		(1.827)		(1.818)
Constant	-2.160***	2.578***	-2.187***	2.426***	-2.198***	2.441***	-2.189***	2.433***
	(-8.344)	(6.505)	(-8.163)	(6.070)	(-8.217)	(6.208)	(-8.164)	(6.196)
Observations	28,031	28,031	28,031	28,031	28,031	28,031	28,031	28,031
FE	Year	Year	Year	Year	Year	Year	Year	Year

Table A5: Product Market Similarity and Corporate Fraud - Industry / Firm FE

This table reports estimates for the incidence of fraud on the average similarity of each firm's 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. Columns 1-3 include three-digit SIC code (SIC3) fixed effects and Columns 4-6 include firm fixed effects. All specifications include the full set of controls as described in Section II. 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) Fraud	(2) Fraud	(3) Fraud	(4) Fraud	(5) Fraud	(6) Fraud
Avg Similarity Score	-0.175***	-0.171***	-0.171***	-0.089*	-0.085*	-0.092*
	(-3.755)	(-3.915)	(-3.936)	(-1.750)	(-1.660)	(-1.785)
TNIC HHI		0.006	0.006		0.002	0.002
		(0.986)	(0.968)		(0.384)	(0.362)
R&D		-0.011	-0.011		-0.005	-0.009
		(-1.114)	(-1.083)		(-0.346)	(-0.626)
R&D dummy		-0.003	-0.003		0.009	0.008
		(-0.919)	(-0.930)		(1.271)	(1.177)
Ln number analysts			0.000			0.009***
			(0.253)			(3.922)
log NCOMP TNIC	0.001	0.002	0.002	0.002	0.002	0.002
	(0.729)	(1.159)	(1.107)	(0.812)	(0.803)	(0.691)
RSST accruals	0.002	0.001	0.001	0.002	0.002	0.002
	(0.434)	(0.335)	(0.324)	(0.421)	(0.406)	(0.407)
Change AR	0.016	0.016	0.016	-0.004	-0.004	-0.003
	(1.306)	(1.305)	(1.297)	(-0.271)	(-0.282)	(-0.232)
Change Inventory	0.020	0.020	0.020	0.004	0.004	0.004
	(0.929)	(0.950)	(0.956)	(0.258)	(0.275)	(0.278)
Pct Soft Assets	0.019***	0.018***	0.019***	0.011	0.011	0.013
	(4.079)	(3.637)	(3.805)	(1.357)	(1.357)	(1.611)
Change in Cash Sales	0.005**	0.005**	0.005**	0.003**	0.003**	0.003**
	(2.172)	(2.244)	(2.232)	(2.383)	(2.391)	(2.408)
Change in ROA	-0.021***	-0.021***	-0.021***	-0.017***	-0.018***	-0.017***
	(-5.870)	(-5.791)	(-5.882)	(-4.161)	(-4.046)	(-3.855)
Change in employee	-0.003	-0.003	-0.003	-0.000	-0.000	-0.001
	(-1.507)	(-1.431)	(-1.438)	(-0.077)	(-0.048)	(-0.316)
Ln Age	-0.009***	-0.009***	-0.008***	-0.014**	-0.014**	-0.013**
	(-3.607)	(-3.659)	(-3.510)	(-2.079)	(-2.088)	(-2.048)
Dummy Security Issue	0.002	0.002	0.002	0.003	0.003	0.003
	(0.697)	(0.825)	(0.752)	(1.069)	(1.035)	(0.896)
Stock Industry Return R2	0.016	0.017	0.017	0.022**	0.022**	0.022**
	(1.481)	(1.513)	(1.516)	(1.995)	(2.004)	(2.038)
Whited-Wu Index	-0.001	-0.000	-0.000	0.002	0.002	0.002
	(-0.156)	(-0.063)	(-0.064)	(0.481)	(0.479)	(0.443)
Ln Asset	0.006***	0.006***	0.006***	0.014***	0.013***	0.009***
	(3.864)	(3.713)	(2.906)	(5.739)	(5.301)	(3.647)
Observations	37,144	37,144	37,144	36,380	36,380	36,380
R-squared	0.034	0.034	0.034	0.437	0.437	0.437
FE	Year x SIC3	Year x SIC3	Year x SIC3	Year x Firm	Year x Firm	Year x Firm

Table A6: Product Market Similarity and Corporate Fraud - Non-Linear Specifications

This table reports estimates for the incidence of fraud on the average similarity of each firm's 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 Similarity 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

Table A7: Product Market Similarity and Corporate Fraud
Alternate Similarity Scores (Standardized)

This table reports estimates for the incidence of fraud on alternative standardized similarity scores of each firm's 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 variables of interest (not controls) are standardized for purposes of comparison. In Column 1, we report results for the main dependent variable used throughout our analysis. In Columns 2-4, we replace *Average Similarity Score* with a firm's product market similarity score averaged across its closest 5, 10, and 15 competitors, respectively. In Column 5, we replace *Average Similarity Score* with the *Precision* measure outlined in section III. In Columns 6-8, *Average Similarity Score* is replaced with the (natural log of) number of a firm's rivals in the top 75th, 90th and 95th percentile of similarity scores, respectively in the full cross section of firms in year *t*. 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) Fraud	(2) Fraud	(3) Fraud	(4) Fraud	(5) Fraud	(6) Fraud	(7) Fraud	(8) Fraud
Standardized Avg Similarity Score	-0.005*** (-4.248)							
Standardized Avg Top5 Similarity		-0.007*** (-3.753)						
Standardized Avg Top10 Similarity			-0.007*** (-3.383)					
Standardized Avg Top15 Similarity				-0.006*** (-3.202)				
Standardized Precision					-0.005*** (-4.231)			
Standardized Ln NCOMP TNIC 75^{th}						-0.004* (-1.778)		
Standardized Ln NCOMP TNIC 90^{th}							-0.004** (-2.265)	
Standardized Ln NCOMP TNIC 95^{th}								-0.004** (-2.314)
Observations	37,144	37,144	37,144	37,144	37,144	37,144	37,144	37,144
R-squared	0.016	0.016	0.016	0.016	0.016	0.015	0.015	0.016
Controls	Full	Full	Full	Full	Full	Full	Full	Full
FE	Year	Year	Year	Year	Year	Year	Year	Year

Table A8: Product Similarity 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 Num	-0.072**	-0.146*	-0.221***	-0.511***
	(-2.354)	(-1.850)	(-2.625)	(-3.564)
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
Year FE	YES	YES	YES	YES

Table A9: Product Market Similarity and Financial Statement Comparability

This table reports estimates for the incidence of fraud on the average similarity of each firm's 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 3, including Year#Sic3 fixed effect transformations. The specification in Column 1 is the specification from Table 3 where Accounting Comp does not equal missing. The specification in Column 2 includes control 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** (-2.32)	-0.150** (-2.34)	0.039 (0.04)	-0.265*** (-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 A10: 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)
VARIABLES	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