Product Similarity, Benchmarking, and Corporate Fraud

Audra Boone, William Grieser, Rachel Li, and Parth Venkat*

Abstract

We document that firms with greater product similarity to their peers exhibit lower rates of financial fraud. We show that peer similarity is associated with better information environments, which is consistent with monitors' enhanced ability to benchmark against other firms. The negative relation between product similarity and fraud remains after controlling for alternative mechanisms including incentive compensation structures, competition, and internal and external governance characteristics. Overall, our findings suggest that greater peer similarity increases the marginal cost of fraud, and therefore, ex-ante disincentivizing managers from committing fraud.

* Boone (Corresponding author), AUDRA.BOONE@tcu.edu, Texas Christian University Neeley School of Business; Grieser, Texas Christian University Neeley School of Business; Li, U.S. Securities and Exchange Commission; Venkat, U.S. Securities and Exchange Commission. The U.S. Securities and Exchange Commission disclaims responsibility for any private publication or statement of any SEC employee or Commissioner. This article expresses the author's views and does not necessarily reflect those of the Commission, the Commissioners, or members of the staff. This paper was initially released prior to the authors joining the Commission. We thank the editor and an anonymous referee for constructive feedback and Anne Albrecht, Khrystyna Bochkay, Donald Bowen, Jeff Chen, Joey Engelberg, Will Gerken, Umit Gurun, Gerard Hoberg, Andy Imdieke, Simi Kedia, Dan Li, Zack Liu, Michelle Lowry, Tanakorn Makaew, Gonzalo Maturana, Karen Nelson, Jordan Neyland, Joerg Picard, Vesa Pursiainen, Nathan Swem, Xinxin Wang, Jared Wilson, participants at the 2019 Conference on Financial Market Regulation, the 2018 FMA Asia Annual Conference, the 2018 FMA Annual Meeting, the 2018 Australasian Banking and Finance Conference, the 2018 New Zealand Finance Conference, the 2019 Midwest Finance Association Conference, and the 2022 Boca Corporate Finance and Governance Conference, as well as seminar participants at Drexel University, Clemson University, Southern Methodist University, Texas Christian University, Universidad de los Andes, University of Nevada Las Vegas, the U.S. Securities and Exchange Commission, and Villanova University for helpful comments. We further thank Gerry Martin for graciously sharing data on the incidence of financial misreporting.

I. Intro

A manager's decision to commit fraud involves trading off the perceived financial incentives against the potential costs from commission (see Becker (1968)). Compensation schemes with a greater incentive component (i.e., stock and option compensation), for example, could entice managers to manipulate performance metrics to inflate the firm's stock price (see Goldman and Slezak (2006) and Johnson, Ryan, and Tian (2009)).

The potential costs of fraud commission include the perceived likelihood of getting caught and the severity of punishments. Karpoff, Lee and Martin (2008) find that when financial fraud is uncovered, each dollar of inflated value corresponds to more than four dollars of market capitalization losses, resulting from both direct penalties and fines and reputational losses. Other work has investigated how various factors, which can aid fraud detection, may affect the likelihood of fraud (see, e.g., Jones (1991), Dechow and Dichev (2002), McNichols (2000), Dechow, Ge, Larson, and Sloan (2011), and Karpoff (2021)). Importantly, a greater ability by external parties to identify fraudulent activity can curb the commission of fraud (see Khanna, Kim and Lu (2015)).

Drawing on prior work showing that peer disclosures can improve a firm's information environment and lead to better board decisions (see Murphy (1986)), improved analyst forecasts (see De Franco, Kothari and Verdi (2011)), and reduced earnings management (see Sohn (2016)), we conjecture that peer firms operating in a similar product market space offer sharper context for monitors to assess a firm's economic circumstances and detect fraud. In turn, managers should be more disciplined and have lower incentives to commit fraud (see Tirole

(2010)). We test this notion and show that the incidence of fraud exhibits a negative relation with the quality of peer firm benchmarks.

To capture whether firms have better benchmarks to assess economic circumstances, we rely on text-based pairwise product similarity scores developed by Hoberg and Phillips (2016) with a higher score suggesting two firms share a more similar product space. To measure a firm's overall product similarity, we calculate the average pairwise similarity score between each firm and all its peers. Therefore, a higher average similarity score suggests that a firm operates in product spaces that overlap with its peers to a greater extent, making it easier to compare financial disclosures with its peers. If product similarity disciplines managers through enhanced benchmarking, we expect the comparison to closer peers to be more valuable in detecting financial abnormalities that could indicate fraud. Therefore, we also use a tighter measure of product similarity using the average similarity score of each firm's 15 closest peers in its product space.

The product similarity measures are based on the words firms use in their business descriptions that reflect their operating environment and competitive landscape. Importantly, the words in and of themselves do not determine or change the information environment. Instead, overlapping terms denote how readily outsiders can assess a firm's economic situation and ascertain whether its financials comport with its peer firms. As with all disclosure-based measures, one may be concerned with the possibility of manipulation. Hence, we also test our hypothesis using alternative proxies for product similarity including Parrino's (1997) industry homogeneity measure and stock-return co-movement.

Consistent with our conjecture that closer peers offer sharper context to identify fraudulent activity, our analyses reveal that a one standard deviation increase in the various product similarity measures is associated with an approximately 11-22% decline in the rate of fraud in the following year. We further show that these results are robust to (a) alternative estimation procedures (i.e., conditional logit regressions), (b) models with different fixed effects and additional control variables (e.g., variables related to fraud, industry characteristics, and industry competition), and (c) using an alternative sample after applying a matching procedure to improve the comparability between the treatment and control groups. These results suggest that firms with similar peers, and thus enhanced benchmarking, are less likely to commit fraud unconditionally, plausibly due to the ex-ante increased risk of being caught.

An empirical challenge in studying fraud is that only detected, rather than all committed, fraud is observable (see Dyck, Morse and Zingales (2023) and Dimmock and Gerken (2012)). Hence, to further confirm our main findings, we employ the predicted probability of fraud as the dependent variable (see Alawadhi, Karpoff, Koski, and Martin (2023)), as well as a bivariate probit estimation that simultaneously estimates the commission rate and detection rate. We continue to find a significantly robust negative relation between product similarity with peers and the incidence of financial fraud.

We postulate that product similarity curtails managers' incentives to engage in fraudulent reporting through an enhanced information environment and is, therefore, more effective when the marginal value of benchmarking is higher. Organizationally complex firms that operate across many segments (e.g.,) are more difficult to monitor, and consequently are associated with more costly governance and a greater likelihood of accounting misstatements (see, e.g., Bushman, Chen, Engel, and Smith (2004), Cohen and Lou (2012), Peterson (2012), and Hoitash and Hoitash (2018)). For these firms, the marginal information provided by peer comparisons should impact fraud detection more than in simple firms where the ex-ante monitoring ability is more effective.

Consistent with our expectation, we find a more negative relation between product similarity and fraud, both statistically and economically, for more complex firms than for less complex firms. We further provide corroborating evidence that product similarity improves information environment by showing that new information released by a closely related peer firm is associated with lower information asymmetry.

Even though our findings are consistent with the benchmarking channel, we acknowledge the empirical challenges in ruling out all other mechanisms that could (partially) explain our results. Managerial incentives – including the decision to commit fraud – are endogenously determined by a firm's product market, information environment, and corporate governance. Therefore, product similarity can correlate with fraud through various channels such as incentive compensation, competition, board oversight, and external monitoring. While we cannot completely rule out all alternative channels, we provide evidence that the negative relation between product similarity and fraud remains after controlling for these other factors.

First, the expected gains and losses from fraud can affect managers' incentives to commit fraud. On the one hand, a higher percentage of equity-based compensation can increase a manager's incentive to commit fraud because an inflated valuation would increase CEO's near-term personal financial gains. Further, higher equity-based compensation can induce managers to conceal negative news about future growth options and to choose suboptimal investment policies to support the pretense (see Benmelech, Kandel, and Veronesi (2010)). On the other hand, equity-based compensation could disincentivize managers from committing fraud. As managers accumulate a higher equity stake in the firm, the expected costs of fraud become higher for them due to the large loss in market capitalization after the revelation of fraud (see Karpoff et al. (2008)) Therefore, the fraud rate can be an equilibrium outcome of these trade-offs. We show that the negative relation between product similarity and fraud is robust controlling for various measures of CEO financial incentives, measured by well-aligned CEO compensation plans, CEO compensation packages, and CEO pay-performance sensitivity.

Second, we examine the fraud rate following a shock to competition that does not significantly alter the information environment, which we capture via large tariff reductions.[1] Intense competition can potentially shift managers' incentives to commit fraud as it can create pressure to raise capital when facing intense competition (see Shleifer (2004)). In contrast, intense competition can also lower the likelihood of fraud by reducing managerial rent extraction (see Nickell (1996) and Ades and Di Tella (1999)); accordingly, the negative relation between product similarity and fraud could reflect shifts in competition. While we do not observe consistent direct effects of competition on fraud rate, we find that product similarity continues to be associated with a lower incidence of fraud even after controlling for a competition shock as well for other proxies for competition.

Lastly, the relation between product similarity and the incidence of fraud could reflect better internal governance or external monitoring rather than an improved information environment.[2] For example, in equilibrium, firms that have similar peers could operate in a more

[1] Tariff reductions increase the intensity of foreign competition (see Fresard, 2010) without affecting the quantity and quality of available public information.
[2] For example, better internal governance is shown to reduce the incidence of fraud (see Beasley (1996) and Ege (2015)).

competitive product space, and therefore require a higher quality board of directors that more effectively prevents fraudulent reporting. Similarly, when firms have more external monitoring, the marginal value of higher product similarity might be low.

Controlling for high-quality internal governance through board independence, as measured by independent directors and CEO-chairman duality, we continue to observe a robust relation between product similarity and fraud. Likewise, proxies for better external governance, such as institutional ownership, analyst coverage, or having a top auditor, do not explain our findings and product similarity is still associated with a lower incidence of fraud. Thus, our results suggest that product similarity offers a novel source of information due to better benchmarking with peers. Overall, our findings are consistent with product similarity increasing detection and providing supplementary discipline to other governance mechanisms.

II. Data and Key Variables

A. Product Similarity Scores and Measures

Our use of the product similarity score has several advantages. Consider Nike, Inc., which has multiple product lines but has a firm-level SIC code of 3021, "Rubber and Plastic Footwear." Nike also earns 30% of its revenue from apparel (SIC codes 2300-2399), so often peer firms operating in this product space would be classified as unrelated. Using the product similarity score, it is easy to demonstrate that Nike shares a product relationship with Under Armour (0.096) and Columbia (0.094) but has lower similarity with Lululemon (0.017). Nike also receives 6% of their revenue from equipment, predominately from golf-related items. In 2014, Nike had a 0.033 product similarity score with Top Golf Calloway (SIC 3949) but a low similarity to Head (also SIC 3949) whose products consist mainly of tennis and ski-related products.[3] As companies become increasingly operationally complex, the benefit of similarity scores prevails as conventional industry classifications often fail to provide comprehensive representation of the product markets.

Therefore, we construct our proxies for product similarity using a measure developed by Hoberg and Phillips (2010), (2016) that captures the pairwise similarity of a given firm's product market descriptions with other firms filing a 10-K each year. Specifically, the authors calculate the cosine similarity, which measures the angle between two-word vectors on a unit sphere between two firms' business descriptions in their 10-K annual filings. The pairwise similarity is higher when the product market descriptions between the two firms exhibit greater overlap. The measure ranges from zero (no similarity) to one (perfect similarity).

[Insert Table 1]

To capture the extent to which a firm operates in overlapping product spaces, we first propose a simple proxy, *Product Similarity (All-Firms),* which is the average of the pairwise similarity scores of all peer firms that have some overlapping business description (i.e., a cosine value larger than 0) each year. *Product Similarity (All-Firms)* ranges from 0.016 to 0.085 with a mean of 0.035 and a median of 0.031, as shown in Panel A of Table 1. While this measure is intuitive and considers the information provided by all peers that exhibit some degree of overlap with the focal firm in its operating environment and competitive landscape, there can be substantial variation in the number of peers and closeness of peers being averaged across each firm. Consequently, this equal-weighted measure might not accurately reflect the incremental

[3] Further, while many firms have product overlap with Nike, they may or may not compete directly in the same product space. For instance, Crocs shares a high similarity to Sketchers (Score of 0.131) but should have no direct competitive relationship with the equipment companies (Top Golf or Head).

value of information by each peer. For example, comparative disclosure from a closer peer should provide greater information value in assessing a firm's economic situation, compared to a distant competitor. Because more closely related peers should provide the greatest information externalities, we propose an alternative, and tighter, measure that focuses on each firm's closest 15 peers (*Product Similarity (Top-15))*. Consistent with our expectation, and as shown in In Panel B of Table 1, the dispersion of common accounting ratios is indeed smaller among a firm's 15 closest peers than for those of all peers, suggesting that a firm's closest peers provide more valuable information and better benchmarks to the focal firm, compared to distant peers. *Product Similarity (Top-15)* ranges from 0.071 to 0.474 with a mean of 0.186 and a median of 0.168.

B. Industry Classifications

Because the rate of fraud varies by industry, it is important that our analyses account for any omitted variables at the industry level that affect the outcome of fraud. Following Hoberg and Phillips (2016), we categorize industries based on the text-based network industry classification (TNIC), which contains granular information that results in a classification scheme that is less constrained than either Standard Industrial Classification (SIC) or North American Industry Classification System (NAICS).

One benefit of using the TNIC is that it is updated annually, leading to improved classification accuracy. 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 pairwise similarity with all the competitors in its TNIC (from 0.028 to 0.021 for *Product Similarity (All-Firms)* and 0.17 to 0.08 for *Product Similarity (Top-15)*). However, the divestment resulted in no change to Exxon's SIC code. A second benefit is that TNIC is intransitive, compared to traditional industry classification, which allows for more complex characterizations and intensive margin. Using the same example of Nike (SIC code 3021), the only other firm with the same SIC code is Crocs, and in 2014, Crocs and Nike had a relatively high similarity score of 0.085 (above the 75th percentile). Nike, however, is also related to many other firms that belong to a different SIC code; for example, Nike has a similarity score of 0.057 with Sketchers, which belongs to SIC 3140 - Footwear, except rubber.

In sum, the product similarity score and the text-based network industry classification provide more accurate, dynamic, and comprehensive depiction of firms' product market space and the competitive landscape, making them more suitable than other competition measures to study the relation between benchmarking and managerial incentives to commit fraud.

C. Alternative Similarity Measures

We contend that measuring product similarity based on firms' business description directly reflects their operating environment and competitive landscape with peers, and thus provides a robust way to assess a firm's underlying economic comparability. Like all other disclosure-based measures, however, a potential concern is that firms could intentionally adjust the terms in their business descriptions to appear more or less close to other companies. To confirm that the chosen words in the disclosure do not drive our results, but rather the underlying economic similarity, we use two alternative similarity proxies in addition to the text-based similarity measures.

The first proxy is industry homogeneity measure developed by Parrino (1997), which measures the correlation between common stock returns within two-digit SIC industries. By construction, this measure assigns an identical value to all firms in the same industry, which leads to lower variation. As shown in Panel A of Table 1, the dispersion of Industry Homogeneity, measured as standard deviation scaled by mean, is 0.25, which is much lower compared that of our text-based product similarity measures, whose values are 0.47 and 0.44 respectively. While this measure cannot directly capture the variation in managerial incentives at the firm level, if enhanced fraud detection works through the product comparability channel, we expect to observe a lower incidence of fraud across firms.

The second measure we use is stock-return co-movement because it is, at least in part, linked to the degree that firms share economic fundamentals such as cash flows and risk (see, e.g., Hameed, Morck, Shen, and Yeung (2015) and de Bodt, Eckbo, and Roll (2024)). Specifically, we start by estimating a rolling-60-month pairwise monthly stock return correlation among all firms in our sample. Similar to the text-based similarity measure, for each firm, we then calculate a simple average of pairwise stock return correlation among 15 firms with which the firm has highest correlations. These market-based correlation measures likely reflect the market's perception of firms' underlying economic similarity and are less susceptible to manipulation. However, we caveat that co-movement measures often contain noticeable noise and can be affected by market frictions and investor preference (see Barberis, Shleifer, and Wurgler (2005)), which may amplify measurement error in estimations.

[Insert Figure 1]

As shown in Figure 1, our text-based product similarity measures are positively correlated with these alternative proxies. Table 2 reports the pairwise correlation between all product similarity proxies. As shown, the correlation between Product Similarity (All-Firms) and Product Similarity (Top 15) is 0.89; the correlation between Product Similarity (All-Firms) and

Industry homogeneity measure is 0.39 and the correlation between Product Similarity (Top 15) and Industry homogeneity measure is 0.46; and the correlation between Product Similarity (Top 15) and Co-movement (Top 15) is 0.04.

[Insert Table 2]

D. Financial Fraud

Consistent with prior work, we define corporate financial fraud as “the intentional, material misstatement of financial statements that causes damages to investors," based on regulatory enforcement actions and class action lawsuits (see Donelson, Kartapanis, McInnis, and Yust (2021)). Because it takes time for fraud-related lawsuits and investigations to occur, there often exists a lag between a firm’s violation period and when it is targeted for enforcement actions. For our study, we classify fraud years as those when the accounting fraud occurred, rather than when it was detected and disclosed.[4] Furthermore, to avoid truncation bias due to this lag, we calibrate and evaluate the prediction models using data on misrepresentation that occurs only through 2014.

For our regulatory enforcement actions, we use two sources. First, we obtain Accounting and Auditing Enforcement Releases (AAER) data for the sample period 1996-2014 from 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 misstatement investigations in our sample occur because of bribery, fraud, inflated assets, financial reporting

[4] Figure IA1 provides a timeline illustrating a fraud incident in our sample, including the commission and detection periods and the resulting punishment.

related enforcement actions concerning civil lawsuits brought in federal court, and orders concerning the institution and/or settlement of administrative proceedings.[5] Second, following Alawadhi et al. (2023), we also include Department of Justice (DOJ) enforcement actions for financial misrepresentation.

For class action lawsuits, we construct our sample following the work of Thompson and Sale (2003), Griffin, Grundfest, and Perino (2004), Choi, Nelson, and Pritchard (2009), and Jayaraman and Milbourn (2009). We download all class action lawsuits from the Securities Class Action Clearinghouse (SCAC) for 1996 through 2014 and only include 10b-5 class action lawsuits, which eliminates lawsuits that occur for non-financial reasons.

Our primary dependent variable, *Fraud*, is a binary variable equal to one for all firm years in which fraud occurred that was later disclosed via regulatory enforcement action or class action lawsuits where managers engage in actions that are intentionally misleading and damaging to investors.[6] We exclude firms headquartered outside the United States, ADRs, REITs, firms with assets less than $1M, penny stocks and unit offerings, and firms with missing assets or sales items in the Compustat database. Our final sample of corporate fraud events contains 2,217 of 83,412 firm-years flagged as fraudulent from 1996 to 2014 that impact 712 unique firms. These statistics are closely in line with those of Dyck, Morse, and Zingales (2010). As shown in Panel A of Table 1, the incidence of discovered fraud in our sample is 2.66%.

[5] 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.
[6] Though Karpoff et al. (2017) note that each potential database on misconduct only encapsulates a portion of such events, we use a combination of public and private enforcement actions and class action lawsuits to increase the likelihood of capturing the most egregious and intentional misreporting. We also run additional checks such as whether the SCAC was dismissed to capture instances of fraud more accurately.

In addition, we use an alternate set of fraud-related dependent variables created by Alawadhi et al. (2023). The first variable is a continuous predicted probability of fraud estimated using parsimonious logistic model with regulatory enforcement actions for financial misrepresentation, which has a mean of 1.8%. Alawadhi et al. (2023) construct two additional fraud indicators comparing the predicted probability of fraud to an optimal cut-off for one and three-consecutive years. Although these indicators are also binary left-hand-side variables, these implied fraud indicators have an unconditional mean of 41.1% and 26.5%, respectively. Therefore, they are less susceptible to the misspecification issues of linear probability models with low fractions of observed fraud. Descriptive statistics of these variables are presented in Panel A of Table 1.

E. Additional Data

Throughout our analyses, we include control variables that are shown to affect the likelihood of fraud such as firm size, firm age, profitability and growth rates, etc. We also construct measures for information asymmetry, proxies for corporate governance and CEO incentive compensation to explore the possible mechanism through which product similarity may affect the outcome of fraud. All variable definitions are presented in the Appendix, and we present the descriptive statistics of financial and accounting controls in Panel A of Table 1. Our financial data comes from CRSP and Compustat and IPO data comes from Thomson Reuters. Analyst coverage is from IBES and the duties and customs values are collected from the U.S. International Trade Commission. Characteristics of the board and executives are from BoardEx, and the blockholder data is from Schwartz-Ziv and Volkova (2023) obtained via Wharton Research Data Services (WRDS). We detail each control variable during the discussion of our results, and definitions are available in the Appendix.

F. Matching

It is possible that firms with particular characteristics are more likely to have greater product similarity to peers, as well as higher propensities to commit fraud, potentially confounding any observed relation between product similarity and financial misreporting outcomes. Additionally, our binary dependent variable, *Fraud,* has a highly skewed distribution with a large number of zeros (no fraud) and a small fraction of ones (fraud), which could lead the linear-probability-model (LPM) to be biased and inconsistent.

To address such concerns, we create a control sample using a matching procedure. Specifically, for each firm with reported financial fraud, we identify a set of firms within the same industry (measured by three-digit SIC code). We then refine the matches based on age quintiles and then select five firms closest in size (measured by market capitalization) to our focal firm to be our control group.[7]

III. Product Similarity and Financial Fraud

The incentives to commit fraud depend on both the severity of punishment conditioned on getting caught and the probability of detection (see Becker (1968)). That is, fraud commission can be dissuaded by an improvement in fraud detection without necessarily increasing the severity of punishment (and vice versa).

Firms operating in more homogeneous product markets where peers are subject to similar demand and cost conditions are likely to have higher product similarity (see Tirole (1988)).

[7] We report the statistical tests on firm characteristics of matched sample in Panel B of Table IA1. We also use alternatively matching selections (1-to-1 and 1-to-3) and find qualitatively similar results.

Consequently, comparing disclosures among closer peers likely provides sharper context to assess the firm's economic circumstances and increases the probability of fraud detection, and hence, also elevates the expected costs of committing fraud. Accordingly, we posit that managers are less likely to commit financial fraud when they anticipate that external observers can more effectively benchmark their firm's disclosures with those of their peers, which leads us to our first hypothesis:

H1: Ceteris paribus, product similarity is negatively associated with the incidence of financial fraud.

An alternative hypothesis (H1a) is that firms with greater peer similarity are positively associated with the incidence of fraud. One explanation is that the incidence of fraud is jointly determined by both commission and detection, this association could occur if benchmarking from product similarity helps create more effective fraud detection without reducing the ex-ante incentive to misreport. It is also possible that we would observe no relation if peer firm disclosures have no systematic effect on the likelihood of fraud detection or punishment, and therefore does not affect managerial behavior.

[Insert Figure 2]

In Figure 2, we observe a negative relation between fraud and various product similarity measures. These are binned scatter plots of the incidence of fraud against product similarity. The plots control for firm size and age as these factors have been associated with factors such as financial statement readability and firm complexity, which could also be correlated with the propensity to commit fraud. This relation provides initial evidence that product similarity is associated with a lower incidence of fraud. We next formally test our hypotheses in a regression setting.

[Insert Table 3]

Table 3 reports partial effects from a Linear Probability Model (LPM) where we formally test the hypotheses outlined above. In the firm-year panel, we investigate whether and how product similarity in year *t* affects the incidence of fraud or financial misreporting in year *t+1*.[8] In column 1, controlling only for firm *Size* and firm *Age*, we observe a large negative relation between a firm's average product similarity and the incidence of fraud.

In column 2 we add several control variables based on the work by Alawadhi et al. (2023), as well as industry-by-year (TNIC x year) fixed effect transformations. The control variables include SIC3 Herfindahl-Hirschman Index (*HHI SIC3*), *Change in Gross Margin*, *Sales Growth*, percentage of soft assets (*Soft Asset)*, the natural logarithm of the number of reported geographic segments (*Geographic Segments)*, and indicator variables for whether a firm in a given year has *Operating Lease* obligations, *Security Issued*, had net losses (*Losses)*, had an *Audit Opinion*, was audited by a top auditor (*BigN Auditor)* or was in either the *Business Equipment* or *Telecom* Fama-French Industry. We continue to observe a large negative relation between a firm's average product similarity score and the incidence of fraud. We use this specification in most of our tests moving forward.[9]

[8] The unit of observation is at the firm-year level, right hand side variables are lagged one year, and the t-statistics are calculated from standard errors clustered by firm.

[9] We exclude five variables suggested by Alawadhi et al. (2023). Each of these variables has missing values for at least 10,00 observations. ALS accruals (see Allen, Larson, and Sloan (2013)) and business segment HHI significantly reduce the sample size but add no explanatory power to the regressions. The other three variables: change in SG&A, change in receivables, and change in inventory, do have some explanatory power for fraud but due to missing observations, reduce the sample substantially and could be components of accounting manipulation.

In column 4, we present regression coefficients using standardized regressors from column 2 as both the mean and standard deviation of *Product Similarity (All-Firms)* (0.035 and 0.016, respectively) are close to 0, making it difficult to interpret the coefficients. In the standardized regressions, we find that a one standard deviation change in *Product Similarity (All-Firms)* is associated with a 0.417% decrease in the incidence of fraud, which corresponds to a 15.7% reduction relative to the unconditional mean (2.66%). This coefficient is not only statistically and economically significant, but the partial effect on fraud from either measure of product similarity is larger than most other control variables in our regression.

In columns 3 and 5, we repeat the tests from columns 2 and 4 but rely on a tighter product similarity measure that computes the average similarity score based exclusively on a firm's 15 closest peers, *Product Similarity (Top-15)*. If product similarity offers greater context to assess economic circumstances, then closer peers should offer even more ability to evaluate reported performance than more distant ones. Using this alternative measure, our results continue to hold, are of similar magnitude, and are more precisely estimated (i.e., lower standard errors). The former is unsurprising as the two measures have a correlation coefficient of 89% in the sample as shown in Panel A of Table IA2 of the Appendix. In Panel B of Table IA2, we also show specifications for alternate independent variable constructions of product similarity using both the top 10 and top 5 peers, and the results remain similar. Thus, in future specifications, we present our tests using *Product Similarity (All-Firms)* as well as *Product Similarity (Top-15)*.

[Insert Figure 3]

Most fraud events occur over multiple years, and thus, to mitigate the concerns of serial correlation in standard errors, we cluster by firm in the linear probability estimations. In addition, in Figure 3 we display the partial effects estimation of pure cross-sectional regressions each year with the control variables from above and TNIC fixed effects. The negative relation between product similarity (using both measures) and fraud is consistently negative. These findings support the idea that our results are unlikely driven by serial correlation or fraud waves and are robust across long time periods. Furthermore, in Table IA3, we present the cross-sectional relation between product similarity and first-year fraud in a matched sample by industry, age, and size and confirm that higher product similarity is associated with a lower probability of fraud.

One concern regarding our estimation is that our primary dependent variable, *Fraud*, is a binary outcome with a low fraction of observations that equal one, potentially causing LPMs to be mis-specified. For this reason, we employ three empirical strategies to verify our primary findings. Specifically, we use (a) a conditional logit model, (b) a matching procedure, and (c) a continuous predicted probability of fraud (rather than indicator) and more balanced indicator variables as the left-hand-side variable, as designed by Alawadhi et al. (2023).

[Insert Table 4]

The non-linear nature of the conditional logit model better manages dependent indicator variables containing mostly zero outcomes that make LPMs susceptible to biases and inconsistencies. It also can control for all fixed characteristics within a group (often referred to as fixed-effects logit). Columns 1-4 in Panel A of Table 4 display the odds ratios (as opposed to partial effects) from the conditional logit regressions using the specification analogous to column 2 of Table 3. As shown in column 1, we continue to observe a robust negative relation between *Product Similarity (All-Firms)* and the incidence of fraud, but with greater statistical significance than the LPM estimates, which is likely due to the added efficiency of the conditional logit estimator.[10]

The odds ratio in column 1 is close to zero, which is interpreted as a change in *Product Similarity (All-Firms)* from 0 to 1 (an unrealistically large change) results in a 100% reduction in the probability of fraud. To ease interpretation, we use standardized regressors in columns 2-4. The coefficient on *Product Similarity (All-Firms)* suggests that a higher product similarity is associated with a lower incidence of fraud by 17% (odds ratio of 0.83, which is 17% less than 1), which is similar to, and slightly stronger than, the partial effect estimated by the OLS regressions. In column 3, we use *Product Similarity (Top-15)* and find a stronger partial effect of a 22.2% reduction (odds ratio of 0.778, which is 22.2% less than 1) in the incidence of fraud. In column 4, we confirm that *Product Similarity (Top-15)* dominates *Product Similarity (All-Firms)* in a horse race between the two. For consistency, we use the conditional logit model as our primary estimation method for the majority of our remaining tests.

To further address the potential LPM misspecification, we create a matched sample to the treated fraud firms to achieve a more balanced distribution of the dependent variable. As noted in subsection 2.6, we select matches based on industry, firm size, and firm age and include the entire time-series for those matched firms and the firms with at least one fraud year. Using the matched group, 11.78% of the firm-years are labeled as fraud (compared to 2.66% in the full sample). The results shown in columns 5 and 6 confirm the negative relation between both measures of product similarity and fraud. Notably, the partial effect is significantly larger than when using the full sample, suggesting that LPM could bias against our findings.

[10] Because the conditional logit model requires some degree of variation within a group (i.e., at least some firm-years with fraud and others without fraud within each TNIC-Year), the number of observations drops significantly.

Another potential concern is that firms could manipulate their product market descriptions, and therefore, obfuscate the interpretation of the relation between product similarity and incentives to commit fraud through the changes in information environment. Although firms have discretion over some aspects of their disclosures, these product descriptions are legally required to be accurate based on Item 101 of Regulation S-K, and firms must update them each fiscal year.[11] As shown in Figure 1 and Table 2, the text-based product similarity measures are positively correlation with both the industry homogeneity measure (see Parrino (1997)) and stock-return co-movement measure, suggesting that our text-based similarity measures are reflective of a firm's underlying economic conditions.

Nevertheless, we next show that our results are robust to alternative measures of product similarity in Panel B of Table 4. Columns 1 and 2 use LPM estimation with matched sample, and columns 3-6 use conditional logit estimation similar to those in Panel A Table 4; Columns 1-4 confirms that higher *Industry Homogeneity (Stock-return Co-movement)* is associated with a decrease in the incidence of fraud. These findings provide corroborative evidence that it is product similarity, rather than a particular form of proxy, that disincentivizes managers from committing fraud.

More importantly, in columns 5 and 6, we include all three product similarity measures in the same specification and the text-based similarity measures, both *Product Similarity (All-Firms)* and *Product Similarity (Top-15)* exhibit the strongest partial effects, both statistically and economically. These findings are consistent with our conjecture that product similarity measures

[11] To reject the concern that CEOs are incentivized to manipulate their product information changed once the Hoberg and Phillips dataset was made public, we run our main results on only data prior to 2010 and all results hold (untabulated).

based on Hoberg and Phillips (2010) capture a firm's economic fundamentals, while less susceptible to some empirical shortcomings of other similarity measures as we discussed in section 2 above.

Finally, in Panel C of Table 4 we use an alternate and novel set of dependent variables created by Alawadhi et al. (2023). These alternative proxies include a continuous predicted probability of fraud, one indicator that equals one when the predicted probability of fraud is higher than the optimal cutoff based on their model, and another indicator when the predicted probability of fraud is higher than the optimal cutoff for three consecutive firm-years. Because most control variables employed in our other tests have been incorporated in the predicted probabilities, we only include firm size, age, and industry-by-year fixed effects as controls in these tests. In contrast to observed fraud, these indicator variables have a high proportion of values equal to one (41.1% and 26.5%, respectively), making them less likely to suffer from the concerns of an unbalanced distribution of the outcome variable. In all three models using both measures of *Product Similarity,* we continue to find a large and significantly negative relation, further supporting our conclusion that firms with higher average product similarity are less likely to commit fraud.

Collectively, these findings support the notion that our results are robust to different proxies of product similarity, estimation methods, and alternative measures of fraud. In the remaining part of the paper, we explore the possible channels through which product similarity potentially serves to help reduce fraud commission and, to the best of our ability, show that our results are unlikely to be driven by alternative mechanisms. Based on the limitation of *Industry homogeneity* and *Stock-return Co-movement* we discussed in Section 2 we focus our analyses on the text-based *Product Similarity* proxies as we view them to best capture the relevant information in peer-firm comparison.

IV. Product Similarity and Benchmarking

Disclosures from closely related peers should enhance a firm's information environment (see, e.g., Badertscher, Shroff, and White (2013), Kim, Kraft, and Ryan (2013), Hsu, Li, Ma, and Phillips (2017), and Engelberg, Ozuguz, and Wang (2018)), thus improving external monitoring. Therefore, we expect that a firms' information environment improves when a firm has more similar peers, and as a result, exhibits lower information asymmetry.

H2-a: Ceteris paribus, a firm's product similarity is negatively associated with information asymmetry.

Furthermore, when a peer firm conducts an IPO, the new peer disclosures should provide a discrete increase in publicly available information about a firm's economic fundamentals, which can help bolster external monitors' ability to assess a focal firm's economic circumstances (see Holmstrom (1982) and Nalebuff and Stiglitz (1983)).[12] If peers undergoing an IPO have similar products to a focal firm that are greater than the focal firm's existing product similarity, the IPO disclosures should offer a positive shock to overall product similarity and information environment accordingly.

H2-b: Ceteris paribus, the changes of a firm's product similarity due to firms conducting IPOs are negatively associated with information asymmetry.

[12] Additionally, because the peers were already operating in the same product space prior to the IPO, such events should primarily be a shock to information rather than a change to competition.

In Panel A of Table 5, we first report the relation between information asymmetry and product similarity using Fama-MacBeth regressions. Following the literature, we measure information asymmetry using the average yearly bid-ask spread and average yearly multi-market information asymmetry (Johnson and So, 2018). We find that firms with higher product similarity, using both *Product Similarity (All-Firms)* and *Product Similarity (Top-15)*, exhibit a lower level of information asymmetry, indicating a more transparent information environment.

[Insert Table 5]

Next, we define an indicator variable, *Information Shock*, which takes value of 1 if the average similarity score of IPO firms is higher than a firm's average similarity score. In Panel B of Table 5, we report the OLS estimation of information asymmetry on product similarity following a positive information shock and find that these increases in product similarity have an associated decline in information asymmetry for both of our proxies.

If product similarity facilitates better information about a firm's economic circumstances, the negative relation between product similarity and financial fraud is, therefore, likely driven by a benchmarking channel and the impact of product similarity should be greater when the marginal value of public information is higher. Based on prior work, we examine how product similarity impacts simple firms versus complex firms whose disclosures can be more multifaceted, and therefore, more difficult to monitor. Supporting our contention, Bushman et al. (2004) show that firm complexity, measured via product line diversification, is associated with more costly governance. Similarly, Cohen, and Lou (2012) argue that firms with multiple operating segments require more complicated analysis to impound information into share prices. In addition, Peterson (2012) and Hoitash and Hoitash (2018) find consistent empirical evidence that firm complexity is associated with a greater likelihood of accounting misstatements. Thus, all else equal, the impact of product similarity should be stronger for complex firms where information should have a larger marginal effect, leading to our next hypothesis:

H3: Ceteris paribus, product similarity has a stronger negative association with financial fraud for complex firms.

We define *complexity* as the number of unique industries (using three-digit SIC codes) in which a firm operates each year by summing the number of distinct industries spanned by a firm's competitor set.[13] For example, if a firm has three peers that each operates in a different three-digit SIC code, then we consider that firm to be operating in three distinct product markets.[14] In Table 6, we split the sample based on whether *complexity* is above or below the median and then estimate our main specification for the relation between *Fraud* and *Product Similarity* separately for each group. We do not find a consistently significant relation between *Product Similarity* and the incidence of fraud for less complex firms but observe both statistically and economically significant association for more complex firms.[15] Overall, product similarity is associated with a lower incidence of fraud for complex firms, which supports Hypothesis H3.

[Insert Table 6]

[13] As discussed in Section 2.2, the TNIC better identifies a firm's close peers. However, each of the firm's SIC codes provides a clear depiction of the operating segment, which is more appropriate in determining the complexity of a firm's business operations.
[14] Our measure of complexity builds on the intuition provided by Cohen and Lou (2012) and Bushman et al. (2004) who define firm complexity using the number of segments in which a firm operates.
[15] The partial effect for the complex firms is also much larger in magnitude than that for less complex firms.

One common concern in studies of financial fraud is that measures of fraud capture the joint outcome of a firm committing fraud and being caught, making it difficult to interpret whether estimated effects are due to changes in detection or commission (see Dimmock and Gerken (2012) and Dyck et al. (2023)). The negative association we document can, therefore, be interpreted as product similarity either reducing commission or lowering detection. However, we contend that it is difficult to ascertain a plausible explanation for why product similarity decreases outsiders' ability to detect reporting manipulations. We have shown that product similarity is associated with better information environments, which suggests that similar rivals should provide more informative assessment of economic circumstances (see Tirole (2010)). If higher product similarity is unlikely to reduce detection, the observed lower rate of fraud should be interpreted as reduced commission.

One may also argue that product similarity to peer firms can shift over time, and consequently, the information environment when fraud is detected can differ from when a manager decides to commit fraud. While we contend that higher similarity increases fraud detection (conditionally) after a fraud is committed, our findings suggest a (unconditional) negative relation between product similarity and the incidence of fraud, which is consistent with our interpretation that high similarity deters managers from committing fraud in the first place.

To further substantiate our interpretation, we estimate a bivariate probit model employed by Wang and Winton (2014), which exploits timing differences between detection and commission. The model estimates the probability of commission and the probability of detection conditional on commission simultaneously. In Table IA4, the estimates suggest that product similarity is associated with a decline in fraud commission and positively related to enhanced detection which is consistent with our expectation.

V. Alternative Explanations

While our findings are consistent with the notion that product similarity helps mitigate fraud commission by increasing the probability of detection, we acknowledge the empirical challenges in ruling out all other mechanisms that could (partially) explain our results. Managerial incentives – including the decision to commit fraud – are endogenously determined by a firm's product market, information environment, and corporate governance; therefore, product similarity can correlate with fraud through various channels such as competition, board oversight, and external monitoring. In this section, we run several tests to address these concerns and provide supportive evidence that negative relation between product similarity and fraud operates through an information channel and rule out, as best as possible, the alternative channels by showing that they are unlikely to fully explain our results.[16]

A. Product Similarity and CEO Compensation

CEO compensation structure and how it relates to firm performance can affect CEO incentives. However, the effect of CEO pay structure on their incentives to commit fraud is ambiguous. On the one hand, a higher percentage of equity-based compensation can increase the CEO's incentive to commit fraud because inflated valuation would increase their personal financial gains (see Davidson (2022)). A higher equity-based compensation can also induce managers to conceal negative news about future growth options and to choose suboptimal investment policies to support this pretense (see Benmelech et al. (2010).) On the other hand, equity-based compensation could disincentivize CEOs from committing fraud. When they have a higher equity stake of the firm, the expected costs of fraud are disproportionately greater for

[16] We present the result using *Product Similarity (Top-15);* in untabulated results, we confirm our findings holds with *Product Similarity (All-Firms)*.

these managers due to a large loss in market capitalization after the revelation of fraud (see Karpoff et al. (2008)) Additionally, CEO compensation structure is an outcome of a firm's internal governance, which can relate to a firm's product space. Therefore, the fraud rate can be an equilibrium outcome of these trade-offs.

[Insert Table 7]

To account for CEO compensation structure and its effects on managerial incentives, we first use the ratio of stock grants and options to total compensation.[17] Following the literature, we also include CEO pay-performance sensitivity (Delta) and CEO wealth to stock volatility sensitivity (Vega) to directly measure CEO incentives (see Coles, Daniels, and Naveen (2006)). Lastly, we control for whether CEOs are "Longholders" (see Malmendier and Tate (2005)), which are CEOs holding more of their stock and options than they otherwise should if they value diversification. As shown in Panel A of Table 7, CEO compensation, compensation structure, and pay for performance sensitivity do not fully explain our findings. Further, product similarity continues to exhibit a strong negative association with fraud. Taken together, these results continue to suggest that product similarity does not operate through a governance channel, but instead, by substantially increasing the expected costs of fraud and changing managers' tradeoff calculation in committing fraud.

B. Product Similarity and Competition

One potential motivation for financial misreporting is the pressure to raise capital when facing intense competition (see Shleifer (2004)). However, the sales-based HHI based on a firm's primary three-digit SIC code – a measure of competition that is amongst the most widely

[17] Our sample size drops because many firms are not in the Execucomp database.

used in academic research – does not show a consistent effect on fraud. In contrast, we observe a robust negative relation between product similarity and fraud, which is not consistent with managers committing more fraud due to intense competition. To further illustrate that product similarity is unlikely to proxy for the effect of competition on fraud, we show that controlling for additional measures of competition have no impact on our main results.

Specifically, column 1 in Panel B of Table 7 shows that including the sales-based HHI based on a firm's TNIC industry classification has virtually no impact on our results. In column 2 we also show that including market fluidity -- a measure of how intensively the product market around a firm is changing -- has virtually no impact on our results, suggesting that product similarity affects managerial incentives different from the dynamics and change of the product market. In column 3, we include another measure of competition based on the occurrence of competition-related words in a firm's 10-K measure (see Li, Lundholm, and Minnis (2013)) and show that product similarity exhibits a consistent negative effect on fraud.

One could also argue that the negative relation between product similarity and fraud can reflect reduced managerial slack (see Machlup (1967)). For example, competition can cause managers to exert more effort by diminishing the benefits of shirking (see Hart (1983)) and reducing resources available for rent extraction (see, e.g., Nickell (1996) and Ades and Di Tella (1999)). Extending this concept to financial fraud, competition potentially reduces the perceived benefits, such as extracting economic profits through reporting manipulations.[18] Tariff reductions have been shown to directly affect foreign peer firms' ability to offer competitive prices and increase the intensity of foreign competition (see Fresard (2010)), which can ultimately reduce

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

managerial slack (see Hart (1983)). Importantly, competition from foreign peers does not directly affect firm's information environment,[19] providing a useful setting to analyze the intermediate effects of changes in competition that are likely independent of the benchmarking channels.

Following the literature, industry tariff rates are calculated as duties collected by U.S. Customs divided by the value of U.S. imports for consumption and the values are aggregated from ten-digit U.S. Harmonized System codes to each three-digit SIC, using the concordance table provided by Pierce and Schott (2012).[20] We then construct an indicator variable, *Big Tariff Reduction*, that takes the value of 1 if the 4-year percentage change in the tariff rate is the bottom tercile, and 0 otherwise. As shown in column 4 of Panel B of Table 7, the estimated coefficient for *Product Similarity (Top-15)* is consistent with the magnitude found throughout our other analyses, suggesting that our results are not likely driven by competition reducing managerial slack.

C. Product Similarity and Governance

Prior research has suggested that better internal governance reduces the commission of fraud (see Beasley (1996) and Ege (2015)). For example, in equilibrium, higher product similarity could suggest the firm operates in a more competitive product space, and therefore needs to have a higher quality board of directors. To ensure our results are not driven by the relation between internal governance and fraud, we confirm that the association between product similarity and the incidence of fraud is robust to the inherent quality of firm governance.

[19] Tariffs do not directly affect the quality or quantity of available information through financial disclosures in 10-Ks.
[20] We thank Chotibhak Jotikasthira for kindly sharing the methodology to calculate tariff reductions.

Internal governance quality is greatly influenced by the board of directors due to their role in selecting a CEO, designing their compensation structure, and approving major corporate decisions. We proxy for board quality based on the percentage of independent directors and whether the CEO is also the chair of the board. Column 1(2) of Panel C of Table 7 presents conditional logit regressions including Independent Director% (CEO-Chairman) as an additional control variable. As shown in the table, *Product Similarity (Top-15)* continues to show a significantly negative relation with fraud. While a more independent board is also shown to discipline managers, CEO-chairman duality does not appear to significantly affect the fraud rate.

External parties also can serve an important monitoring role. Dyck et al. (2010) suggest that institutional investors, analysts, and auditors are more important than internal governance actors for detecting fraud.[21] As it is with internal governance, one might be concerned that a better information environment from enhanced benchmarking correlates with better external governance, which can also reduce fraud.

Institutional owners are key external governance participants and as the legal owners of the firm can influence board composition and governance. We use the number of institutional blockholders and the concentration of institutional blockholders as proxies for stronger institutional monitoring (Hartzell and Starks, 2003). Monitoring by institutional blockholders often emanates from engaging directly with management, rather than relying on public disclosures (see Almazan, de Motta, and Titman (2005) and McCahery, Sautner, and Starks (2016)). However, if institutional blockholders monitor firms through higher quality

[21] In their sample, only 30% of cases were initiated by internal governance actors, specifically management or directors, and even "the vast majority of [those] cases are associated with either a managerial turnover or an economic or financial crisis." The rest were initiated by external governance actors including analysts, auditors, employees, industry regulators, the media, and short sellers.

benchmarks, we expect the partial effect to attenuate. As shown in columns 3 and 4 in Panel C of Table 7, product similarity continues to exhibit a strong negative relation with fraud. Further, the number of institutional blockholders does not appear to significantly affect the fraud rate, but a more concentrated ownership structure is associated with a lower level of fraud.

Sell-side analysts are another form of external monitor who often rely on information besides public disclosures (e.g., conducting their own research or having access to management). If the information environment facilitated though product similarity overlaps with analyst coverage, the marginal effects of benchmarking could be reduced. To the contrary, using the number of analysts as another proxy for external monitoring, column 5 shows that product similarity is still significantly negatively associated with the incidence of fraud.

Research also indicates that Top-4 auditors provide higher audit quality (see Che, Hope, and Langli (2020) and Defond and Zhang (2014)). Therefore, if higher-quality auditors create better information environments that are similar to benchmarking, we expect product similarity to have a diminishing effect on fraud. Once again, column 6 shows that the partial effects from product similarity on *Fraud* are quantitatively unaffected.

In summary, our results show that product similarity augments the ability to assess a firm's economic circumstances, and this is likely different from other forms of external monitoring documented in the literature. Further controlling for the existing forms of external governance does not affect our results and product similarity continues to be associated with a lower incidence of fraud. These results suggest that product similarity is an additional novel mechanism through which managerial incentives to commit fraud could be affected.

VI. Conclusion

Fraudulent financial reporting can have significant negative consequences, both financially and reputationally, for firms, managers, and their shareholders. While detected fraud comprises only two to three percent of firms per annum, the actual rates are likely much higher (see Dyck et al. (2023) and Alawadhi et al. (2023)). Consequently, understanding the factors that could influence the incentives to commit fraud has important economic implications.

We show that product similarity is associated with a reduced incidence of financial fraud. Our evidence is consistent with the notion that product similarity helps enhance the information environment and facilitate monitoring through better benchmarking. In turn, managers have lower incentives to commit fraud ex ante. This factor has explanatory power beyond rent reduction from competition and better internal and external governance controls.

The negative association of product similarity with fraud is economically large, and thus our findings illustrate another important predictor of fraud to reduce omitted variable concerns for fraud prediction models. Finally, because regulators have limited resources to detect fraud, our findings that greater product similarity helps create more market discipline may offer insight into more efficient resource allocation in fraud detection.

Appendix. Variable Definition

Variable	Definitions
Fraud	Equals 1 for firm-years for which firms have settled with the Securities and Exchange Commission, the Department of Justice, and securities class action lawsuits for corporate fraud. Note: This is not the actual settlement year, which is often a at least few years after the alleged fraud occurred, but the year in which the fraud allegedly occurred.
Product Similarity (All-Firms)	Average Pairwise Similarity Score (see Hoberg and Phillips (2010)) for all peers within each firm-year
Product Similarity (Top-15)	Average Pairwise Similarity Score (see Hoberg and Phillips (2010)) for the 15 peers with the highest similarity score within each firm-year
Industry Homogeneity	The correlation between common stock returns within two-digits SIC industries (see Parrino (1997))
Stock-return Co-movement	For each firm, we then calculate a simple average of pairwise stock return correlation among 15 firms with which the firm has highest correlations, estimated using a rolling-60-month pairwise monthly stock return correlation among all firms.
Predicted Prob of Fraud	From the Misrepresentation Prediction Model (Logistic Regression) developed by Alawadhi et al. (2023)
Predicted Prob > optimal cutoff	From the Misrepresentation Prediction Model (Logistic Regression) developed by Alawadhi et al. (2023) with cutoff
3-consec.-year Prob > optimal cutoff	From the Misrepresentation Prediction Model (Logistic Regression) developed by Alawadhi et al. (2023)
HHI (SIC3)	Herfindahl-Hirschman Index calculated based on Sales for SIC3 Industries
Change in Gross Margin	[(Sales(t-1) - COGS (t-1))/Sales (t-1)] / [(Sales (t) - COGS (t))/Sales (t)]
Ln (Age)	Natural Log of the number of years the firm has been in Compustat
Size (Ln MktCap)	Natural Log of the Total market capitalization as of fiscal year-end
Sales Growth	[Sales(t) - Sales(t-1)]/Sales(t-1)
Security Issue	Dummy equals 1 if the firm issued securities during the year.
Geographic Segments	Log (# of Geographic Segments) set to 1 when missing
Auditor Opinion	Dummy equals 1 if Auditor's Opinion variable (AUOP) is one of the following: ("0", "2", "3", "5").
Business Equipment	From Fama French 10-industry portfolio definitions.
Telecom Industry	From Fama French 10-industry portfolio definitions.
Soft Asset	(Total Assets - PP&E- Cash and Cash Equivalent)/Total Assets
Operating Lease	Dummy equals 1 if future operating lease obligations are greater than zero.
Losses	Dummy equals 1 if net income is negative.
BigN Auditor	Dummy equals 1 if Auditor variable (AU) is one of the following: ("1", "2", "3", "4", "5", "6", "7", "8").
Bid-Ask Spread	Average bid-ask spread for the previous fiscal year
MultiMarket Information Asymmetry	An information asymmetry measure based on looking at abnormal volume in options and stock markets (see Johnson and So (2018))
Percentage Equity Compensation	(tdc2 - total_curr)/tdc2 where tdc = Total Compensation (Salary + Bonus + Other Annual + Restricted Stock Grants + LTIP Payouts + All Other + Value of Options Exercised). And total_curr = SALARY + BONUS.

Delta	Dollar changes in wealth associated with a 1% change in the firm's stock price (in $000s) (see Coles et al. (2006))
Vega	Dollar changes in wealth associated with a 0.01 change in the standard deviation of the firm's returns (in $000s) (see Coles et al. (2006))
Longholder	Dummy equals 1 if a CEO (for all their years in the sample), ever holds an option until the last year of its duration and zero otherwise (see Malmendier and Tate (2005))
Big Tariff Reduction	Dummy equals 1 if the 4-year percentage change in the tariff rate of the firms in the 3-digit SIC code is in the bottom tercile

References

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

Alawadhi, A.; J. M. Karpoff; J. L. Koski; and G. S. Martin. “The prevalence and price distorting effects of undetected financial misrepresentation: Empirical evidence.” Working paper, (2023)

Allen, E. J.; C. R. Larson, and R. G. Sloan. “Accrual reversals, earnings and stock returns.” *Journal of Accounting and Economics* 56 (2013), 113-129.

Almazan, A.; A. de Motta; and S. Titman. “Debt, labor markets, and the creation and destruction of firms.” *Journal of Financial Economics*, 118(3) (2015), 636–657.

Armour, J.; C. Mayer; and A. Polo. “Regulatory Sanctions and Reputational Damage in Financial Markets.” *The Journal of Financial and Quantitative Analysis*, 52(4) (2017), 1429–1448.

Badertscher, B.; N. Shroff; and H. D. White. “Externalities of public firm presence: Evidence from private firms’ investment decisions.” *Journal of Financial Economics*, 109(3) (2013), 682–706.

Barberis, N.; A. Shleifer; and J. Wurgler. “Comovement.” *Journal of Financial Economics*, 75(2) (2005), 283-317.

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

Becker, G. S. “Crime and punishment: An economic approach.” *Journal of Political Economy*, 76(2) (1968), 169-217.

Benmelech, E.; E. Kandel; and P. Veronesi. “Stock-based compensation and CEO (dis) incentives.” *The Quarterly Journal of Economics*, 125(4) (2010), 1769-1820.

Bushman, R.; Q. Chen; E. Engel; and A. Smith. “Financial accounting information, organizational complexity and corporate governance systems.” *Journal of Accounting and Economics*, 37(2) (2004), 167–201.

Che, L.; O. K. Hope; and J. C. Langli. “How Big-4 Firms Improve Audit Quality.” *Management Science*, 66(10) (2020), 4552–4572.

Choi, S.; K. Nelson; and A. Pritchard. “The Screening Effect of the Private Securities Litigation Reform Act.” *Journal of Empirical Legal Studies*, 6(1) (2009), 35–68.

Cohen, L., and D. Lou. “Complicated firms.” *Journal of Financial Economics*, 104(2) (2012), 383–400.

Coles, J.; N. Daniel; and L. Naveen. “Managerial incentives and risk-taking.” *Journal of Financial Economics* 79 (2006) 431-468.

Core, J. and W. Guay. “Estimating the value of employee stock option portfolios and their sensitivities to price and volatility.” *Journal of Accounting Research* 40 (2002), 613-630.

Davidson, R. H. “Who did it matters: Executive equity compensation and financial reporting fraud.” *Journal of Accounting and Economics*, 73(2-3) (2022), 101453.

de Bodt, E.; E. Eckbo; and R. Roll. “Competition Shocks, Rival Reactions, and Stock Return Comovement.” *Journal of Financial and Quantitative Analysis*, (2024) forthcoming

De Franco, G.; S. P. Kothari; and R. S. Verdi. “The Benefits of Financial Statement Comparability.” *Journal of Accounting Research*, 49(4) (2011), 895–931.

Dechow, P. M., and I. Dichev. “The Quality of Accruals and Earnings: The Role of Accrual Estimation Errors.” *The Accounting Review*, 77 (2002), 35–59.

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

DeFond, M., and J. Zhang. "A review of archival auditing research." *Journal of Accounting and Economics*, 58(2) (2014), 275–326.

Dimmock, S. G., and W. Gerken. "Predicting fraud by investment managers." *Journal of Financial Economics*, 105(1) (2012), 153–173.

Donelson, D. C.; A. Kartapanis; J. McInnis; and C. Yust. "Measuring Accounting Fraud and Irregularities Using Public and Private Enforcement." *The Accounting Review*, 96(6) (2021), 183–213.

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

Dyck, A.; A. Morse; and L. Zingales. "How pervasive is corporate fraud?" *Review of Accounting Studies*. (2023)

Ege, M. S. "Does Internal Audit Function Quality Deter Management Misconduct?" *The Accounting Review*, 90(2) (2015), 495–527.

Engelberg, J.; A. Ozoguz; and S. Wang. "Know Thy Neighbor: Industry Clusters, Information Spillovers, and Market Efficiency." *Journal of Financial and Quantitative Analysis*, 53(5) (2018), 1937–1961.

Fresard, L. "Financial Strength and Product Market Behavior: The Real Effects of Corporate Cash Holdings." *The Journal of Finance*, 65(3) (2010), 1097–1122.

Goldman, E. and S. L. Slezak. "An equilibrium model of incentive contracts in the presence of information manipulation." *Journal of Financial Economics*, 80(3) (2006), pp.603-626.

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

Hameed, A.; R. Morck; J. Shen; and B. Yeung. "Information, Analysts, and Stock Return Comovement." *The Review of Financial Studies*, 28(11) (2015), 3153-3187.

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

Hartzell, J. C., and L. Starks. "Institutional Investors and Executive Compensation." *The Journal of Finance*, 58(6) (2003), 2351-2374.

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

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

Hoberg, G.; G. Phillips; and N. Prabhala. "Product Market Threats, Payouts, and Financial Flexibility." *Journal of Finance*, 69 (1) (2014), 293-324.

Hoitash, R., and U. Hoitash. . "Measuring Accounting Reporting Complexity with XBRL." *The Accounting Review*, 93(1) (2018), 259–287.

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

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

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

Jones, J. “Earnings Management During Import Relief Investigations.” *Journal of Accounting Research*, 29(2) (1991), 193–228.

Johnson, S. A.; H. Ryan; and Y. Tian. “Managerial incentives and corporate fraud: The sources of incentives matter.” *Review of Finance*, *13*(1) (2009), 115-145.

Johnson, T. L., and E. So. “A Simple Multimarket Measure of Information Asymmetry.” *Management Science*, 64(3) (2018), 1055–1080.

Karpoff, J. M. “On a stakeholder model of corporate governance.” *Financial Management*, 50(2) (2021), 321–343.

Karpoff, J. M.; A. Koester; D. Lee; and G. Martin. “Proxies and Databases in Financial Misconduct Research.” *The Accounting Review*, 92(6) (2017), 129–163.

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

Karpoff, J. M.; D. Lee; and G. Martin. “The consequences to managers for financial misrepresentation.” *Journal of Financial Economics*, 88(2) (2008), 193–215.

Khanna, V.; E. Kim; and Y. Lu. “CEO connectedness and corporate fraud.” *The Journal of Finance*, *70*(3) (2015), 1203-1252.

Kim, S.; P. Kraft.; and S. Ryan. “Financial statement comparability and credit risk.” *Review of Accounting Studies*, 18(3) (2013), 783–823.

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

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

Malmendier, U., and G. Tate. “CEO Overconfidence and Corporate Investment.” *The Journal of Finance*, 60(6) (2005), 2661–2700.

McCahery, J. A.; A. Sautner; and L. Starks. “Behind the Scenes: The Corporate Governance Preferences of Institutional Investors.” *The Journal of Finance*, 71(6) (2016), 2905–2932.

McNichols, M. F. “Research design issues in earnings management studies.” *Journal of Accounting and Public Policy*, 19(4) (2000), 313–345.

Murphy, K. J. “Incentives, Learning, and Compensation: A Theoretical and Empirical Investigation of Managerial Labor Contracts.” *The RAND Journal of Economics*, 17(1) (1986), 59–76.

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

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

Parrino, R. “CEO Turnover and outside Succession: A Cross-Sectional Analysis.” . *Journal of Financial Economics*. (1997)

Peterson, K. . “Accounting complexity, misreporting, and the consequences of misreporting.” *Review of Accounting Studies*, 17(1) (2012), 72–95.

Pierce, J. R., and P. Schott. . “A concordance between ten-digit U.S. harmonized system codes and SIC/NAICS product classes and industries.” *Journal of Economic & Social Measurement*, 37(1/2) (2012), 61–96.

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

Schwartz-Ziv, M., and E. Volkova. “Is Blockholder Diversity Detrimental?” *Management Science*. (2024)

Shleifer, Andrei. “Does Competition Destroy Ethical Behavior?” *American Economic Review*, 94 (2) (2004): 414-418.

Sohn, B. C. “The effect of accounting comparability on the accrual-based and real earnings management.” *Journal of Accounting and Public Policy*, 35(5) (2016), 513–539.

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

Tirole, J. “The Theory of Industrial Organization.” *MIT Press*. (1988)

Tirole, J. “The Theory of Corporate Finance.” *Princeton University Press*. (2010)

Wang, T. Y., and A. Winton. “Product market interactions and corporate fraud.” Available at SSRN 2398035. (2014)

Figure 1. Product Similarity Measures

This figure presents a scatter plot exhibiting the relation between the average text-based product similarity and Industry Homogeneity (see Parrino, 1997), and stock-return co-movement by industry.

Figure 2. Product Similarity and Fraud

This figure presents binned scatter plots exhibiting the relation between fraud and product similarity after controlling for firm size and age.

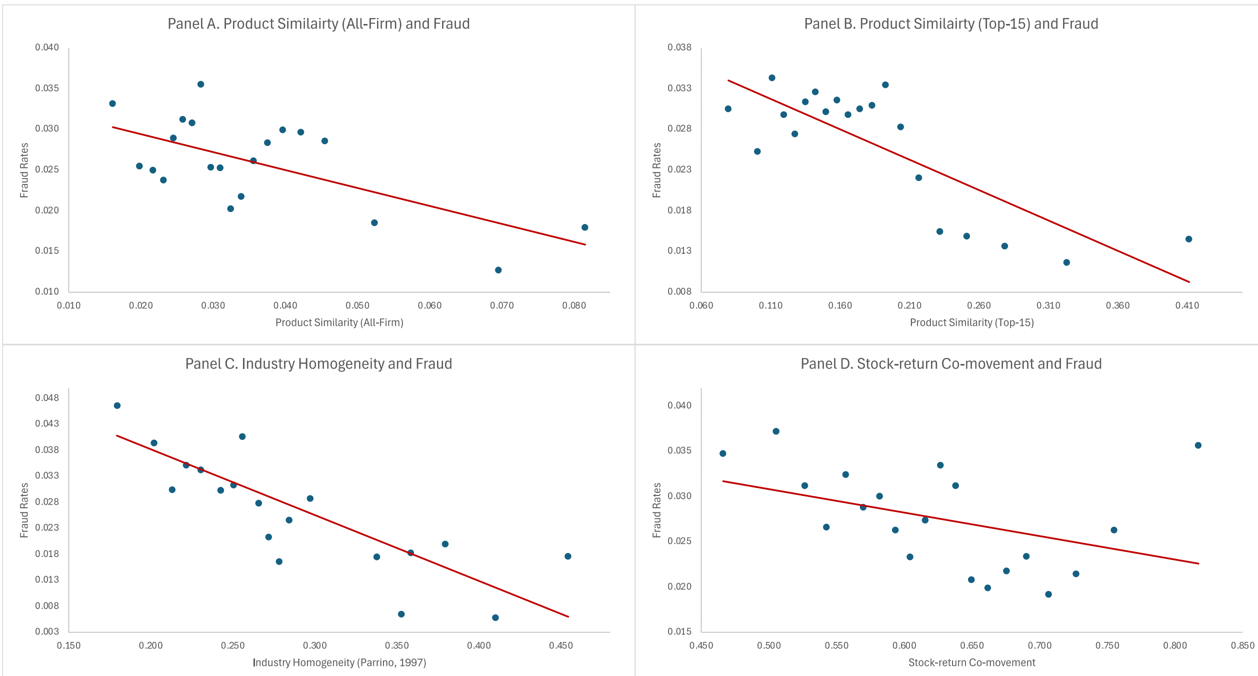


Figure 3. Product Similarity and Fraud (Independent Cross-Sectional Regressions)

This figure plots coefficient estimates independent cross-sectional linear regressions of fraud on product similarity for each year for both the *All-Firm* and *Top-15* measures of product similarity. The coefficient estimates for each year are obtained using the control variables from our main model specification (columns 3 and 5) in Table 2 and TNIC fixed effects.

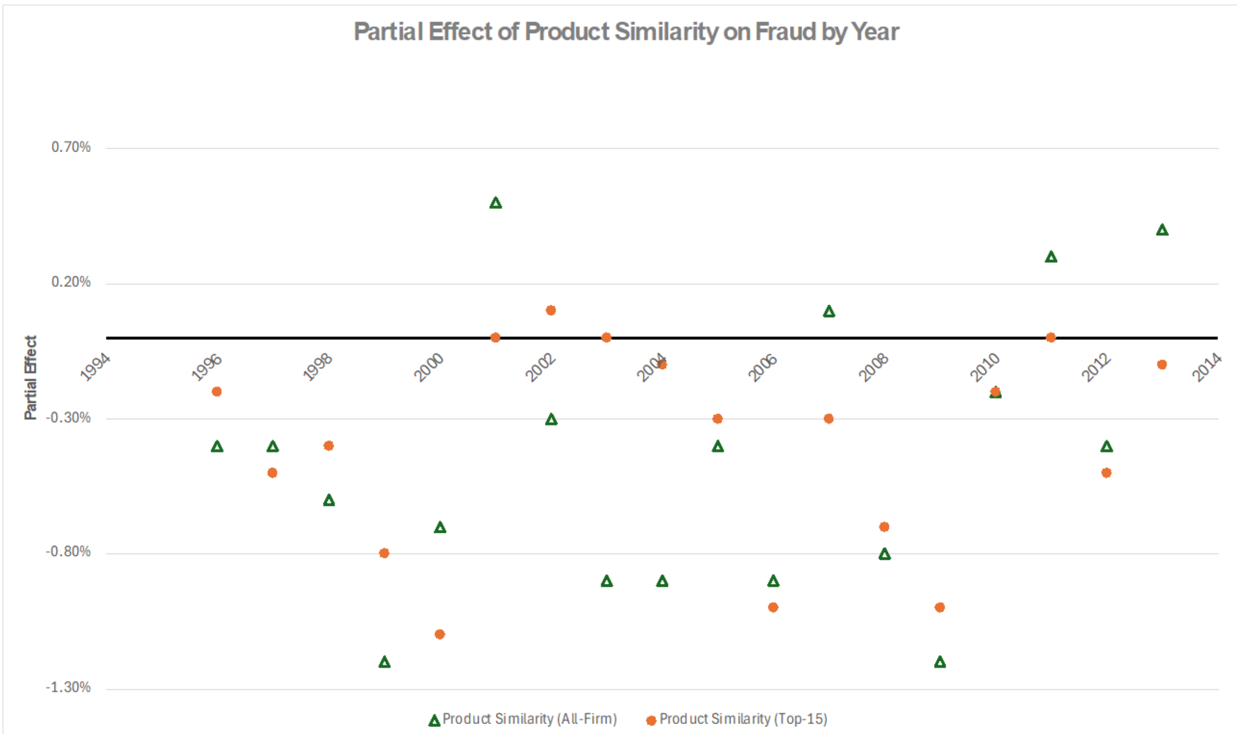


Table 1. Summary Statistics

Panel A reports summary statistics of firm characteristics at the firm-year level for the full sample. Variable definitions are provided in the Appendix. Our sample spans 1996 through 2014. Panel B reports paired t-tests of accounting dispersion between Top 15 peers and all peers for the full firm-year panel. Accounting dispersions are calculated as the standard deviation divided by the mean. The accounting ratios of interest include sales/assets, operating expense/asset, and unsigned abnormal accruals/sales. Unsigned abnormal accruals is defined as the absolute value of the difference between a firm's accruals (see Allen et al. (2013)) and the industry average accruals according to three-digit SIC. Statistical significance (two-sided) at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

Panel A. Summary Statistics of Firm Characteristics

	N	Mean	Std. Dev.	p10	p50	p90
Fraud	83,412	0.027	0.161	0.000	0.000	0.000
Product Similarity (All-Firms)	83,412	0.036	0.017	0.020	0.031	0.062
Product Similarity (Top-15)	83,412	0.186	0.082	0.100	0.168	0.300
Predicted Prob of Fraud	64,429	0.018	0.022	0.003	0.011	0.040
Predicted Prob > Optimal Cutoff	64,429	0.411	0.492	0.000	0.000	1.000
3-consec.-year Prob > Optimal Cutoff	64,429	0.265	0.442	0.000	0.000	1.000
Industry Homogeneity	72,115	0.286	0.071	0.202	0.263	0.394
Stock-return Co-movement	57,548	0.625	0.092	0.509	0.620	0.750
Size (Ln MktCap)	83,412	5.610	2.115	2.890	5.563	8.408
Ln (Age)	83,412	2.333	0.973	1.099	2.485	3.555
HHI (SIC3)	83,412	0.149	0.127	0.046	0.102	0.304
Change in Gross Margin	83,412	0.991	0.743	0.745	0.996	1.223
Sales Growth	83,412	0.219	0.670	-0.172	0.086	0.593
Soft Asset	83,412	0.598	0.273	0.188	0.630	0.942
Geographic Segments	83,412	1.072	0.444	0.693	1.099	1.792
Operating Lease	83,412	0.786	0.410	0.000	1.000	1.000
Security Issue	83,412	0.856	0.351	0.000	1.000	1.000
Losses	83,412	0.317	0.465	0.000	0.000	1.000
Auditor Opinion	83,412	0.001	0.022	0.000	0.000	0.000
BigN Auditor	83,412	0.737	0.440	0.000	1.000	1.000
Business Equipment	83,412	0.191	0.393	0.000	0.000	1.000
Telecom Industry	83,412	0.027	0.161	0.000	0.000	0.000
Bid-Ask Spread	79,899	1.073	1.585	0.037	0.459	2.867
MultiMarket Information Asymmetry	27,376	0.421	0.096	0.291	0.425	0.531
Delta	34,924	729.7	1,378	63.17	281.6	1,634
Vega	34,958	111.2	196.4	0.382	37.98	301.4
Longholder	22,972	0.392	0.488	0	0	1
Big Tariff Reduction	83,412	0.403	0.491	0	0	1

Panel B. Paired t-tests of Accounting Dispersion between Top-15 Peers and All-Peers

Dispersion (std. dev./ mean)	All-Firm	Top-15	Diff
Sales/Asset	0.941	0.556	0.385***
Operating Expense/Asset	0.945	0.573	0.372***
Unsigned Abnormal Accruals/Sale	4.485	1.639	2.846***

Table 2. Correlation of Product Similarity Measures

This table reports correlations between product similarity using all peers and product similarity only using the top 15 closest peers, Industry Homogeneity measure by Parrino (1997), and the stock-return co-movement measure.

	Product Similarity (All-Firms)	Product Similarity (Top 15)	Industry Homogeneity	Stock-return Co-movement
Product Similarity (All-Firms)	1.00			
Product Similarity (Top-15)	0.89***	1.00		
Industry Homogeneity	0.39***	0.46***	1.00	
Stock-return Co-movement	0.05***	0.04***	0.07***	1.00

Table 3. Product Similarity and Financial Fraud

This table reports linear-probability-model estimates for Product Similarity on the incidence of fraud. Our proxy for financial fraud includes settled SEC and DOJ enforcement actions and Securities Class Actions from the Stanford University Lawsuit Database. In columns 1, 2, and 4 our product similarity measure includes the average similarity with all firms whereas in columns 3 and 5 it only includes the average similarity of the top 15 peers. The specification in Column 1 includes only size and age as control variables and includes year fixed effects. The specifications in Columns 2-5 include controls selected from Alawadhi et al. (2023) as well as TNIC×Year fixed effects. In Columns 4 and 5 we report standardized regressions. All specifications are run at the firm-year level, and include explanatory variables lagged by one year. The t-statistics, calculated from standard errors clustered at the Firm level, are reported in parentheses below the coefficient estimates. Statistical significance (two-sided) at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

	1	2	3	4	5
	Fraud (t+1)				
	LPM(OLS)			Standardized Regression	
Product Similarity (All-Firms)	-0.184***	-0.255*		-0.004*	
	(-3.246)	(-1.934)		(-1.935)	
Product Similarity (Top-15)			-0.049**		-0.004**
			(-2.302)		(-2.302)
Size	0.011***	0.012***	0.012***	0.027***	0.027***
	(14.111)	(12.820)	(12.828)	(12.820)	(12.828)
Age	-0.002	-0.000	-0.000	-0.000	-0.000
	(-1.491)	(-0.058)	(-0.146)	(-0.058)	(-0.146)
HHI (SIC3)		0.015	0.015	0.002	0.002
		(1.451)	(1.409)	(1.451)	(1.409)
Change in Gross Margin		0.000	0.000	0.000	0.000
		(0.317)	(0.315)	(0.317)	(0.315)
Sales Growth		0.005***	0.005***	0.004***	0.004***
		(4.948)	(4.906)	(4.948)	(4.906)
Soft Asset		0.027***	0.026***	0.007***	0.007***
		(4.607)	(4.537)	(4.607)	(4.537)
Geographic Segments		0.012***	0.012***	0.005***	0.005***
		(3.213)	(3.149)	(3.213)	(3.149)
Operating Lease		-0.003	-0.003	-0.003	-0.003
		(-0.867)	(-0.859)	(-0.867)	(-0.859)
Security Issue		0.003	0.003	0.003	0.003
		(0.402)	(0.399)	(0.402)	(0.399)
Losses		0.002	0.002	0.002	0.002
		(1.155)	(1.128)	(1.155)	(1.128)
Auditor Opinion		0.042	0.042	0.042	0.042
		(0.835)	(0.843)	(0.835)	(0.843)
BigN Auditor		-0.012***	-0.012***	-0.012***	-0.012***
		(-4.495)	(-4.404)	(-4.495)	(-4.404)
Business Equipment		0.015***	0.015***	0.015***	0.015***
		(2.688)	(2.640)	(2.688)	(2.640)
Telecom		0.004	0.005	0.004	0.005
		(0.432)	(0.508)	(0.432)	(0.508)
Constant	-0.024***	-0.060***	-0.059***	0.030***	0.030***
	(-4.258)	(-5.485)	(-5.588)	(4.384)	(4.357)
Observations	83412	83297	83297	83297	83297
Adj/Pseudo R-squared	0.024	0.044	0.044	0.044	0.044
FE	Year	TNIC×Year	TNIC×Year	TNIC×Year	TNIC×Year

Table 4. Product Similarity and Financial Fraud – Alternative Specifications

Panel A reports Conditional Logit Odds Ratios in columns 1-4 using the full sample and Linear Probability Model estimates using matched sample in columns 5-6 for Product Similarity on the incidence of fraud. Our proxy for financial fraud includes settled SEC and DOJ enforcement actions and Securities Class Actions. In columns 1, 2 and 5 our measure of product similarity includes the average similarity with all firms whereas in columns 3 and 6 we only include the average similarity of the top 15 peers and in column 4 we include both measures. All specifications include controls selected from Alawadhi et al. (2023). In columns 2-4 we report standardized regressions. In Panel B, columns 1-4 report regressions with alternate proxies for product similarity -- Industry Homogeneity measure by Parrino (1997), and the stock market return co-movement. Columns 5-6 also include average similarity with all firms and only the top 15. Panel C reports Linear Probability Model estimates on the full sample but with alternate dependent variables based on Alawadhi et al. (2023) and only include size and age controls. All specifications are run at the firm-year level, include TNIC×Year fixed effects, and include explanatory variables lagged by one year. The t-statistics, calculated from standard errors clustered at the firm level, are reported in parentheses below the coefficient estimates. Statistical significance (two-sided) at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

Panel A. Regression Estimates of Main Measures of Product Similarity on the Incidence of Fraud

	1	2	3	4	5	6
	Fraud (t+1)					
	C-logit	Standardized C-logit			Matched Sample OLS	
Product Similarity (All-Firms)	0.000***	0.830***		1.035	-1.318**	
	(-2.998)	(-2.998)		(0.376)	(-2.028)	
Product Similarity (Top-15)			0.778***	0.761***		-0.250**
			(-4.453)	(-3.274)		(-2.280)
Size	1.554***	2.608***	2.632***	2.631***	0.031***	0.031***
	(29.138)	(29.138)	(29.126)	(29.197)	(9.941)	(9.890)
Age	0.928**	0.931**	0.925**	0.926**	-0.006	-0.006
	(-2.284)	(-2.284)	(-2.482)	(-2.469)	(-1.008)	(-1.005)
HHI (SIC3)	1.771**	1.076**	1.069**	1.069**	0.056	0.053
	(2.549)	(2.549)	(2.345)	(2.350)	(1.311)	(1.248)
Change in Gross Margin	1.034	1.026	1.026	1.026	-0.000	0.000
	(1.043)	(1.043)	(1.052)	(1.052)	(-0.001)	(0.008)
Sales Growth	1.228***	1.154***	1.153***	1.153***	0.029***	0.029***
	(6.023)	(6.023)	(6.043)	(6.032)	(5.338)	(5.344)
Soft Asset	3.226***	1.383***	1.373***	1.373***	0.093***	0.090***
	(8.159)	(8.159)	(8.046)	(8.024)	(3.980)	(3.874)
Geographic Segments	1.378***	1.151***	1.147***	1.147***	0.027**	0.027**
	(5.313)	(5.313)	(5.181)	(5.171)	(2.179)	(2.138)
Operating Lease	0.906	1.151***	1.147***	1.147***	-0.019	-0.018
	(-1.027)	(5.313)	(5.181)	(5.171)	(-0.938)	(-0.905)
Security Issue	1.258	1.258	1.239	1.237	0.006	0.006
	(1.109)	(1.109)	(1.014)	(1.006)	(0.222)	(0.208)
Losses	1.000	1.000	1.001	1.000	-0.006	-0.005
	(-0.004)	(-0.004)	(0.013)	(-0.001)	(-0.652)	(-0.625)
Auditor Opinion	3.633**	3.633**	3.683**	3.677**	0.279*	0.276*
	(2.146)	(2.146)	(2.208)	(2.210)	(1.680)	(1.668)
BigN Auditor	0.782***	0.782***	0.790***	0.791***	-0.034***	-0.033***
	(-2.820)	(-2.820)	(-2.715)	(-2.700)	(-2.697)	(-2.633)
Business Equipment	1.585***	1.585***	1.568***	1.563***	0.016	0.015
	(4.975)	(4.975)	(4.933)	(4.860)	(0.806)	(0.752)
Telecom	1.068	1.068	1.109	1.114	0.010	0.013
	(0.288)	(0.288)	(0.465)	(0.482)	(0.349)	(0.448)
Observations	49107	49107	49107	49107	17819	17819
Adj/Pseudo R-squared	0.117	0.117	0.118	0.118	0.075	0.075
TNIC×Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Panel B. Regression Estimates of Alternative Measures of Product Similarity on the Incidence of Fraud

	1	2	3	4	5	6
	Fraud (t+1)					
	Matched Sample		C-logit			
Industry Homogeneity	-0.243*		-1.683*		-0.975	-0.983
	(-1.862)		(-1.943)		(-1.020)	(-1.031)
Stock-return Co-movement		-0.106*		-0.651	-0.361	-0.346
		(-1.674)		(-1.519)	(-0.790)	(-0.755)
Product Similarity (All-Firms)					-10.781**	
					(-2.140)	
Product Similarity (Top-15)						-1.986**
						(-2.268)
Size	0.032***	0.030***	0.446***	0.438***	0.447***	0.448***
	(9.537)	(8.086)	(27.254)	(22.790)	(22.319)	(22.067)
Age	-0.003	0.006	-0.064*	-0.122**	-0.136**	-0.139**
	(-0.517)	(0.569)	(-1.810)	(-2.306)	(-2.433)	(-2.465)
HHI (SIC3)	0.035	0.012	0.403	0.394	0.367	0.362
	(0.706)	(0.259)	(1.555)	(1.508)	(1.291)	(1.277)
Change in Gross Margin	0.002	-0.003	0.049	0.010	0.025	0.025
	(0.539)	(-0.762)	(1.361)	(0.225)	(0.512)	(0.509)
Sales Growth	0.027***	0.027***	0.191***	0.275***	0.284***	0.283***
	(4.701)	(3.406)	(5.224)	(4.901)	(4.619)	(4.625)
Soft Asset	0.110***	0.120***	1.327***	1.585***	1.645***	1.630***
	(4.146)	(4.112)	(8.397)	(8.836)	(8.467)	(8.381)
Geographic Segments	0.035***	0.032**	0.399***	0.364***	0.488***	0.485***
	(2.654)	(2.185)	(6.033)	(4.901)	(6.130)	(6.055)
Operating Lease	-0.030	-0.018	-0.241**	-0.177	-0.313***	-0.310***
	(-1.377)	(-0.729)	(-2.318)	(-1.559)	(-2.798)	(-2.757)
Security Issue	0.012	-0.009	0.314	0.037	0.174	0.165
	(0.392)	(-0.278)	(1.446)	(0.158)	(0.776)	(0.726)
Losses	-0.005	-0.004	-0.028	0.021	0.024	0.023
	(-0.601)	(-0.354)	(-0.402)	(0.267)	(0.292)	(0.279)
Auditor Opinion	0.284*	0.280	1.468**	1.398	1.915*	1.913*
	(1.789)	(1.555)	(2.331)	(1.549)	(1.740)	(1.770)
BigN Auditor	-0.032**	-0.020	-0.232**	0.047	-0.028	-0.018
	(-2.334)	(-1.465)	(-2.375)	(0.394)	(-0.228)	(-0.148)
Business Equipment	0.010	0.020	0.454***	0.506***	0.459***	0.443***
	(0.425)	(0.885)	(4.229)	(4.578)	(3.658)	(3.565)
Telecom	0.012	0.037	0.020	0.428	0.236	0.272
	(0.388)	(1.323)	(0.078)	(1.555)	(0.739)	(0.860)
Observations	15742	12856	43474	29859	27059	27059
Adj/Pseudo R-squared	0.081	0.083	0.130	0.136	0.147	0.147
TNIC×Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Panel C. Regression Estimates of Main Measures of Product Similarity on Alternative Fraud Measures

	1	2	3	4	5	6
	Predicted Prob of Fraud		Predicted Prob > optimal cutoff		3-consecutive-year Pred Prob > optimal cutoff	
Product Similarity (All-Firms)	-0.058***		-1.420***		-0.989***	
	(-3.062)		(-3.713)		(-2.656)	
Product Similarity (Top-15)		-0.015***		-0.376***		-0.197***
		(-5.094)		(-6.471)		(-3.508)
Size	0.006***	0.006***	0.127***	0.127***	0.113***	0.113***
	(39.272)	(39.148)	(99.826)	(99.830)	(83.025)	(82.631)
	0.001***	0.001***	0.006**	0.005*	0.060***	0.059***
Age	(6.856)	(6.555)	(2.189)	(1.670)	(21.413)	(21.117)
	-0.018***	-0.017***	-0.275***	-0.259***	-0.469***	-0.468***
	(-15.834)	(-16.715)	(-17.413)	(-19.271)	(-30.175)	(-35.548)
Constant	0.006***	0.006***	0.127***	0.127***	0.113***	0.113***
	(39.272)	(39.148)	(99.826)	(99.830)	(83.025)	(82.631)
Observations	66541	66541	66541	66541	66541	66541
Adjusted R-squared	0.493	0.494	0.468	0.469	0.454	0.454
FE	TNIC×Year	TNIC×Year	TNIC×Year	TNIC×Year	TNIC×Year	TNIC×Year

Table 5. Product Similarity and Measures of Information Environment

Panel A reports Fama-Macbeth estimates for product similarity on a firm's information asymmetry and Panel B displays the OLS estimates of a shock to product similarity on a firm's information asymmetry. Our proxies for information asymmetry are the annual average of daily bid-ask spread and MIA (see Johnson and So (2018)). Both Panels include controls selected from Alawadhi et al. (2023). All specifications are run at the firm-year level and include explanatory variables lagged by one year. The t-statistics, calculated from standard errors clustered at the firm level, are reported in parentheses below the coefficient estimates. Statistical significance (two-sided) at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

Panel A. Regression Estimates of Product Similarity on Information Asymmetry Measures

	1	2	3	4
	Bid-Ask Spread	MIA	Bid-Ask Spread	MIA
Product Similarity (All-Firms)	-5.055**	-0.180*		
	(-2.674)	(-2.066)		
Product Similarity (Top-15)			-0.780**	-0.018
			(-2.598)	(-1.380)
Size	-0.348***	-0.023***	-0.350***	-0.024***
	(-7.848)	(-13.462)	(-7.804)	(-13.685)
Age	0.087***	-0.002*	0.092***	-0.002
	(7.807)	(-1.886)	(7.504)	(-1.549)
HHI (SIC3)	0.131***	-0.015**	0.142***	-0.015**
	(5.548)	(-2.740)	(5.160)	(-2.596)
Change in Gross Margin	-0.023***	0.003***	-0.023***	0.003***
	(-3.999)	(3.213)	(-3.924)	(3.224)
Sales Growth	-0.042***	-0.011***	-0.043***	-0.011***
	(-6.073)	(-9.362)	(-6.262)	(-9.441)
Soft Asset	0.223***	0.039***	0.189***	0.038***
	(6.461)	(16.672)	(5.755)	(14.931)
Geographic Segments	0.042*	-0.008***	0.045	-0.007***
	(1.873)	(-2.989)	(1.735)	(-2.976)
Operating Lease	-0.078***	-0.015***	-0.068**	-0.015***
	(-3.068)	(-6.478)	(-2.231)	(-6.717)
Security Issue	-0.092***	-0.012	-0.101***	-0.013
	(-4.229)	(-1.190)	(-4.195)	(-1.214)
Losses	0.285***	-0.009***	0.279***	-0.009***
	(4.928)	(-4.838)	(4.954)	(-4.955)
Auditor Opinion	0.090	-0.002	0.093	-0.002
	(0.580)	(-0.310)	(0.625)	(-0.326)
BigN Auditor	-0.257***	0.016***	-0.237***	0.018***
	(-9.542)	(3.957)	(-8.254)	(4.939)
Business Equipment	-0.069*	-0.024***	-0.101**	-0.025***
	(-1.981)	(-7.133)	(-2.382)	(-6.682)
Telecom	0.127***	0.013***	0.144***	0.013***
	(3.721)	(4.203)	(3.841)	(4.355)
Constant	2.927***	0.612***	2.905***	0.608***
	(8.042)	(34.792)	(8.174)	(33.440)
Observations	79899	27376	79899	27376
Avg R-squared	0.4845	0.1948	0.4839	0.1942

Panel B. Estimates of Shock to Firms' Product Similarity and Information Asymmetry

	1	2
	Bid-Ask Spread	MIA
Information Shock	-0.037***	-0.003**
	(-3.009)	(-2.024)
Size	-0.343***	-0.024***
	(-56.282)	(-22.690)
Age	0.062***	-0.004***
	(6.984)	(-2.972)
HHI (SIC3)	0.125	-0.028**
	(1.608)	(-2.243)
Change in Gross Margin	-0.014*	0.002**
	(-1.950)	(2.520)
Sales Growth	-0.021**	-0.005***
	(-2.286)	(-5.115)
Soft Asset	0.187***	0.026***
	(4.682)	(4.270)
Geographic Segments	0.075***	-0.002
	(3.946)	(-0.615)
Operating Lease	0.029	-0.016***
	(0.945)	(-3.166)
Security Issue	-0.128**	-0.002
	(-2.314)	(-0.188)
Losses	0.266***	-0.004*
	(16.932)	(-1.792)
Auditor Opinion	-0.045	0.012
	(-0.251)	(0.395)
BigN Auditor	-0.174***	0.015***
	(-7.083)	(3.568)
Business Equipment	-0.002	-0.008
	(-0.058)	(-1.433)
Telecom	0.176**	0.010
	(2.140)	(0.887)
Constant	2.723***	0.608***
	(39.430)	(41.482)
Observations	69334	28290
Adjusted R-squared	0.496	0.223
FE	TNIC×Year	TNIC×Year

Table 6 Product Similarity and Information Environment: Complexity

This table reports Conditional Logit estimates for the full sample for product similarity on the incidence of fraud with the sample split by above and below median complexity. Our proxy for financial fraud includes settled SEC and DOJ enforcement actions and Securities Class Actions from the Stanford University Lawsuit Database. In columns 1 and 2 the product similarity measure includes the average similarity with all firms whereas in columns 3 and 4 it only includes the average similarity of the top 15 peers. We define complexity as the number of unique SIC codes spanned by peer firms as defined by Hoberg and Phillips (2016). All specifications include the controls selected from Alawadhi et al. (2023). The t-statistics, calculated from standard errors clustered at the firm level, are reported in parentheses below the coefficient estimates. Statistical significance (two-sided) at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

	1	2	3	4
	Fraud (t+1)			
Complexity	Low	High	Low	High
Product Similarity (All-Firms)	-2.305	-18.841***		
	(-0.382)	(-3.185)		
Product Similarity (Top-15)			-1.862*	-3.861***
			(-1.810)	(-3.950)
Size	0.441***	0.440***	0.445***	0.443***
	(19.600)	(20.504)	(19.781)	(20.633)
Age	-0.172***	-0.024	-0.177***	-0.027
	(-3.683)	(-0.537)	(-3.798)	(-0.614)
HHI (SIC3)	0.283	0.859***	0.253	0.812***
	(0.737)	(2.735)	(0.658)	(2.630)
Change in Gross Margin	0.094*	-0.015	0.095*	-0.016
	(1.830)	(-0.318)	(1.853)	(-0.347)
Sales Growth	0.220***	0.171***	0.220***	0.171***
	(4.099)	(3.762)	(4.116)	(3.775)
Soft Asset	1.120***	1.163***	1.092***	1.129***
	(4.903)	(6.239)	(4.776)	(6.160)
Geographic Segments	0.416***	0.211***	0.410***	0.203**
	(4.463)	(2.661)	(4.399)	(2.560)
Operating Lease	-0.336***	0.133	-0.340***	0.113
	(-2.615)	(0.861)	(-2.675)	(0.742)
Security Issue	-0.245	0.556*	-0.267	0.556*
	(-0.876)	(1.782)	(-0.943)	(1.780)
Losses	-0.198*	0.142*	-0.192*	0.137
	(-1.885)	(1.664)	(-1.833)	(1.609)
Auditor Opinion	-9.871***	1.299*	-9.606***	1.307*
	(-16.498)	(1.753)	(-15.937)	(1.815)
BigN Auditor	-0.106	-0.373***	-0.094	-0.363***
	(-0.870)	(-2.727)	(-0.777)	(-2.693)
Business Equipment	0.577***	0.362***	0.580***	0.344***
	(3.940)	(2.980)	(3.998)	(2.910)
Telecom	-0.289	0.209	-0.260	0.251
	(-0.626)	(0.791)	(-0.566)	(0.968)
Observations	17717	21369	17717	21369
Pseudo R-squared	0.148	0.111	0.148	0.112
FE	TNIC×Year	TNIC×Year	TNIC×Year	TNIC×Year

Table 7. Alternative Explanations

This table reports Conditional Logit estimates for the full sample for product similarity on the incidence of fraud with additional controls. Our proxy for financial fraud includes settled SEC and DOJ enforcement actions and Securities Class Actions from the Stanford University Lawsuit Database. In Panel A, we add additional controls for CEO compensation measures including (1) CEO equity compensation percentage, (2) CEO "Longholders" (see Malmendier and Tate (2005)), and (3) CEO pay-performance sensitivity (Delta) and CEO wealth to stock volatility sensitivity (Vega) (see Coles et al. (2006)). In Panel B, we add additional controls for competition including (1) TNIC-based sales HHI, (2) Product Fluidity (see Hoberg, Phillips, and Prabhala (2014)), (3) 10-k based measure of competition from (see Li et al. (2012)), and (4) large year-over-year tariff reductions. In Panel C, we add additional controls for internal governance quality including (1) board independence, (2) CEO-Chairman duality and external monitoring including (3) number of institutional blockholders, (4) institutional ownership concentration (see Hartzell and Starks (2003)), (5) analyst coverage, and (6) Top-4 auditor. All specifications include the controls selected from Alawadhi et al. (2023). The t-statistics, calculated from standard errors clustered at the firm level, are reported in parentheses below the coefficient estimates. Statistical significance (two-sided) at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

Panel A. Regression Estimates of Product Similarity on the Incidence of Fraud with CEO Compensation measures

	1	2	3
		Fraud (t+1)	
Product Similarity (Top-15)	-2.103**	-2.843**	-2.418**
	(-2.234)	(-2.113)	(-2.496)
Equity Compensation %	0.155		
	(1.178)		
Longholder		-0.305***	
		(-3.991)	
Delta			0.000
			(1.106)
Vega			0.000*
			(1.798)
Observations	17719	9024	17107
Pseudo R-squared	0.092	0.111	0.089
Additional Controls	Yes	Yes	Yes
FE	TNIC×Year	TNIC×Year	TNIC×Year

Panel B. Regression Estimates of Product Similarity on the Incidence of Fraud with Additional Competition Controls

	1	2	3	4
		Fraud (t+1)		
Product Similarity (Top-15)	-3.356***	-3.422***	-1.863**	-3.029***
	(-4.247)	(-4.558)	(-1.971)	(-4.437)
HHI (TNIC)	-0.124			
	(-0.806)			
Product Fluidity		0.013		
		(1.078)		
Competition 10K			0.090	
			(1.288)	
Big Tariff Reduction				0.024
				(0.283)
Observations	49081	48685	26932	49107
Pseudo R-squared	0.118	0.118	0.112	0.118
Additional Controls	Yes	Yes	Yes	Yes
FE	TNIC×Year	TNIC×Year	TNIC×Year	TNIC×Year

Panel C. Regression Estimates of Product Similarity on the Incidence of Fraud with Additional Governance Controls

	1	2	3	4	5	6
			Fraud (t+1)			
Product Similarity (Top-15)	-2.014**	-2.052**	-3.053***	-3.270***	-3.506***	-3.038***
	(-2.172)	(-2.174)	(-4.469)	(-4.288)	(-4.743)	(-4.456)
Independent Director %	-1.074***					
	(-4.811)					
CEO-Chairman Duality		-0.007				
		(-0.091)				
Num Institution Block			0.009			
			(0.504)			
Institutional Ownership Concentration				-0.610***		
				(-2.659)		
Num Analysts					0.006	
					(1.163)	
Top-4 Auditor						-0.005
						(-0.052)
Observations	23015	23187	49107	37277	39264	49107
Pseudo R-squared	0.085	0.081	0.118	0.115	0.110	0.118
Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes
TNIC×Year FE	Yes	Yes	Yes	Yes	Yes	Yes

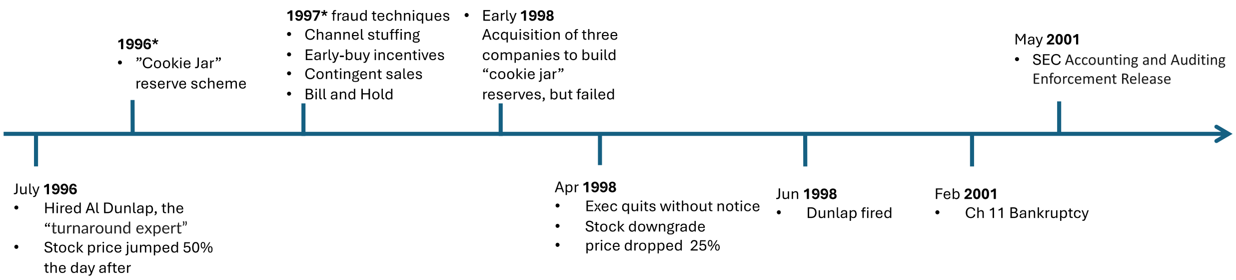
Internet Appendix for
Product Similarity, Benchmarking, and Corporate Fraud

Additional Variable Definitions

Variable	Definitions
Abnormal Industry Litigation	Yearly deviation from the industry average litigation intensity measured by the ln (MktCap) of the litigated firms in the industry in a given year as in Wang and Winton (2014).
Abnormal Return Volatility	The difference between the yearly standard deviation of a firm's stock returns and its time-series average as in Wang and Winton (2014).
Abnormal Stock Turnover	The difference between the monthly share turnover from the firm's time-series average as in Wang and Winton (2014).
Disastrous Stock Return	Dummy equals 1 if annual return in Compustat is in the bottom 10% of all firms as in Wang and Winton (2014).
Unsigned Abnormal Accruals/Sale	the absolute value of the difference between a firm's accruals (see Allen et al. (2013)) and the industry average accruals according to three-digit SIC.

Figure IA1. An Example of the Timeline of Fraud

This figure presents an example of the fraud timeline including the time when a fraud is committed, is uncovered, and is resolved. Sunbeam Corp. hired Albert Dunlap, a "turnaround expert," as CEO who made several questionable accounting decisions to reduce earnings in 1996 (the year he was hired) and increase earnings in 1997 to inflate growth in hopes of selling the firm. The scheme unraveled quickly in 1998, and the firm filed for bankruptcy in 2001. The SEC announced an AAER in 2001 for misstating earnings in 1996 and 1997. In our dataset 1996 and 1997 would be considered the fraudulent firm-years.



*1996 and 1997 were indicated as fraud years for Sunbeam in our dataset.

Table IA1. Additional Summary Statistics

Panel A of this table provides summary statistics of firm characteristics at the firm-year level for the full sample. Variable definitions are provided in the Internet Appendix. Our sample spans 1996 through 2014. Panel B of This table reports the mean and median of firm characteristics at the firm-year level separately for Fraud firms and Control firms in the matched sample. Variable definitions are provided in the Appendix. Matching procedure is described in section 2.6. Statistical significance (two-sided) at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

Panel A: Summary Statistics of Additional Variables

	1	2	3	4	5	6
	N	Mean	Std. Dev.	p10	p50	p90
Abnormal Industry Litigation	72,850	0.514	4.29	-2.5	0	4.25
Abnormal Return Volatility	66,826	0.002	0.0582	-0.061	-0.006	0.076
Abnormal Stock Turnover	67,101	-0.000	0.018	-0.013	-0.001	0.012
Disastrous Stock Return	67,101	0.080	0.272	0	0	0
Unsigned Abnormal Accruals/Sale	65,022	0.711	3.218	0.0069	0.060	0.859

Panel B: Summary Statistics of Matched Sample

	Control Firm		Fraud Firm			
	Mean	Median	Mean	Median	ΔMean	t-value
Size	6.240	5.521	7.155	7.032	-0.915***	-19.6
Age	2.425	2.485	2.425	2.565	0	0
Product Similarity (All-Firm)	0.034	0.031	0.034	0.031	0.001	1.1
Product Similarity (Top-15)	0.175	0.168	0.172	0.164	0.003**	2.05
Industry Homogeneity	0.275	0.263	0.268	0.263	0.007***	4.35
Stock-return Co-movement	0.639	0.620	0.641	0.631	-0.002	-0.8

Table IA2. Product Similarity and Financial Fraud – Alternative Product Similarity

This table reports Conditional Logit estimates for the full sample for Product Similarity on the incidence of fraud using alternative constructions of our primary independent variable. Our proxy for financial fraud includes settled SEC and DOJ enforcement actions and Securities Class Actions from the Stanford University Lawsuit Database. Panel A reports correlations between Product Similarity using all peers and Product Similarity only using the top 15, top 10 and top 5 closest peers. Panel B replaces Product Similarity with all peers with Product Similarity with firms' similarity averaged across its closest 15, 10, and 5 competitors, respectively. All specifications include controls selected from Alawadhi et al. (2023). All specifications are run at the firm-year level, include TNIC×Year fixed effects, and include explanatory variables lagged by one year. The t-statistics, calculated from standard errors clustered at the firm level, are reported in parentheses below the coefficient estimates. Statistical significance (two-sided) at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

Panel A. Correlations

	Product Similarity (Top-5)	Product Similarity (Top-10)	Product Similarity (Top-15)	Product Similarity (All-Firms)
Product Similarity (Top-5)	1.00			
Product Similarity (Top-10)	0.99***	1.00		
Product Similarity (Top-15)	0.97***	1.00***	1.00	
Product Similarity (All-Firms)	0.82***	0.86***	0.88***	1.00

Panel B. Regressions

	1	2	3	4	5	6
	Fraud (t+1)					
	LPM			C-Logit		
Product Similarity (Top-15)	-0.049**			-3.039***		
	(-2.302)			(-4.452)		
Product Similarity (Top-10)		-0.039*			-2.523***	
		(-1.917)			(-3.945)	
Product Similarity (Top-5)			-0.017			-1.291**
			(-0.877)			(-2.379)
Observations	83297	83297	83297	49107	49107	49107
Pseudo/Adj R-squared	0.044	0.044	0.044	0.118	0.118	0.117
Additional Controls	Yes	Yes	Yes	Yes	Yes	Yes
FE	TNIC×Year	TNIC×Year	TNIC×Year	TNIC×Year	TNIC×Year	TNIC×Year

Table IA3. Product Similarity and Financial Fraud – Cross-sectional

This table reports LPM estimates for the matched sample for Product Similarity on the incidence of fraud. Specifically, for each firm with fraud incident, we include in the sample only the first year of fraud. For each treatment firm-year, we then find five control firms in the same industry, year, and age group with the most comparable size. Our proxy for financial fraud includes settled SEC and DOJ enforcement actions and Securities Class Actions. Columns 2 and 4 include additional control variables used in the Fraud prediction model from Alawadhi et al. (2023) and matching-group FE. All specifications are run at the firm-year level and include explanatory variables lagged by one year. The t-statistics, calculated from standard errors clustered at the firm level, are reported in parentheses below the coefficient estimates. Statistical significance (two-sided) at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

	1	2	3	4
		Fraud (t+1)		
Product Similarity (All-Firms)	-2.412**	-3.551		
	(-2.139)	(-0.787)		
Product Similarity (Top-15)			-0.726***	-1.236*
			(-3.278)	(-1.874)
Size	0.068***	0.262***	0.070***	0.263***
	(10.310)	(15.102)	(10.637)	(15.092)
Age	-0.070***	-0.146*	-0.071***	-0.151*
	(-5.571)	(-1.773)	(-5.816)	(-1.860)
Observations	1219	932	1219	932
Adjusted R-squared	0.070	-0.371	0.074	-0.364
Additional Controls	No	Yes	No	Yes
FE	Year	Group	Year	Group

Table IA4. 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 (2021). Our proxy for financial fraud includes settled SEC and DOJ enforcement actions and Securities Class Actions from the Stanford University Lawsuit Database. The t-statistics, calculated from standard errors clustered at the firm 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
	Fraud (t+1)	
	P(F)	P(D|F)
Product Similarity (Top-15)	-3.660***	2.982***
	(-3.481)	(2.590)
Size	0.359***	-0.210***
	(16.101)	(-3.431)
Age	-0.380***	0.310***
	(-7.161)	(4.402)
Change in Gross Margin	-0.046	0.093
	(-1.013)	(1.540)
Sales Growth	-0.128**	0.261***
	(-2.146)	(3.224)
Business Equipment	-0.098	0.311***
	(-0.804)	(2.605)
Telecom	-0.339	0.181
	(-1.396)	(0.675)
HHI (SIC3)	0.184	
	(1.464)	
Soft Asset	0.333***	
	(3.563)	
Operating Lease	0.114**	
	(2.318)	
Security Issue	-0.030	
	(-0.238)	
Losses	0.137***	
	(3.209)	
Geographic Segments	0.031	
	(1.012)	
Auditor Opinion		0.455
		(0.740)
BigN Auditor		-0.059
		(-1.266)
Abnormal Industry Litigation		0.023***
		(3.867)
Abnormal Return Volatility		0.015
		(0.062)
Abnormal Stock Turnover		2.302**
		(2.470)
Disastrous Stock Return		0.396***
		(3.437)
Observations	61750	61750
FE	Year	Year