ARTICLE IN PRESS

Accounting, Organizations and Society xxx (xxxx) xxx

Contents lists available at ScienceDirect

Accounting, Organizations and Society

journal homepage: www.elsevier.com/locate/aos

Using Google searches of firm products to detect revenue management[☆]

Peng-Chia Chiu [a], Siew Hong Teoh [b,*], Yinglei Zhang [c], Xuan Huang [d]

[a] School of Management and Economics, CUHK Business School, The Chinese University of Hong Kong, Shenzhen, China
[b] UCLA Anderson School of Management, USA
[c] Chinese University of Hong Kong, Hong Kong, China
[d] California State University, Long Beach, USA

A R T I C L E I N F O

Article history:
Received 26 May 2021
Received in revised form
4 April 2023
Accepted 11 April 2023
Available online xxx

Handling Editor: Prof. M. Peecher

Keywords:
Big data analytics
Audit risk assessment
Revenue fraud detection
Financial reporting quality
Auditing theory of evidentiary triangulation
Earnings management
Predicting accounting fraud

A B S T R A C T

We introduce a novel Big Data analytics model to detect upward revenue misreporting. The model uses freely available Google searches of firm products to provide external entity business state (EBS) evidence. The veracity of the reported numbers is enhanced when auditors can obtain external EBS evidence congruent with the reported numbers. The Google search volume index (SVI) of firm products is a good candidate for such EBS evidence because it nowcasts (i.e. predicts present) firm sales and is independent of management control. A large discrepancy such as a high sales growth together with a large decline in the SVI suggests possible manipulation upwards of revenues. We find that an indicator variable, MUP, of a firm in the top sales growth quartile and bottom ΔSVI quartile in each industry-quarter predicts revenue misstatements incrementally to the F_Score, Discretionary-Revenues model, two alternative upward revenue manipulation identifiers, and analyst and media coverages. MUP predictability is stronger in end-user industries and in interim quarters relative to the fourth quarter. We also find corroborating evidence that MUP firms have lower sales growth persistence, larger increases in accounts receivables, and lower allowances for bad debts, consistent with their lower revenue quality.

1. Introduction

Financial misreporting remains frequent despite increased regulatory oversight and standard setters' repeated attempts to simplify and standardize revenue recognition. Accounting frauds and misstated financial statements persist. Dyck, Morse, and Zingales (2023) speculate that the corporate frauds we observe could be just the tip of the iceberg. Peecher, Schwartz, and Solomon (2007) express a heightened concern that auditors remain ineffective in detecting audit risks following major auditing failures globally in the 2000s.

The ongoing technological revolution has excited people with the promise that Big Data can help solve accounting and auditing problems. The American Accounting Association (AAA) hosted the inaugural 'Accounting *IS* Big Data' conference in 2015 with the mission "to explore the role of Big Data and analytics in all areas of the accounting profession and to identify the opportunities for accounting education and research" (https://aaahq.org/Meetings/2015/Accounting-IS-Big-Data). Despite several more such annual conferences, there have been few academic studies on Big Data innovations in auditing and accounting fraud detection (see Section 2.2).

We introduce a novel, simple model that leverages new Big Data analytics to help auditors detect revenue misreporting. Many people search for products on Google before making purchases, and these searches have been found to 'nowcast' (i.e. predict present) firm sales (Da, Engelberg, & Gao, 2011a). Google Trends provides search data related to queried terms in a search volume index (SVI)

☆ We thank two anonymous referees, Mark Peecher (editor), Zhaoyang Gu, Lale Guler, Jeff Hoopes, Steve Kachelmeier, Bin Ke, Ningzhong Li, Allison Nicoletti, Peter Oh, Eddie Riedl, Yan Sun, Xiaoli Tian, TJ Wong, Nir Yehuda, Amy Zang workshop participants at Boston University, Chinese University of Hong Kong, National Taipei University, National Taiwan University, New York University, Ohio State University, Singapore Management University, University of Southern California Brown Bag Workshop, University of Texas Dallas, University of Texas, Austin, Cornell University, Tsinghua University, Sun Yat-sen University and European Accounting Association conference, American Accounting Association Annual meeting, and Conference on Textual Analysis in Accounting and Finance for very helpful comments.

* Corresponding author.
E-mail address: steoh@anderson.ucla.edu (S.H. Teoh).

https://doi.org/10.1016/j.aos.2023.101457

at https://trends.google.com; Appendix A explains how to obtain SVI data from Google Trends. We propose that the SVI for firm products is a prime Big Data candidate for aiding with the detection of sales misreporting. We intuit that an inconsistent pattern such as a high sales growth in conjunction with a large SVI decline is suggestive of upward revenue manipulation by managers. In this paper, we show that a simple prototype indicator model, *MUP*, for a firm ranked in the top sales growth quartile and bottom *ΔSVI* quartile each industry-quarter successfully predicts upward revenue misstatements.

Our MUP (manipulation up) model may be viewed as an application of the strategic-system auditing (SSA) framework proposed by the auditing literature (Budescu, Peecher, & Solomon, 2012; Peecher et al., 2007). Auditors perform evidentiary triangulation of data to assess fraud risks using three fundamental sources of evidence: entity business states (EBS), management information intermediaries (MII), and management business representations (MBR). An apparent agreement between MBR and MII evidence is not itself conclusive about the veracity of financial statements because management has control over both sources of evidence. Rather, a third component of external EBS-based evidence, which is not easily manipulated by management, is essential. Auditors can more reliably assess the veracity of the numbers reported in financial statements when all three of EBS, MII, and MBR evidence are congruent with each other (Peecher et al., 2007; Trotman & Wright, 2012).

One promise of Big Data analytics for improving audit quality is this potential to provide the external EBS evidence. Because Google searches serve as externally generated evidence of firm activities and transactions, the *MUP* model using the large deviations between the demand for a firm's products as implied by Google searches and the reported GAAP sales numbers could serve as a red flag that warns auditors of potential revenue misreporting. This basis for the successful detection of misreporting is different from that described in prior research relying on nonfinancial measures (e.g., Brazel, Jones, & Zimbelman, 2009, references in Section 2.2). Most of the non-financial items considered by previous studies are still reported by management and so are not external sources of evidence as required by the SSA framework.

The *MUP* model has other advantages. Google is the most popular search engine in the U.S., so auditors and other stakeholders can utilize its large amount of data to probe the risk of financial misstatements for a large number of auditees in a wide variety of industries.[1] Furthermore, anyone can access Google Trends data for free and in close-to-real time, which presents an opportunity for the *SVI* data related to a firm's products to provide useful and timely information about a firm's demand for its products. If so, the change in SVI (*ΔSVI*) of firm products would be a useful correlate to capture implied sales growth (Da et al., 2011a). In other words, low numbers of Google searches would indicate a low demand for a firm's products and signal low *actual* sales; we validate this assumption that *ΔSVI* nowcasts actual sales growth in the contemporaneous quarter in our sample. Alternative Big Data in previous studies that are correlates of firm activities and transactions, such as cellphone activity and parking lot traffic satellite images at retail stores, are available mainly in proprietary datasets sold to professional traders and hedge funds or are limited to specific industries (see Section 2).

The *MUP* model can be especially useful for auditors because public company auditors have been increasingly foregoing "substantive analytical procedures on large income statement accounts, such as revenue, due to criticisms from regulatory inspectors that such procedures are not capable of providing useful substantive evidence" (Glover, Prawitt, & Drake, 2015, p. 161). Additionally, while increased regulatory oversight has improved the quality of firms' financial reports, these improvements have been disproportionately less significant for filings with less auditor involvement (Boyle, Lewis−Western, & Seidel, 2021). Given that interim quarterly filings are only reviewed by the auditor and that review procedures consist primarily of analytical procedures and inquiries, the use of *MUP* could be particularly useful in auditors' quarterly review procedures.

We focus on revenue misreporting to demonstrate the suitability of *MUP* as a Big Data supplemental auditing tool because regulators worldwide have acknowledged the importance of revenue reporting. The opening statement in the FASB's 2018 announcement of the converged FASB/IASB standard for revenue recognition states that "Revenue is one of the most important measures used by investors in assessing a company's performance and prospects." Revenues have also been one of the most frequently misreported financial statement items (Nelson, Elliot, & Tarpley, 2002, 2003; Turner, Dietrich, Anderson, & Bailey, 2001); see Section 2.2 for further evidence of the importance of revenue reporting. We focus on revenue misstatements in the upward direction (i.e., "*MUP*"), as it is much more common than in the downward direction.

The simple binary indicator model, *MUP*, may be used by auditors or others to check the veracity of manager-reported revenue numbers. We test whether *MUP* can successfully predict upward revenue misstatements identified from material, unintentional, or fraudulent restatements from the Audit and Analytics database.[2] We find that the incremental odds of a revenue misstatement are 165% for a *MUP* firm relative to non-*MUP* firms after controlling for a battery of determinants of misstatements and industry-quarter fixed effects. Our results are robust to sorting observations into quintiles or deciles.

In practice, auditors check the veracity of *pre*-audit revenues, which are unobservable by us as researchers, so we use reported sales as a proxy for pre-audit revenue in our main analysis. To demonstrate the robustness of our main finding, we use analyst sales forecasts as an alternative proxy for pre-audit revenues. Within each industry-quarter, we rank both *ΔSVI* and pre-audit sales growth derived from analyst revenue forecasts and define *MUPpre-audit* as 1 if a firm's *ΔSVI* belongs to the highest quartile but its forecasted audit sales growth belongs to the lowest quartile each industry-quarter, but is otherwise zero. We find that *MUPpre-audit* is also a strong predictor of upward revenue misstatement.

A potential concern with our proposed approach is that auditors might not have the resources or time to assess the revenue growth information for their auditees' industry peers to perform the within-industry ranking required to construct *MUP* as we do. Accordingly, we examine an alternative heuristic fraud detector, *MUPsimple*, which equals 1 for a firm whose current change in sales

[1] The amount of search data is enormous: Google has 90.46% of the search engine market share worldwide, and processes over 63,000 search queries worldwide every second on average, with over 5.4 billion searches per day as of September 23, 2018: https://seotribunal.com/blog/google-stats-and-facts/. The integrity of the search data is crucial for Google to be able to monetize the data, so Google has put in place extensive safeguards and detection tools to prevent attempts to manipulate search data. In short, it would be difficult and costly for firms to manipulate Google search results, which makes SVI the ideal source of external EBS evidence.

[2] We focus on the more common upward revenue manipulation. We do not investigate the opposite discrepancy between the highest *ΔSVI* quartile and the lowest *ΔSales* quartile because such cases are likely associated with product recalls, which are salient bad news events that attract investor attention.

is higher than 15%, but whose *ΔSVI* is lower than −15%, and is otherwise zero.[3] We show that *MUPsimple* can also strongly identify upward revenue manipulators.

We hasten to add that the purpose of the *MUP* model is to serve as a feasible prototype diagnostic tool for auditors to collect EBS evidence to perform evidentiary triangulation. We do not intend for the *MUP* model to substitute for other more traditional indicators of fraud risk. Nevertheless, we examine whether *MUP*'s ability to detect upward revenue manipulation is incremental to a large set of firm characteristics controls and four manipulation detector measures suggested by the literature or motivated by audit field practice.

We demonstrate *MUP*'s incremental predictability of revenue misstatements with respect to two key misreporting predictors in the literature: Dechow, Ge, Larson, and Sloan (2011)'s fraud score (*F_Score*) and Stubben (2010)'s discretionary revenues (*DiscretionaryRev*). We also show *MUP*'s incremental predictability of revenue misstatements with respect to two common benchmarks auditors may use to access the appropriateness of current period revenues. The first benchmark is the growth in revenues relative to same-quarter prior-year revenues. We construct an indicator variable, *AMUP*, to identify auditees having a large deviation between current sales growth from sales growth four quarters ago. The second benchmark uses non-financial data such as headcount to predict financial misreporting (e.g., Brazel et al., 2009). We construct the indicator variable, *HMUP*, to identify firms with a large deviation between sales growth and headcount growth. Prior sales growth and headcount changes are internally generated and management-controlled, so they cannot serve as the external EBS source of evidence.

We also explore the cross-sample variation in *MUP*'s ability to detect revenue misstatement. We find that *MUP* is most effective for firms in the retail and business-to-customer industries where customers are more likely to search for product information before their purchase, and thus that *ΔSVI* can capture the change in customer demand with less noise. As discussed earlier, the benefits of *MUP* may be larger in interim quarters when regulatory oversights such as mandatory audits are absent. Consistent with this prediction, we find that *MUP* is less effective during the fourth fiscal quarter when auditors' substantive audit testing may have required corrections to pre-audit revenues for reporting revenues. If corrections did occur during the audit, reported sales growth would be a noisier proxy for pre-audit sales growth in the fourth quarter than in the interim quarters. These additional cross-sample findings provide further confidence that the ability of *MUP* to assess revenue misstatement risk comes from Google search data as valid EBS evidence to verify the MBR assertion (in this case, asserted revenues).

Finally, sales growth reversals are normal for high sales growth firms, but firms that have manipulated revenues upwards to obtain high sales growth are expected to suffer larger reversals when the misreporting is corrected in the future. Consistent with this prediction, we find that *MUP* firms have larger sales growth reversals than non-*MUP* firms, controlling for the magnitude of sales growth. We also find that, relative to non-*MUP* firms, *MUP* firms have larger increases in accounts receivables and lower allowances for bad debts, which are common channels for upward revenue management. These additional findings provide evidence that the EBS evidence from Google search is congruent with the MBR evidence to identify the misstated revenues.

In commenting on the AAA's pronouncement that ‘Accounting *IS* Big Data’, Sharma (2015) notes that internal and external auditors need to combine Big Data analytics to understand a firm's business and improve audit quality.[4] In the academic literature, Earley (2015) and Appelbaum, Kogan, and Vasarhelyi (2017) discuss the promises and challenges of using Big Data and analytics for improving audit quality, and Appelbaum et al. (2017) call for more research on Big Data applications for auditing. Teoh (2018) discusses the promise that Big Data will be part of a monitoring mechanism to constrain managerial opportunism in firm operations and financial reporting. Our research responds to these calls for Big Data innovation in auditing and in deterring financial misreporting.

Our evidence contributes to the auditing literature by showing that external Big Data can be a useful source of EBS evidence to help improve auditors' assessment of financial misstatement risks and overall audit quality. Our evidence also suggests a similar approach for investors, accounting regulators, and policymakers to detect and hence potentially deter financial misreporting. We urge further research by the academic profession to consider the feasibility and net benefits of establishing an information technology infrastructure that aggregates *external* Big Data correlates of firm activities and transactions with user-friendly software, similar to our *MUP* prototype approach, to provide external EBS evidence in order to enhance audit quality and help deter accounting misreporting.

2. Background

2.1. The SSA framework and the triangulation of three sources of audit evidence

There is a heightened expectation of the level and nature of auditor responsibility for financial-statement fraud (Peecher et al., 2007). Regulations such as SAS 99, AICPA 2002, and PCAOB 2007, have all emphasized such auditor responsibility (Trotman & Wright, 2012). Recent regulations such as SAS 134 and SAS 135 have added to auditors' responsibilities to detect misstatements. However, Dyck et al. (2023) estimate that the likelihood of detection of financial fraud by auditors is only about 29%. Auditing approaches have evolved in response to changes in informational needs, business organization value-creation processes, intangible value drivers, and accounting and auditing technologies (Peecher et al., 2007).

A new approach to public company audits called the strategic-system auditing (SSA) framework, has emerged. Under the SSA framework, auditors would perform evidentiary triangulation in fraud risk assessments (Bell, Peecher, & Solomon, 2005; Peecher et al., 2007) using three fundamental sources: EBS, MII, and MBR.[5] Ongoing research suggests that auditors experience difficulties extracting value from evidentiary triangulation when they assess fraud risk (Trotman & Wright, 2012). A key point of triangulation highlights the inadequacy of testing the veracity of management's financial statement account assertions when auditors compare MBR evidence (e.g., journal entries) with MII evidence (e.g., underlying electronic and paper documentation). Apparent

[3] The threshold is based on the descriptive statistics for the top 25% cutoff for the change in sales and the bottom 25% cutoff for the *ΔSVI*.

[4] See http://aaahq.org/Meetings/2016/Accounting-Is-Big-Data-Conference and the HBS blog by Sharma, a principal at the Ernst and Young Center for Board Matters https://corpgov.law.harvard.edu/2015/10/24/big-data-and-analytics-in-the-audit-process/. Sharma notes that “Insights gleaned from such data can and should extend beyond risk assessment” and that the relevant information “now extend far beyond traditional financial transactional data in a company's general ledgers and extends into data from email, social media, video, voice, texts—mountains of unstructured data.” (Sharma, 2015). See also Zhu (2019).

[5] EBSs are all of entity's business strategies, conditions, processes, and economic actions/events, as well as past, current, and future business relations with other economic entities (Bell et al., 2005).

agreement between MBR and MII evidence, both of which are controlled by management, does not by itself provide reasonable assurance of the material correctness of reported balances. Evidentiary triangulation benefits auditors most when EBS evidence, which is outside the control of management, is gathered on the underlying economic or business states of the company (Peecher et al., 2007). Recent research provides evidence that auditors can leverage EBS-based evidence to improve the justifiability of their evidence-driven, belief-based risk assessments about fraud and improve overall audit quality.

Our novel approach for auditors and capital market stakeholders to obtain crowd-sourced EBS-based evidence is a timely contribution to the auditing literature and capital markets research. By leveraging recently available Big Data sources, such as Google searches, we show how to obtain relevant EBS evidence that is low-cost and widely available to auditors and capital market participants for the purpose of assessing potential revenue-related misreporting risks.

2.2. Research on revenue management

2.2.1. Fraud and revenue management detection models using financial items

Scandals, large and small, continue to occur despite regulatory actions to constrain the pertinent behavior. Dyck et al. (2023) estimate that around 40% of companies misrepresent their financial reports, but only a small percentage of them are detected. Among various financial misstatements, revenues have been shown to be one of the most frequently misreported items on the income statement (Beasley, Hermanson, Carcello, & Neal, 2010; Nelson et al., 2002, 2003; Turner et al., 2001).

Dechow et al. (2011) report that around 54%–60% of the SEC's AAERs involve misstated revenues (page 29, Panel E of Table 1). Similar to the analysis of SEC enforcement releases, Nelson et al. (2002) find that revenue manipulation is the second most common type of earnings manipulation, based on surveys of auditors. Using more recent data, Albrecht, Kim, and Lee (2020) study changes in accounting estimates (CAEs) disclosed in annual or quarterly financial reports. While CAEs involve various types of accounting estimates, Albrecht et al. report that revenue recognition accounts for 30% of all cases and that it is the most frequent type of CAE in their sample. In our sample, around 43% of the restatements involve misstated revenues, more than any other category of misstatement.

Two widely used models for detecting financial misreporting are Dechow et al. (2011) and Stubben (2010). The Dechow et al. (2011) model estimates a financial fraud score using a combination of financial ratios and non-financial measures that predict financial fraud. The Stubben (2010) model uses the residuals in the regression of change in accounts receivable on change in revenues and the interaction of change in revenues with a set of firm and industry characteristics. Both studies have shown that their measures have explanatory power in predicting AAERs. Subsection 3.3 details how we adapt these measures in demonstrating the incremental predictability of the *MUP* model over these two measures. Recall the advantages of *MUP* over these measures: *MUP* is easy for an auditor to implement, complements existing analytical procedures, and incorporates data from an external source.

2.2.2. Fraud and revenue management detection models using non-financial measures (NFMs)

The literature in accounting and finance uses NFMs to forecast/nowcast firm earnings and revenues. Some of the studies use sector-specific NFM measures, such as the number of cellular phone subscribers, Twitter users' feedback on products, internet-related usage, and satellite data on car counts in the parking lots of U.S. retailers to forecast sales and earnings (Amir & Lev, 1996; Bartov, Mohanram, & Seethamraju, 2002; Hughes, 2000; Kang, Stice-Lawrence, & Wong, 2021; Riley, Pearson, & Trompeter, 2003; Tang, 2018; Trueman, Wong, & Zhang, 2001).[6] These data may be voluntarily disclosed on the company's website or financial statements, or reported to regulatory agencies such as the U.S. Department of Transportation or the U.S. Department of Energy. The proprietary nature of some of these data sources limits their accessibility to auditors and average stakeholders.

Two recent studies propose using NFMs to detect reporting manipulation. For example, Brazel et al. (2009) use the number of retail outlets, the amount of warehouse space, and headcount to examine financial misreporting. Allee, Baik, and Roh (2021) find that, amongst Korean firms, an inconsistent growth pattern between accounting performance and electricity consumption is a useful indicator for detecting financial misreporting.

Our study differs from prior NFM studies in several important ways. Specifically, our data source from Google Trends is provided by an entity that is independent of and external to the firm, is directly related to sales, covers a large number of industries, and is made freely available to the public almost instantaneously. Therefore, it is a suitable source of EBS evidence auditors can use to probe the revenue misstatement risk.

2.3. Evidence of the relevance of google search data

Studies in economics have found Google search data useful in nowcasting a wide range of economic activities, including unemployment (Askitas & Zimmermann, 2009; Suhoy, 2009; D'Amuri & Marcucci, 2017), private consumption (Vosen & Schmidt, 2011), exchange rates (Bulut, 2018), and some monthly economic indicators, such as automobile sales, unemployment claims, travel destination planning, and consumer confidence (Choi & Varian, 2012).

In the accounting and finance fields, web search data have also been used to provide proxies for investor sentiment and investor attention. Past studies have found that these proxies are correlated with contemporaneous and subsequent returns (Da, Engelberg and Gao 2011b, 2015) and nowcast growth in personal consumption and retail sales (Della Penna & Huang, 2010). Drake, Roulstone, and Thornock (2012) use Google searches of a firm's stock as a proxy for investor demand for information about the firm. Chi and Shanthikumar (2016) find that firms are searched more intensively by individuals closer to a firm's headquarters. Da et al. (2011a) find that changes in the *SVI* for a firm's products strongly nowcast revenue surprises. This finding justifies Google searches of firm products as a correlate of the implied demand for a firm's products and, therefore, of a firm's sales. Our study goes beyond the research objective of these studies to examine whether the discrepancy between the implied sales growth from the change in the *SVI* related to queries about the firm's products and the reported sales growth can provide EBS-based evidence that assists auditors in assessing revenue fraud risk.

[6] The examples in this literature include the number of cell phone subscribers, population coverage, market penetration, web traffic, pollutant emissions, and customer satisfaction. More recent studies examine NFMs such as product reviews on Amazon.com (Huang, 2018), employees' predictions of their employer's business outlook in Glassdoor.com (Huang, Li, & Markov, 2020), employee satisfaction in Glassdoor.com (Hales, Moon, & Swenson, 2018), proprietary data on mobile phone activity (Froot, Kang, Ozik, & Sadka, 2017), and satellite images of cars in parking lots (Kang et al., 2021).

Table 1
Summary statistics.

Panel A: Descriptive Statistics (Google Search Sample)

Variable	Mean	SD	p25	p50	p75
ΔSVI	0.01	0.29	−0.14	−0.02	0.11
ΔSales	0.06	0.23	−0.04	0.05	0.14
Size	7.88	2.07	6.56	7.97	9.41
BTM	0.54	0.58	0.25	0.45	0.75
Lev	3.06	6.34	0.70	1.45	3.41
Loss	0.20	0.40	0.00	0.00	0.00
BIG4	0.86	0.35	1.00	1.00	1.00
OPCycle	3.04	19.24	0.00	0.00	0.03
Age	26.13	19.65	12.00	21.00	36.00
Sale_Vol	0.03	0.04	0.01	0.02	0.04
IO_Own	0.50	0.40	0.00	0.61	0.86
Special	0.57	0.50	0.00	1.00	1.00
Ret_Vol	0.02	0.01	0.02	0.02	0.03
Past Ret	0.11	0.42	−0.13	0.08	0.30
#Analysts	9.33	8.31	2.00	7.00	15.00
MUP	0.05	0.22	0.00	0.00	0.00
Misstate_Rev	0.01	0.10	0.00	0.00	0.00
F_Score	0.50	0.42	0.25	0.42	0.62
DiscretionaryRev	0.00	0.02	−0.01	0.00	0.01
AMUP	0.05	0.21	0.00	0.00	0.00
HMUP	0.09	0.29	0.00	0.00	0.00
AccRev	0.01	0.05	0.00	0.00	0.02
DefRev	0.00	0.02	0.00	0.00	0.00
Allowance	0.04	0.06	0.01	0.02	0.05

Panel B Industry Composition (Google Search Sample vs. COMPUSTAT Sample)

Industry	% of Sample	% of COMPUSTAT	Diff
Consumer Nondurables	7.75%	3.60%	−4.15%***
Consumer Durables	3.92%	1.92%	−2.00%***
Manufacturing	8.89%	8.16%	−0.73%***
Energy	2.17%	5.04%	2.87%***
Chemicals and Allied Products	2.42%	2.07%	−0.36%***
Business Equipment	15.06%	17.67%	2.61%***
Telecommunication	4.30%	2.81%	−1.49%***
Utilities	2.90%	2.27%	−0.64%***
Wholesales and Retails	14.32%	6.80%	−7.52%***
Healthcare	6.10%	14.05%	7.95%***
Finance	19.93%	22.10%	2.17%***
Others	12.23%	13.51%	1.28%***

Panel C Descriptive Statistics (Google Search Sample vs. COMPUSTAT Sample)

Variable	Sample		COMPUSTAT		Difference	
	Mean	Med	Mean	Med	Mean	Med
ΔSales	0.06	0.05	0.10	0.07	***	***
Size	7.88	7.97	5.97	5.90	***	***
BTM	0.54	0.45	0.66	0.50	***	***
Lev	3.06	1.45	2.54	1.06	***	***
Loss	0.20	0.00	0.36	0.00	***	***
BIG4	0.86	1.00	0.50	1.00	***	***
OPCycle	3.04	0.00	117.06	0.08	***	***
Age	26.13	21.00	13.08	9.00	***	***
Sale_Vol	0.03	0.02	0.04	0.02	***	***
IO_Own	0.50	0.61	0.35	0.25	***	***
Special	0.57	1.00	0.28	0.00	***	***
Ret_Vol	0.02	0.02	0.03	0.02	***	***
Past Ret	0.11	0.08	0.10	0.04	***	***
#Analysts	9.33	7.00	3.49	1.00	***	***

Panel D Pearson Correlations

	Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1)	*ΔSVI*	1.00												
(2)	*ΔSales*	**0.11**	1.00											
(3)	*Size*	**0.03**	**0.09**	1.00										
(4)	*BTM*	**−0.02**	**−0.14**	**−0.17**	1.00									
(5)	*Lev*	**−0.01**	−0.01	**0.06**	**0.18**	1.00								
(6)	*Loss*	**−0.06**	**−0.17**	**−0.35**	**0.14**	**−0.03**	1.00							
(7)	*BIG4*	0.01	**−0.02**	**0.45**	**−0.08**	**−0.03**	**−0.10**	1.00						
(8)	*OPCycle*	**−0.02**	**−0.02**	**−0.31**	**−0.04**	**−0.02**	**0.14**	**−0.30**	1.00					
(9)	*Age*	−0.01	**−0.10**	**0.28**	**−0.05**	**−0.04**	**−0.10**	**0.10**	**−0.06**	1.00				

(*continued on next page*)

Table 1 (*continued*)

Panel D Pearson Correlations

	Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(10)	*Sale_Vol*	**0.03**	**0.06**	**−0.28**	**−0.09**	**−0.16**	**0.17**	**−0.08**	**0.11**	**−0.13**	1.00			
(11)	*IO_Own*	**0.00**	**0.05**	**0.13**	**−0.06**	**0.01**	**−0.10**	**0.14**	**−0.15**	**0.03**	**−0.15**	1.00		
(12)	*Special*	**−0.05**	**−0.05**	**0.12**	−0.01	0.00	**0.09**	**0.13**	**−0.10**	**0.07**	**−0.09**	**0.10**	1.00	
(13)	*Ret_Vol*	**−0.03**	**−0.15**	**−0.47**	**0.25**	**0.03**	**0.38**	**−0.11**	**0.07**	**−0.20**	**0.24**	**−0.14**	**0.03**	1.00
(14)	*Past Ret*	**0.04**	**0.25**	**0.17**	**−0.28**	**−0.02**	**−0.20**	**0.03**	**−0.02**	−0.01	**0.02**	**0.04**	**−0.03**	**−0.19**
(15)	*#Analysts*	**0.02**	**0.05**	**0.61**	**−0.13**	**−0.02**	**−0.17**	**0.29**	**−0.17**	**0.15**	**−0.12**	**0.24**	**0.08**	**−0.22**
(16)	*MUP*	**−0.23**	**0.25**	−0.01	**−0.02**	**0.02**	0.00	**−0.03**	0.01	**−0.05**	**0.04**	0.01	−0.01	**0.02**
(17)	*Misstate_Rev*	−0.01	**0.02**	**−0.03**	0.00	0.00	**0.01**	**−0.03**	**−0.01**	**−0.05**	**0.02**	**0.03**	0.00	**0.03**
(18)	*F_Score*	**0.03**	**0.27**	**0.02**	**0.03**	**0.10**	**−0.09**	**−0.06**	−0.01	**−0.06**	**0.01**	**0.05**	**0.06**	**−0.05**
(19)	*DiscretionaryRev*	0.00	**0.01**	**0.01**	0.00	0.00	**−0.04**	**0.01**	**−0.02**	0.01	**−0.02**	−0.01	**−0.02**	0.01
(20)	*AMUP*	0.00	**0.30**	**−0.02**	**0.03**	**0.01**	**0.02**	**−0.02**	**0.03**	**−0.01**	**0.04**	**−0.04**	0.00	**0.04**
(21)	*HMUP*	**−0.06**	**−0.30**	**−0.16**	**0.07**	0.01	**0.17**	**−0.03**	**0.07**	0.00	**0.04**	**−0.05**	**0.05**	**0.12**
(22)	*AccRev*	**0.07**	**0.45**	**0.02**	**−0.05**	**0.10**	**−0.14**	**−0.12**	**0.02**	**−0.08**	−0.01	**0.03**	**−0.03**	**−0.14**
(23)	*DefRev*	**0.01**	**0.16**	**0.03**	**−0.06**	0.00	0.00	0.01	0.00	**−0.07**	0.01	0.01	**−0.04**	**−0.02**
(24)	*Allowance*	**−0.02**	**−0.04**	**−0.15**	−0.01	−0.01	**0.10**	**−0.03**	**0.04**	**−0.11**	0.01	**0.02**	**−0.03**	**0.19**

	Variable	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
(14)	*Past Ret*	1.00										
(15)	*#Analysts*	**0.04**	1.00									
(16)	*MUP*	**0.05**	**−0.02**	1.00								
(17)	*Misstate_Rev*	0.00	−0.01	**0.01**	1.00							
(18)	*F_Score*	**0.01**	**−0.04**	**0.08**	−0.01	1.00						
(19)	*DiscretionaryRev*	0.00	0.01	0.01	0.00	0.01	1.00					
(20)	*AMUP*	**0.05**	**−0.06**	**0.22**	0.00	0.00	0.01	1.00				
(21)	*HMUP*	**−0.07**	**−0.10**	**−0.07**	−0.01	**−0.08**	**−0.02**	**−0.07**	1.00			
(22)	*AccRev*	**0.11**	**−0.02**	**0.12**	0.01	**0.36**	**0.14**	**0.07**	**−0.15**	1.00		
(23)	*DefRev*	**0.05**	**0.07**	**0.05**	0.01	**0.07**	0.01	**0.02**	**−0.08**	**0.14**	1.00	
(24)	*Allowance*	**−0.04**	−0.01	**−0.02**	**0.08**	**−0.12**	0.00	−0.01	**0.05**	**−0.09**	−0.01	1.00

Note: This table reports summary statistics of the main variables and firm characteristics. The sample period is from 2004 to 2020. Panel A reports the mean, median, standard deviation, and the first- and the third-quartile statistics for main variables of our Google search sample. Panel B compares the industry composition of the Google search sample and the overall Compustat sample. Panel C compares the summary statistics of the firm characteristics of the Google search sample and the overall Compustat sample. Panel D reports Pearson correlations among variables. *** indicates $p < 0.01$; ** $p < 0.05$; * $p < 0.1$, $p < 0.05$. In Panel D, $p < 0.05$ is bolded. All variables are as defined in Appendix B.

3. Sample and nowcasting of google searches

3.1. Search volume index (SVI)

The Google Trends platform normalizes *SVI* data to be between 0 and 100, enabling this measure to be comparable across firms of different sizes.[7] We obtain monthly *SVIs* for the brand names of the products in our sample from January 2004 to December 2020. We illustrate how to use the Google Trends platform to obtain the *SVI* in Appendix A.

To identify a firm's main brand product names for search queries, we obtain the brand names from Nielsen Media Research and Ad$pender.[8] Following Da et al. (2011a), we select the brand with the largest advertising units (expenditures) for each firm for the search queries. This selection method ensures that the search queries are restricted to only products of sufficient importance to a firm's revenue to warrant a high advertising budget. The selection of search terms could potentially be modified to accommodate more targeted goals, such as an external auditor who has more extensive information about firm brands and their associated revenues when auditing an auditee's revenues. We hand-match to obtain the set of firms that are publicly traded and covered by COMPUSTAT database. This procedure yields a list of 1,872 firms.

The search term that consumers type into the query box need not be the exact brand name. Following Da et al. (2011a), we ask student research assistants how they would search for each product/brand as a consumer. For big name retailers such as Target or Walmart, the search term consumers use is often the retailer's name. When their search terms differ, we use all of their suggestions as related searches on Google Trends. Google Trends returns the top-searched related brand names, and we choose the top related query. For each firm-quarter, we calculate seasonally adjusted *ΔSVI* as the percentage change of *SVI* over the same quarter in the previous year. Historical financial data are from COMPUSTAT and stock returns data are from CRSP. To facilitate comparisons and interpretation, we winsorize *ΔSales* and *ΔSVI* at−/+100 percent.[9] The final sample has 46,739 firm-quarter observations, covering 2004 to 2020. The sample size varies across tests because of different data requirements.

3.2. Upward revenue manipulation indicator MUP based on google search data

For each firm-quarter, we calculate seasonally adjusted sales growth (*Δsales*) as the percentage change in sales for the current quarter over the sales of the same quarter in the prior year. We then sort firm observations in each Fama-French 48 industry-calendar

[7] See https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052 for details.

[8] We acquired brand name information of each firm between 2004 and 2014 through a one-time purchase from Nielsen Media Research, which tracks firms' television advertising. To update the sample to the more recent period between 2015 and 2020, we obtained additional brand name data from Ad$pender. We thank Chuchu Liang for providing us with the additional data necessary to facilitate the analysis.

[9] In our sample, only 730 observations (i.e., 1.47%) are winsorized, so the nowcasting results remain robust if we truncate rather than winsorize the sample. Neither winsorization nor truncation affects *ΔSales* or*ΔSVI* rankings, and therefore does not affect the construction of *MUP*. Consequently, the misstatement results are unaffected by winsorization or transaction.

quarter separately by the change in search volume, *ΔSVI*, and the change in sales, *ΔSales*. Observations in the lowest *ΔSVI* quartile and the highest *Δsales* quartile are assigned *MUP* value 1, with *MUP* set to 0 otherwise.[10] Most audit failures and SEC enforcement actions involve revenue manipulation—especially upward (see for example Files, 2012; Palmrose, Richardson, & Scholz, 2004; Teoh, Wong, & Rao, 1998). Therefore, we use *MUP* to identify upward manipulation. Our results are robust if *MUP* is constructed using quintiles or a modified decile ranking in each industry-quarter.[11]

3.3. Alternative revenue manipulation detectors

We examine whether *MUP*'s ability to detect upward revenue misstatements is incremental to four alternative revenue manipulation detectors that are known to auditors and other stakeholders. In the first alternative predictor, we calculate the quarterly *F_Score* values for our sample using the estimated coefficients in model 3 (page 68) of Dechow et al.'s (2011) annual *F_Score* model and apply them to our quarterly variables. The *F_Score* model uses a comprehensive set of determinants consisting of financial variables (accruals, Δreceivables, Δinventory, percentage of soft assets, Δcash sales, Δreturn on assets, and an issuance indicator), non-financial and off-balance-sheet variables (abnormal change in the number of employees and an indicator showing the existence of operating leases), and market-based variables (past returns and book-to-market ratios).

The second alternative predictor estimates quarterly discretionary revenues (*DiscretionaryRev*) following Stubben (2010)'s conditional revenue model (Equation (5), page 702), but using the quarterly instead of annual values for the model determinants. Each quarter, the change in accounts receivables is regressed on the change in revenues and on the interaction of the change in revenues with control variables, including firm size, age, and its square, separate positive and negative industry-median adjusted growth rates in revenues, and industry-median adjusted gross margin and its square. The discretionary revenues are the regression residuals.

The third alternative predictor *AMUP* uses sales growth four quarters ago in order to benchmark current reported sales growth. For each calendar quarter, firms' current quarter sales growth $\Delta Sales_t$ and sales growth from the same quarter last year ($\Delta Sales_{t-4}$) are sorted independently into quartiles within each Fama-French 48 industry classification. The indicator *AMUP* is set to 1 for firms that are in the bottom $\Delta Sales_{t-4}$ quartile and the highest $\Delta sales_t$ quartile, and is 0 otherwise.

The final alternative predictor variable, *HMUP*, is designed to identify firms with a large deviation between reported sales growth and a change in the number of employees. For each calendar year, firms are ranked independently into quartiles by *ΔSales* and by the most recently available growth in the number of employees within each Fama-French 48 industry classification.[12] *HMUP* firms belong to the quartile with the lowest change in the number of employees and the quartile with the highest *ΔSales*.

3.4. Other variables

The key dependent variable *Misstate_Rev* equals 1 for the misstated periods, which are the actual periods being restated (not the restatement announcement period), and 0 otherwise. Following Bartov, Marra, and Momenté (2021), we use "big R" restatements from the Audit Analytics dataset for firm-years involved in a misstated period during which firms disclose the filing of Form 8-K Item 4.02. These big R misstatements refer to material (unintentional) or fraudulent (intentional) errors in financial statements. For each revenue misstatement event, we rely on the cumulative restated net income in the Audit Analytics dataset to identify the direction of revenue misstatement. As such, we assume that the direction of manipulated revenue is the same as that of manipulated income. Because the restatement amount for each individual quarter is not available, we assume that firms manipulate revenue in each quarter covered in the restated period.[13]

We examine the effects on potential revenue management for *MUP* firms by examining changes in certain accrual accounts, including changes in accounts receivables (*AccRev*) and deferred revenues (*DefRev*). We also examine whether *MUP* firms reserve lower allowances for bad debts (*Allowance*) in order to supplement accrued revenue manipulation. These variables are scaled by beginning-quarter total assets.

3.5. Summary statistics

Table 1, Panel A presents the mean, median, standard deviation, and the first- and third-quartile statistics for the main variables of the sample firms. The typical firm has a mean of one percent in the quarterly change of Google search volume, *ΔSVI*. At the first (third) quartile, there is about a −14% (11%) decrease (increase) in *ΔSVI*, and the standard deviation is about 29%. The average quarterly sales growth is 6%. *MUP* firms account for around 5% of total observations. About one percent of the observations in our sample exhibit upward revenue manipulation, consistent with prior studies (Huang & Hairston, 2021; Lennox, Lisowsky, & Pittman, 2013). The descriptive statistics of all other control variables used in our model are consistent with prior literature (Da et al., 2011a; Dechow et al., 2011; Demerjian, Lev, Lewis, & McVay, 2013).

Panel B compares the industry composition of our Google Trends sample and the overall COMPUSTAT population over the sample period. Unsurprisingly, our sample firms have a higher representation than the COMPUSTAT sample in end-user industries, such as Wholesale and Retail, Consumer Nondurables, and Business Equipment.

Panel C compares firm characteristics used as controls in later analyses between our sample and the overall COMPUSTAT population. In general, our sample firms are larger and have lower sales growth rates, lower book-to-market ratios, and higher leverage.

Panel D reports Pearson correlations among key variables. The correlation between *ΔSVI* and *ΔSales* is significantly positive at 0.11, consistent with the ability of the Google search volume of a firm's products to nowcast same-quarter revenues.[14] Consistent with the broad literature, revenue misstatement firms tend to be small, young, audited by non-big four, and have higher sales volatility and

[10] On average, we have 15 observations for each industry-quarter. As a robustness test, we drop observations that have fewer than 5 observations for each industry-quarter. The results are quantitatively similar.

[11] Decile rankings produce too few *MUP* = 1 observations (less than 0.5% of our sample). In keeping with the spirit of decile rankings, we modify the *MUP* procedure to obtain a sufficient discrepancy between the *ΔSales* and *ΔSVI* decile ranks. *MUP* = 1 if *ΔSales* decile rank exceeds *ΔSVI* decile rank by at least 7, otherwise *MUP* = 0.

[12] Headcount data is only available on an annual basis. In order to convert the data to a quarterly frequency, we use the most recent available headcount data for each firm-quarter observation.

[13] Our assumption may introduce noise in the *Misstate_Rev* variable. The measurement error of the dependent variable is likely to reduce the power of our test and make it more difficult to obtain test significance.

[14] Untabulated results show a similar pattern for the autocorrelations of *ΔSVI* and of *ΔSales*; that is, + + + -. This suggests that revenues and *SVI* likely share an underlying generating process. In other words, common fundamental factors drive both *ΔSVI* and *ΔSales*, so the instantaneous availability of *ΔSVI* could be useful for nowcasting sales growth before sales information is released.

higher incidence of losses (e.g., Callen, Robb, & Segal, 2008). As a preliminary univariate test of *MUP* as a potential misstatement indicator, we find that *Misstate_Rev* is positively correlated with *MUP* but not with the other misstatement indicators *F_Score, DiscretionaryRev, AMUP* and *HMUP*. Furthermore, the correlations between *MUP* and *F_Score, DiscretionaryRev,* and *HMUP* are small (0.08, 0.01, and −0.07, respectively), suggesting an initial indication that *MUP* may provide incremental predictability of misstatements that is largely independent of past measures.

Appendix C replicates the nowcasting test of Da et al. (2011a), confirming that *ΔSVI* nowcasts *ΔSales* in our larger sample. This result is a necessary validation that *ΔSVI* is a suitable correlate of implied firm sales for us to move on to test whether *MUP* is able to identify upward revenue manipulation. Column (1) shows that the coefficient on *ΔSVI* is positive and strongly significant, 0.080 (t-statistic = 8.22), consistent with *SVI*'s nowcasting of sales. The coefficient on *ΔSVI* remains positive and significant in Columns (2) and (3) when additional control variables are included. Overall, the results are comparable to those in Da et al. (2011a).

4. Main results

4.1. Predicting revenue misstatements by MUP

In our first main analysis, we test whether *MUP* firms are more likely to be upward revenue manipulators by estimating the effect of *MUP* on the likelihood of upward revenue misstatements. We run the following logistic regression.

$$Misstate\_Rev_{jt} = \beta_0 + \beta_1\, MUP_{jt} + other\ controls + \varepsilon_{jt} \quad (1)$$

Misstate_Rev = 1 for the quarter in which the big R restatement from Audit Analytics identifies that a revenue number is overstated. We control for industry fixed effects to control for industry differences in misreporting. We also include lagged sales changes and a similar set of firm characteristic controls as in the nowcasting regression shown in Appendix C Column (3). Since a misstatement usually lasts for a year, once the *MUP* indicator variable becomes 1, we keep *MUP* as 1 for the next three quarters.[15]

Table 2 Panel A reports the regression results. The *MUP* coefficient in Column (1) is significantly positive at 0.764 (z-statistic = 2.51). Column (2) adds additional controls using firm characteristics, and the *MUP* coefficient 0.973 (z-statistic = 2.73) remains incrementally significantly positive. For economic magnitude, this coefficient translates into an incremental odds ratio of misstatement of 165% ($e^{0.973}-1 = 1.65$) for *MUP* firms relative to non-*MUP* firms. These results show that *MUP* successfully predicts revenue misstatement in firms that later are publicly revealed to having restated their revenues. Our results provide direct evidence that Google search data on product demand contains information about the business state of the firm, and thus can serve as useful EBS evidence that auditors could potentially use to assess revenue fraud risk.

4.2. MUP based on alternative pre-audit revenue proxy

At the early stage of an audit engagement, auditors have access to pre-audit revenues to construct *MUP*. Because pre-audit revenues are unobservable by us as researchers, we use reported sales as a proxy in the earlier construction of *MUP*. In this subsection, we examine another candidate, analyst sales forecasts, to proxy for pre-audit revenues. Analyst revenue forecast has several advantages as a proxy for pre-audit revenues in that it is more timely and incorporates industry and macro-wide information (Piotroski & Roulstone, 2004).[16]

We calculate pre-audit sales growth as the difference between the analyst sales forecasts for the fiscal quarter and actual sales four quarters ago, scaled by the latter. Within each industry-quarter, we rank both *ΔSVI* and the analyst-forecasted revenue growth. *MUPpre-audit* = 1 if a firm is in the highest *ΔSVI* quartile and the lowest analyst-forecasted growth quartile, and is 0 otherwise. We rerun our main test of Equation (1) and report the results in Table 2 Panel B Column (1). We find that *MUPpre-audit* is also a strong predictor of upward revenue manipulation with a *MUPpre-audit* coefficient of 0.830 (z-statistic = 2.61).

To the extent that auditors and analysts are independent of management, post-audit reported sales and analyst sales forecasts are likely more in line with true sales than with pre-audit revenues. It suggests that our researcher-built *MUP* and *MUPpre-audit* would have lower power and are therefore biased against finding significance, than the *MUP* model auditors can build using pre-audit revenues in practice.

4.3. A heuristic application of the MUP model

In practice, auditors may not be able to wait to assess the revenue growth information for all industry peers so they can perform the necessary industry rankings to construct *MUP*. Accordingly, in this subsection we propose a simple heuristic approach that could be used to apply our model in practice.

From the descriptive statistics in our overall sample, the top 25% cutoff for *ΔSales* is about 15%, and the bottom 25% cutoff for *ΔSVI* is about −15%. For this reason, we define the alternative fraud detector, *MUPsimple*, as 1 for a firm whose *ΔSales* in the current quarter is higher than 15% but whose *ΔSVI* is lower than −15%, and 0 otherwise. Using a simple fixed cutoff rule is common practice, such as for calculating Altman Z-score or O-score using in-sample coefficients to assess financial default risk or to classify a firm with *F_Score*>1 as “above normal fraud risk” or *F_Score* >2.45 as “high fraud risk.”

The results using *MUPsimple* to predict misstatements are reported in Column (2) of Table 2 Panel B. We find that even such a simple *MUP* implementation successfully identifies revenue misstatements (coefficient = 0.500, z-statistic = 2.29) in our sample. Overall, the evidence demonstrates that it is feasible for an auditor to implement our *MUP* approach to obtain an initial sample of higher risk auditees for more careful substantive testing.

4.4. Explanatory power relative to other fraud predictors

In this subsection, we examine whether *MUP* predictability of revenue misstatement is incremental to several other predictors. When *F_Score* is included in the regression, Table 3 Column (1) reports that *MUP* remains incrementally significant (coefficient = 1.050, z-statistic = 3.06). The *F_Score* coefficient is not significant in our sample.

The test of *MUP* as a revenue predictor when *DiscretionaryRev* is added as an additional regressor is reported in Column (2) of Table 3. Our key variable *MUP* remains statistically significant

[15] The results are robust in a lower power test with a more conservative *MUP* indicator variable that is equal to 1 only for the initial quarter of revenue misstatement indicated in the Audit Analytics.

[16] We control for lagged sales in the misstatement regression, which precludes the random walk proxy. A simple trend proxy has high noise in a large heterogeneous sample. An alternative candidate, management sales forecasts, is available only for a smaller sample.

Table 2
Logistic Regressions of Revenue Misstatement on *MUP*, alternative proxies for *MUP* and Controls.

Panel A Logistic Regressions of Revenue Misstatement on *MUP* and Controls

Dependent Variable	(1)		(2)	
	Misstate_Rev		Misstate_Rev	
	Coefficients	z-statistics	Coefficients	z-statistics
MUP	**0.764****	**(2.51)**	**0.973*****	**(2.73)**
$\Delta Sales_t$			−0.889**	(-2.43)
$\Delta Sales_{t-1}$			−0.192	(-0.68)
$\Delta Sales_{t-2}$			−0.055	(-0.22)
$\Delta Sales_{t-3}$			0.199	(0.54)
Size			0.064	(0.45)
BTM			0.116	(0.40)
Lev			−0.010	(-0.52)
Loss			−0.003	(-0.01)
BIG4			−0.919**	(-1.96)
OPCycle			−0.017**	(-2.00)
Age			−0.012	(-1.29)
Sale_Vol			2.722	(0.67)
IO_Own			0.615	(0.81)
Special			0.167	(0.56)
Ret_Vol			16.305	(1.02)
Past Ret			−0.229	(-1.14)
#Analysts			−0.078*	(-1.87)
# of Obs.	26,194		19,464	
Pseudo. R^2	0.07		0.14	

Panel B Logistic Regressions of Revenue Misstatement on *MUPpre-audit*, *MUPsimple* and Controls

Dependent Variable	(1)		(2)	
	Misstate_Rev		Misstate_Rev	
	Coefficients	z-statistics	Coefficients	z-statistics
MUPpre-audit	**0.830*****	**(2.61)**		
MUPsimple			**0.500****	**(2.29)**
$\Delta Sales_t$	−0.598	(-1.58)	−0.596*	(-1.88)
$\Delta Sales_{t-1}$	−0.127	(-0.42)	−0.1	(-0.40)
$\Delta Sales_{t-2}$	0.175	(0.56)	0.243	(0.90)
$\Delta Sales_{t-3}$	0.569*	(1.85)	0.424	(1.52)
Size	0.005	(0.04)	0.089	(0.68)
BTM	−0.159	(-0.53)	0.144	(0.56)
Lev	−0.005	(-0.23)	−0.009	(-0.44)
Loss	0.14	(0.37)	0.096	(0.29)
BIG4	−1.027**	(-2.37)	−1.011**	(-2.24)
OPCycle	−0.011	(-1.53)	−0.015*	(-1.92)
Age	−0.016	(-1.57)	−0.013	(-1.42)
Sale_Vol	3.37	(0.82)	3.53	(0.96)
IO_Own	0.589	(0.89)	0.768	(1.07)
Special	0.032	(0.12)	0.126	(0.47)
Ret_Vol	3.364	(0.28)	14.497	(1.02)
Past Ret	−0.38	(-1.33)	−0.252	(-1.35)
#Analysts	−0.060*	(-1.67)	−0.067*	(-1.77)
# of Obs.	17,673		20,948	
Pseudo. R^2	0.15		0.14	

Note: This table reports logistic regression results of upward revenue misstatements on *MUP*, alternative proxies for *MUP*, and controls. Regression results on *MUP* are in Panel A, and on alternative proxies, *MUPpre-audit* and *MUPsimple* are in Panel B. The sample period is from 2004 to 2020. The dependent variable *Misstate_Rev* is set to 1 for the quarters in which revenues had to be restated downwards. *MUP*, an indicator variable for a likely upward revenue manipulator, is set to 1 for a firm in the bottom *ΔSVI* quartile and top *ΔSales* quartile, and is zero otherwise; ranking of observations is performed for each industry-calendar quarter. *MUPpre-audit* is an indicator variable set to 1 for a firm in the bottom *ΔSVI* quartile and top quartile of pre-audit sales growth (proxied by analyst revenue forecast), and is zero otherwise. *MUPsimple* is an indicator variable set to 1 for a firm with *ΔSales* higher than 15% but *ΔSVI* lower than −15%, and is zero otherwise. Appendix B defines the variables *ΔSales, ΔSVI*, pre-audit sales growth, and the following control variables: size (*Size*), book-to-market ratio (*BTM*), leverage (*Lev*), a loss indicator (*Loss*), a Big Four indicator (*BIG4*), operating cycle (*OPCycle*), firm age (*Age*), the standard deviation of sales over at least three of the last eight quarters (*Sale_Vol*), institutional ownership (*IO_Own*), special items (*Special*), the standard deviation of the monthly stock returns in the prior year (*Ret_Vol*), the return over the past 12 months (*Past Ret*), and number of analysts following (*#Analysts*). Industry and calendar quarter fixed effects are included in all regressions. Standard errors are clustered by firm and by calendar quarter. *z*-statistics are reported in parentheses. The intercept is included but not tabulated for brevity. *** indicates $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

(coefficient = 0.915, z-statistic = 2.31) and also economically important.

The *AMUP* indicator variable identifies firms with high current-quarter sales growth but low past-year same-quarter sales growth. Column (3) of Table 3 reveals that *MUP* remains economically and statistically significant with the inclusion of *AMUP*. The *HMUP* indicator uses the change in employee growth, measured using headcount, to benchmark sales growth, allowing a prediction of revenue manipulation. Order backlog is also available from COMPUSTAT, but only 10% of our sample has non-missing observations for this variable, so we exclude it for constructing *HMUP*. The result for *MUP* predictability after controlling for *HMUP* is in Column (4) of Table 3. We find that *MUP* remains economically and highly statistically significant, so it is incremental to the inclusion of *HMUP*. In

Table 3
Logistic regressions of revenue misstatement on *MUP* and other revenue manipulation detectors.

Dependent Variable	(1)		(2)		(3)		(4)	
	Misstate_Rev		Misstate_Rev		Misstate_Rev		Misstate_Rev	
	Coefficients	z-statistics	Coefficients	z-statistics	Coefficients	z-statistics	Coefficients	z-statistics
MUP	**1.050*****	**(3.06)**	**0.915****	**(2.31)**	**1.101*****	**(3.03)**	**1.054*****	**(3.12)**
F_Score	−0.466	(-1.23)						
DiscretionaryRev			4.129	(1.17)				
AMUP					−0.132	(-0.33)		
HMUP							−0.499	(-1.16)
$\Delta Sales_t$	−0.878**	(-2.22)	−0.873**	(-2.31)	−0.972**	(-2.31)	−1.076***	(-2.88)
$\Delta Sales_{t-1}$	−0.188	(-0.57)	−0.139	(-0.40)	−0.250	(-0.74)	−0.290	(-0.94)
$\Delta Sales_{t-2}$	−0.015	(-0.05)	−0.046	(-0.16)	−0.076	(-0.24)	−0.191	(-0.70)
$\Delta Sales_{t-3}$	0.166	(0.48)	0.266	(0.65)	0.199	(0.48)	−0.066	(-0.21)
Size	0.144	(1.03)	0.083	(0.54)	0.029	(0.21)	0.106	(0.75)
BTM	0.169	(0.57)	−0.107	(-0.25)	0.219	(0.77)	0.147	(0.49)
Lev	−0.010	(-0.55)	−0.002	(-0.09)	−0.009	(-0.48)	−0.009	(-0.49)
Loss	0.000	(0.00)	0.117	(0.30)	0.098	(0.25)	0.065	(0.18)
BIG4	−0.990**	(-2.06)	−0.837*	(-1.67)	−0.895*	(-1.81)	−0.960**	(-1.99)
OPCycle	−0.016**	(-2.00)	−0.017**	(-2.24)	−0.021*	(-1.85)	−0.017**	(-2.08)
Age	−0.015	(-1.48)	−0.013	(-1.17)	−0.012	(-1.20)	−0.011	(-1.22)
Sale_Vol	3.934	(0.96)	5.022	(1.23)	3.973	(0.91)	3.785	(0.89)
IO_Own	0.575	(0.73)	0.366	(0.42)	0.726	(0.89)	0.579	(0.74)
Special	0.220	(0.72)	0.170	(0.52)	0.227	(0.75)	0.230	(0.78)
Ret_Vol	18.518	(1.17)	−1.568	(-0.13)	9.886	(0.61)	18.603	(1.18)
Past Ret	−0.298	(-1.38)	−0.250	(-0.96)	−0.144	(-0.65)	−0.215	(-1.03)
#Analysts	−0.093**	(-2.12)	−0.097**	(-2.07)	−0.080*	(-1.82)	−0.091**	(-2.07)
# of Obs.	18,824		13,087		17,455		19,040	
Pseudo. R^2	0.15		0.16		0.16		0.15	

Note: This table reports logistic regression results of upward revenue misstatements on *MUP*, other revenue management proxies, and controls. The sample period is from 2004 to 2020. The dependent variable *Misstate_Rev* is set to 1 for the quarters in which the revenue number had to be restated downwards. *MUP* is an indicator variable set to 1 if the firm is a likely upward revenue manipulator, defined as a firm in the bottom *ΔSVI* quartile and the top *ΔSales* quartile; ranking of observations is performed for each industry-calendar quarter. *F_Score* is the fraud detection score as defined in Dechow et al. (2011), adapted to quarterly frequency. *DiscretionaryRev* is the discretionary revenue as defined in the conditional revenue model of Stubben (2010), as shown in our Equation (4). Indicator *AMUP* equals 1 for firms in the bottom *ΔSales* quartile four quarters ago but in the highest *ΔSales* quartile currently, and is zero otherwise. Indicator *HMUP* equals 1 for firms in the lowest quartile of the change in number of employees and the highest *ΔSales* quartile, and is zero otherwise. Control variables are defined analogously as in Table 3 *z*-statistics are reported in parentheses. Industry and calendar quarter fixed effects are included and standard errors are clustered by firm and by calendar quarter. The intercept is included but not tabulated for brevity. *** indicates $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

contrast, the coefficient on *HMUP* is not significant in our sample.

4.5. Cross-sample analyses

To corroborate Google search data as reliable EBS evidence for assessing the veracity of firm reported sales, we next explore cross-sample variation in the ability of *MUP* to detect revenue misstatements. We expect that *MUP* is most effective at identifying revenue misstatements when Google search data can best capture customer demand with less noise or when managers' reporting is less likely to be scrutinized by auditors.[17]

4.5.1. Retail and business-to-customer industries vs. other industries

We expect that Google search exhibits a stronger correlation with product demand for industries in which customers are more likely to search product information before their purchase, that is, where *ΔSVI* nowcasts sales growth well. We examine revenue misstatement predictability in the sample of retail and business-to-customer industries versus other industries, similar to the partitions in Chakravarthy, DeHaan, and Rajgopal (2014). We rerun the main Equation (1) regression in Table 2 for each subsample and report the results in Table 4, Panel A. As predicted, *MUP* misstatement predictive ability is concentrated in the retail and the business-to-consumer industries (coefficient = 1.152, z-statistic = 2.87). In contrast, *MUP* is insignificant in the non-retail and the non-business-to-customer industries.

4.5.2. Interim quarters versus the fourth quarter

Our analyses are at the quarterly level. Substantive audit procedures are typically performed in the latter part of the fiscal year, when unusual or incorrect revenue transactions would be identified and corrected as part of the audit, bringing reported sales growth more in line with actual sales. We therefore expect *MUP* to have a greater ability to identify revenue misstatement during interim quarters than in the fourth fiscal quarter.

We partition our sample into two subsamples: an interim-quarter sample (i.e. first, second and third fiscal quarters) and a fourth quarter—only sample. We then run separate Equation (1) regressions for the two subsamples; the results are reported in Panel B of Table 4. As expected, we find that *MUP* is a strong identifier of the misstatement of revenues for the interim quarters (coefficient = 1.096, z-statistic = 3.21). In contrast, we do not find revenue misstatement predictability in the fourth-quarter sample. We also run Equation (1) regression for each of the interim quarters separately and find that the interim-quarter coefficients for *MUP* are 1.20, 1.24, and 1.21, respectively for Q1, Q2 and Q3, and all are significant at the one percent level.

These findings suggest that auditors are able to exert control over auditees to correct misstated reported sales. They may also suggest that managers are cautious about overstating annual sales and choose on their own to reverse out aggressive interim-quarter

[17] We thank an anonymous referee for suggesting these additional cross-sample tests.

reporting of sales during the fourth quarter because of the annual audit.[18] Therefore, reported sales would be a noisier proxy for pre-audit sales in the fourth quarter than in the interim quarters. Collectively, our findings support the auditing theory of evidential triangulation. *MUP*'s ability to assess revenue fraud risk comes from Google search data providing valid EBS evidence to help verify MBR assertions (in this case, reported revenues).

4.6. Type I and type II error analysis

Models that estimate discretionary manipulation of financial reporting items using regression model residuals present an inherent and well-known statistical "bad model" problem. Specifically, the bad model problem results in false positives, a Type I error that reduces the model's reliability in detecting manipulation when it occurs. Our simple binary indicator model may have certain advantages that complement the accruals-based regression residual measure. *MUP* does not rely on model specifications of accrual behaviors, so it partially sidesteps the bad-model econometric issues. So long as sales increases are accompanied by increases in Google searches of similar magnitude, *MUP* will be less likely to misidentify firms as manipulators, so it has a smaller Type I error rate.

Another advantage of *MUP* is that it can also potentially detect non-accruals-based revenue manipulations such as barter transactions or grossed-up revenues or accruals-based revenue manipulation that are later masked by actions like factoring of account receivables. For example, a firm can artificially boost sales through a liberal or unconditional return scheme. If the sales are in cash, a traditional accruals-based revenue detection approach will fail to discover such manipulation. Even if bartering, grossing-up of revenues, or factoring situations occur, *MUP* could correctly identify manipulation as long as the sales increase is from the false reporting of sales not accompanied by Google searches that reflect genuine interest from consumers in the firms' products.

We evaluate the Type I and Type II error rates for the *MUP* model in our sample. Untabulated results indicate that we correctly classify 85.49% of 26,194 firm-quarter observations as having 13.74% of Type I error and 75.68% of Type II error. For comparison, Dechow et al. (2011) report that predicting manipulation using an *F_Score* with a cutoff of 1.0 correctly classifies 63.71% of 133,461 firm-year observations as having a 36.31% of Type I error and 31.38% of Type II error in their sample. These findings suggest that the *MUP* model is superior in attaining a higher correct classification rate and a lower false positive rate. Section 4.2 describes how our researcher *MUP* model likely has lower testing power (a higher Type II error rate) than the auditor's model because the auditor has pre-audit revenues and other information to improve the testing power of the auditor.

5. Additional corroborative and robustness analyses

We provide additional corroborative and robustness tests for whether *MUP* firms likely have misreported revenues. We first investigate whether the revenue surprises of *MUP* firms have less persistence, a commonly used measure of revenue quality in the literature. To triangulate our EBS evidence with an MBR source of evidence, we then examine whether *MUP* firms are more likely associated with financial statement items that are common channels for revenue manipulation, such as accounts receivables, deferred revenues, and allowance for bad debts. Lastly, we report on some additional robustness analyses.

5.1. Are the revenue surprises of MUP firms less persistent than those of non-MUP firms?

Many studies use the notion of persistence to estimate the quality of earnings or revenues (Atwood, Drake, & Myers, 2010; Baber, Kang, & Kumar, 1998; Dechow, Ge, & Schrand, 2010; Demerjian et al., 2013; Givoly, Hayn, & Katz, 2010). The reasoning is that high-quality revenue surprises are likely sustainable whereas low-quality manipulated revenue surprises tend to reverse. Our primary findings indicate that *MUP* firms tend to have overstated revenue surprises, and therefore would exhibit lower revenue surprise persistence. We run the following regression with the inclusion of *MUP* and its interaction with revenue surprises to examine differences in persistence between *MUP* and non-*MUP* firms:

$$\Delta Sales_{jt+1} = \beta_0 + \beta_1 \Delta Sales_{jt} + \beta_2 MUP_{jt} + \beta_3 MUP_{jt} \times \Delta Sales_{jt} + \text{other controls} + \text{quarter fixed effect} + \varepsilon_{jt} \quad (2)$$

where $\Delta Sales_{jt+1}$ is next-quarter sales growth. The main variable of interest is *MUP* × *ΔSales*. A negative β_3 implies that *MUP* firms have less revenue growth persistence after controlling for the magnitude of revenue growth.

Column (1) of Table 5 shows that the coefficient on *MUP* × *ΔSales* is negatively significant (−0.083, t-statistic = −2.79), consistent with *MUP* firms having lower revenue growth persistence relative to non-*MUP* firms. In terms of economic significance, the sales growth of *MUP* firms is on average about 15% lower than that of non-*MUP* firms (0.542 vs. 0.646). In Column (2) of Table 5, we add the complete set of controls in Column (3) of Appendix C (i.e., the nowcasting model) and the interaction terms between *ΔSales* and the additional controls. The persistence coefficient (β_3) remains significant and negative (coefficient = −0.067, t-statistic = −1.84).

In sum, our finding that the sales revenues of *MUP* firms reverse faster than those of non-*MUP* firms corroborates that *MUP* identifies revenue misreporting.

5.2. How do MUP firms manage revenues upwards?

To test whether additional MBR evidence indicating revenue misstatements corroborates the EBS evidence on how *MUP* firms may have manipulated revenues, we explore the potential revenue management channels that are likely used by *MUP* firms. Caylor (2010) finds that firms manipulate both accrued revenues and deferred revenues to avoid negative earnings surprises. Teoh et al. (1998) find that firms manage earnings using the allowance for bad debts. Therefore, we investigate the behavior of accrued and deferred revenues as well as the allowance for bad debts for *MUP* firms.

To test whether *MUP* firms use accounts receivables to manipulate reported revenues upward, we extend the conditional discretionary accounts receivable model of Stubben (2010) to include our variable of interest *MUP*, as shown in the regression below:

[18] We also examine the robustness of *MUPannual* based on the discrepancy between annual changes in sales growth and annual changes in the *SVI*. The *MUPannual* coefficient is a slightly larger 1.095 (versus 0.93 for quarterly changes) but has weaker statistical significance (z-statistic = 1.51, two-tailed p-value = 0.13). However, *MUPannual* in the retail and the business-to-customer industries remains significant (coefficient = 1.333, z-statistic = 1.90, two-tailed p-value = 0.057). The lower power of the test for *MUPannual* is expected; the number of observations fell by almost 75%.

Table 4
Logistic regressions of revenue misstatement on *MUP* and controls under subsamples.

Panel A Retail and business-to-customer industries vs. Non-Retail and Non-business-to-customer industries

Dependent Variable	(1)		(2)	
	Misstate_Rev *Retail and business-to-customer industries*		*Misstate_Rev* *Non-Retail and non-business-to-customer industries*	
	Coefficients	z-statistics	Coefficients	z-statistics
MUP	**1.152*****	**(2.87)**	0.423	(0.49)
$\Delta Sales_t$	−1.005***	(-2.69)	−0.643	(-0.77)
$\Delta Sales_{t-1}$	−0.353	(-1.05)	−0.019	(-0.04)
$\Delta Sales_{t-2}$	−0.149	(-0.43)	0.372	(0.86)
$\Delta Sales_{t-3}$	0.216	(0.47)	−0.068	(-0.12)
Size	0.037	(0.22)	−0.049	(-0.21)
BTM	0.109	(0.30)	−0.226	(-0.42)
Lev	−0.018	(-0.66)	−0.005	(-0.19)
Loss	−0.188	(-0.46)	0.669	(1.51)
BIG4	−0.801	(-1.54)	−1.923**	(-1.98)
OPCycle	−0.017**	(-2.25)	−0.776*	(-1.78)
Age	−0.010	(-0.93)	−0.026**	(-2.11)
Sale_Vol	1.508	(0.35)	5.311	(0.56)
IO_Own	0.277	(0.29)	1.893**	(2.39)
Special	0.096	(0.26)	0.272	(0.47)
Ret_Vol	2.455	(0.19)	52.577**	(2.55)
Past Ret	−0.086	(-0.31)	−0.387	(-1.15)
#Analysts	−0.078	(-1.57)	−0.045	(-0.64)
# of Obs.	14,360		2,970	
Pseudo. R^2	0.12		0.25	

Panel B First Three Quarters vs. Fourth Quarter

Dependent Variable	(1)		(2)	
	Misstate_Rev Interim Quarters		Misstate_Rev Fourth Quarter	
	Coefficients	z-statistics	Coefficients	z-statistics
MUP	**1.096*****	**(3.21)**	0.448	(1.18)
$\Delta Sales_t$	−0.978**	(-2.13)	−0.826	(-1.20)
$\Delta Sales_{t-1}$	−0.297	(-0.77)	1.154	(1.39)
$\Delta Sales_{t-2}$	0.388	(0.97)	−2.446*	(-1.96)
$\Delta Sales_{t-3}$	−0.077	(-0.22)	1.486*	(1.79)
Size	0.059	(0.40)	0.150	(0.94)
BTM	0.151	(0.52)	0.085	(0.26)
Lev	−0.005	(-0.26)	−0.049**	(-2.05)
Loss	−0.121	(-0.31)	0.551	(1.31)
BIG4	−0.815*	(-1.71)	−1.246***	(-2.66)
OPCycle	−0.016*	(-1.88)	−0.044	(-1.26)
Age	−0.012	(-1.31)	−0.012	(-1.09)
Sale_Vol	2.249	(0.51)	6.302**	(2.05)
IO_Own	0.651	(0.91)	1.023	(1.28)
Special	0.226	(0.77)	0.061	(0.21)
Ret_Vol	18.479	(1.23)	−2.356	(-0.13)
Past Ret	−0.237	(-1.17)	−0.538**	(-2.06)
#Analysts	−0.079*	(-1.78)	−0.114***	(-2.62)
# of Obs.	13,886		3,208	
Pseudo. R^2	0.15		0.19	

Note: This table reports logistic regression results of upward revenue misstatements on *MUP,* and controls in subsamples. Panel A reports results of retail and business-to-customer industries vs. other industries. Panel B reports results of interim quarters (1st, 2nd, and 3rd fiscal quarters) vs. fourth fiscal quarter. The sample period is from 2004 to 2020. The dependent variable *Misstate_Rev* is set to 1 for the quarters in which the revenue number had to be restated downwards. *MUP* is an indicator variable set to 1 if the firm is a likely upward revenue manipulator, defined as a firm in the bottom ΔSVI quartile and top $\Delta Sales$ quartile, and is zero otherwise; ranking of observations is performed for each industry-calendar quarter. Appendix B defines the variables $\Delta Sales$, ΔSVI, and the following control variables: size (*Size*), book-to-market ratio (*BTM*), leverage (*Lev*), a loss indicator (*Loss*), a Big Four indicator (*BIG4*), operating cycle (*OPCycle*), firm age (*Age*), the standard deviation of sales over at least three of the last eight quarters (*Sale_Vol*), institutional ownership (*IO_Own*), special items (*Special*), the standard deviation of the monthly stock returns in the prior year (*Ret_Vol*), the return over the past 12 months (*Past Ret*), and number of analysts following (*#Analysts*). Industry and calendar quarter fixed effects are included in regressions. Standard errors are clustered by firm and by calendar quarter. *z*-statistics are reported in parentheses. The intercept is included but not tabulated for brevity. *** indicates $p < 0.01$; **$p < 0.05$; *$p < 0.1$.

$$AccRev_{jt} = \beta_0 + \beta_1 MUP_{jt} + \beta_2\,\Delta SVI_{jt} + \beta_3 \Delta Sales_{jt} + \Delta Sales_{jt} \times (\beta_4 Size_{jt} + \beta_5 Age_{jt} + \beta_6 AgeSqr + \beta_7 GrowPos_{jt} + \beta_8 GrowNeg_{jt} + \beta_9 GM_{jt} + \beta_{10} GMSqr_{jt}) + other\ controls + \varepsilon_{jt}, \tag{3}$$

In addition to *MUP,* the regressors include ΔSVI, $\Delta Sales$, the interaction of $\Delta Sales$ with the firm's size, age, and its square, indicator variables for positive and negative industry-adjusted growth rate (*GrowPos, GrowNeg)*, gross margins and its square (*GM and GMSqr)*, and other controls. We include the search variable ΔSVI because it proxies for demand for the firms' products and high numbers of searches that translate to actual credit sales would increase accounts receivables. The set of other controls are common firm fundamental characteristics, including size, book-to-market ratio, leverage, a loss dummy, return on assets, and the length of the operating cycle to proxy for the economic determinants of accrued revenues.

Table 5
Persistence of sales changes for *MUP* firms.

Dependent Variable	(1)		(2)	
	$\Delta Sales_{t+1}$		$\Delta Sales_{t+1}$	
	Coefficients	t-statistics	Coefficients	t-statistics
$\Delta Sales_t$	0.646***	(35.51)	0.860***	(8.67)
MUP	0.013*	(1.79)	0.012	(1.39)
***MUP* × $\Delta Sales_t$**	**−0.083*****	**(-2.79)**	**−0.067***	**(-1.84)**
$\Delta Sales_{t-1}$			0.072***	(2.96)
$\Delta Sales_{t-2}$			0.084***	(3.71)
$\Delta Sales_{t-3}$			−0.218***	(-12.46)
Size			0.006***	(4.23)
BTM			−0.016***	(-6.22)
Lev			−0.000	(-0.49)
Loss			−0.002	(-0.77)
BIG4			−0.010**	(-2.30)
OPCycle			0.000	(0.02)
Age			−0.000***	(-6.31)
Sale_Vol			0.050	(1.17)
IO_Own			0.001	(0.44)
Special			−0.003	(-1.20)
Ret_Vol			−0.150	(-1.40)
Past Ret			0.036***	(7.99)
#Analysts			−0.001***	(-2.92)
ΔSales interactions with Controls			Included	
# of Obs.	44,818		32,376	
Adj. R^2	0.45		0.54	

Note: This table reports results of regressions of $\Delta Sales_{t+1}$ on $\Delta Sales_t$, *MUP*, *MUP*×$\Delta Sales_t$, and control variables. The sample period is from 2004 to 2020. The dependent variable, $\Delta Sales_{t+1}$, is the sales change in the next quarter. *MUP* is an indicator variable set to 1 if the firm is in the bottom *ΔSVI* quartile and top *ΔSales* quartile, and is zero otherwise; ranking of observations is performed for each industry-calendar quarter. *MUP*×$\Delta Sales_t$ is the interaction variable between *MUP* and $\Delta Sales_t$. $\Delta Sales_t$ and $\Delta Sales_{t-i}$ (i = 1 to 3) are current and one-to three-quarter-lagged seasonal sales changes. Appendix B defines the variables *ΔSales, ΔSVI*, and the following control variables: size (*Size*), book-to-market ratio (*BTM*), leverage (*Lev*), a loss indicator (*Loss*), a Big Four indicator (*BIG4*), operating cycle (*OPCycle*), firm age (*Age*), the standard deviation of sales over at least three of the last eight quarters (*Sale_Vol*), institutional ownership (*IO_Own*), special items (*Special*), the standard deviation of the monthly stock returns in the prior year (*Ret_Vol*), the return over the past 12 months (*Past Ret*), and number of analysts following (*#Analysts*). Calendar quarter fixed effects are included and standard errors are clustered by firm and by calendar quarter. *t*-statistics are reported in parentheses. The intercept is included but not tabulated for brevity. *** indicates $p < 0.01$; **$p < 0.05$; *$p < 0.1$.

Table 6
Regressions of accrued revenues, deferred revenues, and allowance on *MUP*.

Dependent Variable	(1)		(2)		(3)	
	AccRev		DefRev		Allowance	
	Coefficients	t-statistics	Coefficients	t-statistics	Coefficients	t-statistics
MUP	**0.013*****	**(5.13)**	0.001	(1.51)	**−0.007****	**(-2.22)**
ΔSVI_t	0.004***	(3.56)	−0.000	(-0.22)	−0.006*	(-1.78)
$\Delta Sales_t$	0.241**	(2.33)	0.044***	(7.97)	−0.012	(-0.44)
$\Delta Sales_t$×*Size*	0.019**	(2.42)				
$\Delta Sales_t$× *Age*	0.003	(1.17)				
$\Delta Sales_t$×*AgeSqr*	−0.000	(-1.32)				
$\Delta Sales_t$× *GrowPos*	−0.063	(-0.67)				
$\Delta Sales_t$×*GrowNeg*	−0.141	(-1.44)				
$\Delta Sales_t$×*GM*	0.101**	(2.14)				
$\Delta Sales_t$×*GMSqr*	0.065	(0.73)				
$Sales_t$			−0.006***	(-4.12)	−0.025*	(-1.78)
GrossRec					0.010	(0.41)
StdSales					−0.096***	(-2.59)
RecTurnover					0.002**	(2.22)
Size	−0.000	(-1.52)	0.000	(1.09)	−0.005***	(-4.26)
BTM	0.000	(0.25)	−0.001**	(-2.01)	−0.004	(-1.13)
Lev	0.001***	(4.91)	−0.000	(-0.13)	0.000	(0.03)
Loss	−0.007***	(-5.38)	0.002***	(2.86)	0.008**	(2.58)
ROA	0.010	(0.80)	−0.004	(-0.31)	0.010	(0.22)
OPCycle	0.000	(1.25)	−0.000	(-0.09)	−0.000	(-0.27)
# Obs	30,836		25,089		14,933	
Adj. R^2	0.22		0.03		0.08	

Note: This table reports results of regressions of sales-related accruals variables on *MUP* and control variables. The sample period is from 2004 to 2020. In Column (1), the dependent variable *AccRev* is change in accrued revenues. In Column (2), the dependent variable *DefRev* is change in deferred revenues. In Column (3), the dependent variable *Allowance* is the allowance for uncollectible account receivables. *MUP* is an indicator variable set to 1 if the firm is a likely upward revenue manipulator, defined as a firm in the bottom *ΔSVI* quartile and top *ΔSales* quartile, and is zero otherwise; ranking of observations is performed for each industry-calendar quarter. All variables are defined in Appendix B. Calendar quarter fixed effects are included in regressions and standard errors are clustered by firm and by calendar quarter. *t*-statistics are reported in parentheses. The intercept is included but not tabulated for brevity. *** indicates $p < 0.01$; **$p < 0.05$; *$p < 0.1$.

We report the results in Column (1) of Table 6. The coefficient on *MUP* (β_1) in the accrued revenue regression is positively significant (0.013, t-statistic = 5.13), consistent with our conjecture that *MUP* firms likely manage revenues upward by increasing accrued revenues. The effect of *MUP* is also economically significant. Moving *MUP* from 0 to 1 increases accounts receivables by about 1.1% of lagged total assets, or about 28% of the standard deviation of changes in accounts receivables.

To test whether *MUP* firms use deferred revenues account to manipulate revenues, we use the determinants of deferred revenues in Srivastava (2014) and add our key variables *MUP* and *ΔSVI* as shown in the regression below:

$$DefRev_{jt} = \beta_0 + \beta_1 MUP_{jt} + \beta_2 \Delta SVI_t + \beta_3 \Delta Sales_{jt} + \beta_4 Sales_{jt} + other\ controls + \varepsilon_{jt} \quad (4)$$

High numbers of Google searches may occur around prepayments, but are only associated with actual revenues with a lag. Thus, upward revenue manipulation detected by *MUP* is less likely via the deferred revenue channel. Thus, we do not expect the *MUP* coefficient β_1 in Equation (5) to be significant. Consistent with our expectation, the coefficient in Column (2) of Table 6 is not statistically significant.

We also investigate whether *MUP* firms likely have a lower allowance for uncollectible accounts. Adapting Jackson and Liu (2010), the following regression model is:

$$Allowance_{jt} = \beta_0 + \beta_1 MUP_{jt} + \beta_2 \Delta SVI_{jt} + \beta_3 \Delta Sales_{jt} + \beta_4 Sales_{jt} + \beta_5 GrossRec_{jt} + \beta_6 StdSales_{jt} + \beta_7 RecTurnover_{jt} + other\ controls + \varepsilon_{jt} \quad (5)$$

We include sales level, gross accounts receivable, the standard deviation of sales, receivable turnover, *ΔSVI*, *ΔSales*, and the same set of other common controls as previously described. The coefficient on *MUP* in the allowance regression (Table 6, Column (3)) is negatively significant (−0.007, t-statistic = −2.22). The effect of *MUP* is also economically significant, with a change of *MUP* from 0 to 1 showing a decrease in *Allowance* by about 0.7% of gross receivables, or about 14% of its standard deviation. This result complements the finding from Column (1). *MUP* firms report higher increases in accounts receivables but reserve less for uncollectible accounts than do *non-MUP* firms, consistent with *MUP* firms managing accruals and biasing revenues upward.

5.3. Additional robustness tests

The business press and analysts may also be conduits for whistleblowers to come forward about financial reporting fraud, though such investigations entail significant lag time. Some studies document a positive effect of media on earnings management, mainly via a monitoring role that curbs managerial opportunistic behaviors (Cheng, Liu, & Wei, 2020). The business media has also been shown to play a role in improving the flow of information in capital markets, thereby reducing information asymmetry between managers and outside stakeholders (Bushee, Core, Guay, & Hamm, 2010; Drake, Guest, & Twedt, 2014; Fang & Peress, 2009; Guest, 2021). In summary, media reporting enhances monitoring by drawing the attention of market participants to potential managerial opportunism.

In this subsection, we investigate *MUP*'s ability to predict revenue misstatements incremental to analyst and media coverage. Analyst coverage is already included as a control in our baseline regressions. We add media coverage as an additional control variable in our main regression model to examine the robustness of our results. Using the Ravenpack dataset, we calculate media coverage as the natural log of the times the firm was mentioned by media in the contemporaneous quarter.

We observe that the coefficient on both media coverage and analyst coverage are significantly negative in the regression, suggesting that higher media and analyst coverage are indeed associated with lower revenue restatements (untabulated). Notably, the ability of *MUP* to detect revenue misstatements remains robust and the magnitude of the coefficient on *MUP* is basically unchanged as compared with that in the baseline regression.

5.4. Additional caveats

Despite the extensive corroborative and robustness tests we conduct, we recognize that *MUP* has limitations. Both reported sales and Google search index volume (*SVI*) are imperfect proxies for a firm's true product demand. There are several reasons why Google searches for brands capture actual sales with noise. Not all Google searches translate into actual sales, and not all purchases are preceded by online searches. As a result, the observed correlation between the *ΔSVI* and *ΔSales* may be only modest, which would reduce the test power of the *MUP* predictor. These contaminations likely bias against the ability to detect revenue manipulation.

Recall that our goal is to propose a simple prototype model to assist auditors and other stakeholders in assessing revenue fraud risk. This task does not require that Google Trend's *SVI* be a perfect proxy for the consumer demand for a firm's products, but rather imply only that Google searches contain information that is outside of managerial control about consumer demand for the concurrent fiscal period of sales. An additional contribution is that *MUP* showed incremental predictive power beyond past predictors of restatements.

6. Conclusion

We show that *MUP*—an indicator based on the incongruence between (a) the quarterly change in the Google search volume index and (b) the sales revenue growth as reported by the firm—can serve as useful external EBS evidence that may assist auditors in assessing revenue misstatement risk. Specially, we find that *MUP* firms are more likely to have big R restatements than non-*MUP* firms.

Our results are robust to modifying *MUP* in two ways: using analyst forecasts to proxy for pre-audit revenue that is available to auditors, and using a simple fixed cutoff (±15%) for sales growth and change in Google searches so that the model is easily implementable by auditors. We also show that *MUP* is a stronger and incremental predictor to known fraud detectors such as the *F_Score,* discretionary revenues, two alternative indicator predictors based on current sales growth and its deviation from sales growth in the same quarter prior year and from headcount growth, analyst coverage, and media coverage.

To corroborate the key results, we find that *MUP* predicts revenue misstatements more strongly for firms in end-user industries, for which Google searches are likely to capture demand with less noise. The *MUP* predictability is also stronger during the first three fiscal quarters, when reported sales growth is a better proxy for pre-audit sales growth than that of the fourth quarter. We further find that *MUP* firms tend to have lower revenue surprise persistence, higher increases in accrued revenues, and a lower allowance for bad debt expenses, consistent with the inflated revenue surprises for *MUP* firms.

Our purpose in identifying the *MUP* effects is to demonstrate a simple use of Big Data that is external to management control and available at minimal cost in real time to auditors and other stakeholders for the purpose of assessing revenue fraud risk. Since a simple tool like *MUP* can help in assessing revenue fraud risk, we can suggest a potential policy implication for reducing waste associated with misreporting behaviors. We propose that the SEC and accounting regulators look beyond the rules of mandatory disclosure to the public. We suggest that they expand their role to guide the development of a technology infrastructure that aggregates firm-level information about transactions and activities directly from external parties transacting with the firm and make the information publicly available at low cost to auditors and other stakeholders.

For example, perhaps SEC or accounting regulators could provide the infrastructure with support from Google Trends to organize and aggregate search data about all of a firm's products, as suggested by our *MUP* prototype model. Additional software to automate the nowcasting of sales and compare the forecasts generated from this external source to the reported sales could provide EBS evidence to auditors and make potential misreporting of sales transparent. Facilitating the detection of potential misreporting of firm fundamentals would reduce the benefits of misreporting to firm managers and therefore discourage the practice. Sunshine may indeed be the best disinfectant for fraudulent reporting of revenues.

We understand that many other considerations will need to be worked out before our idea for *MUP* can be implemented, fully functional, and useful. Privacy or proprietary issues pertaining to individual transactions with a public firm may be overcome by using only aggregated, and therefore anonymized, information. Firm disclosures may also be encouraged to provide other legitimate reasons for any large deviations between reported fundamentals and the implied fundamentals suggested by Big Data.

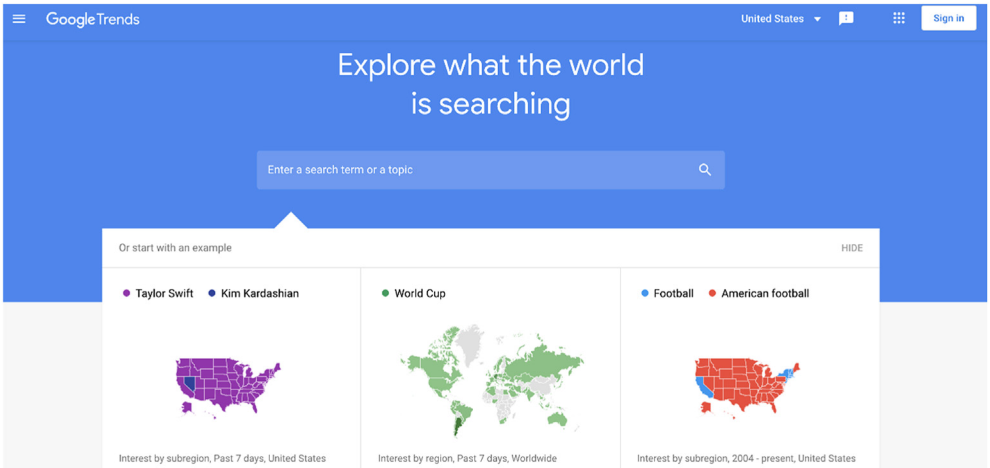
In addition to Google searches, there are other potential Big Data candidates that could be harnessed in a similar manner to constrain the misreporting of revenues. Similarly, Big Data correlates for other firm activities could be utilized to detect misreporting of expenses, not just revenues. We encourage further research into suitable candidates for such external information correlates of major firm activities in order to build a comprehensive external information infrastructure that provides closer to real-time information about firm performance for all investors, perhaps for a small fee to defray the costs of maintaining the information infrastructure. Such an infrastructure may enrich available EBS evidence for auditors and stakeholders so that they can improve the detection of and potentially deter accounting fraud.

Data availability

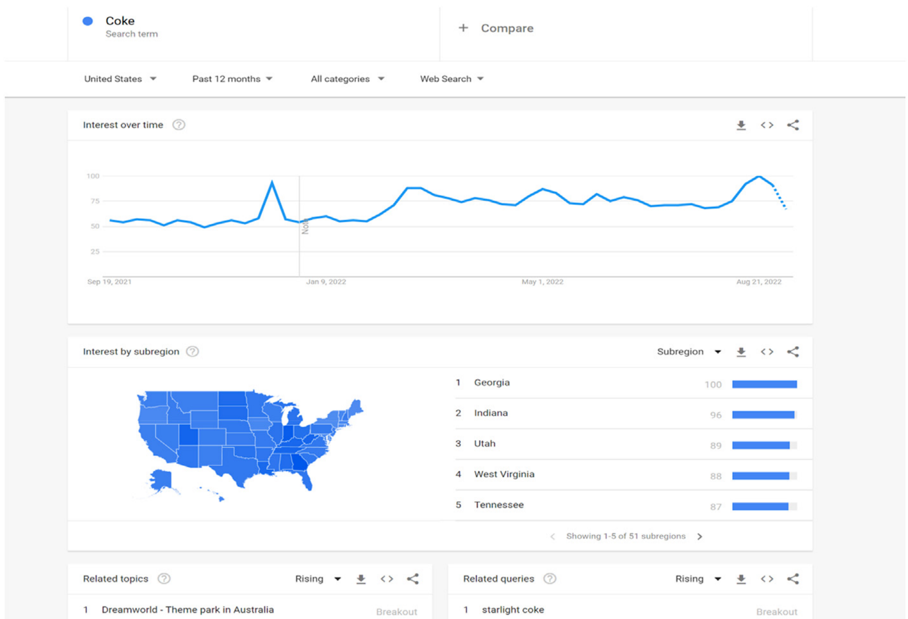
Data from Google Trends are available publicly. All other data sources are available via subscription or purchase as explained in the paper.

Appendix A. How to download Google Trends Search Volume Index (*SVI*) Data

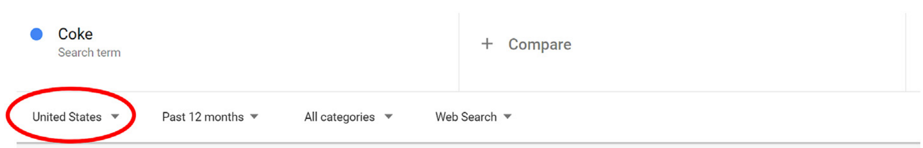
Google Trends website available at https://trends.google.com/is an online search tool that allows the user to see how often specific keywords, subjects and phrases have been queried over a specific period of time.



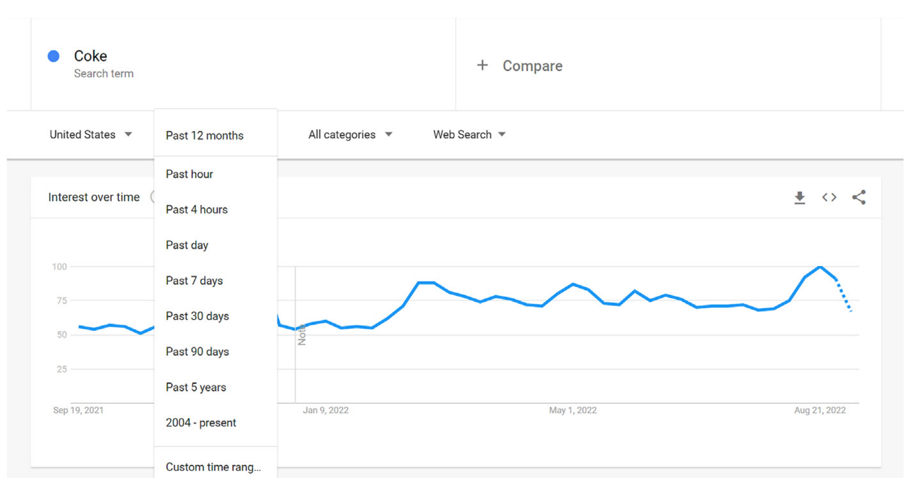
For example, to get the Google search volume index (*SVI*) for Coke, we type "Coke" in the search bar and will get the following screenshot. The blue line shows the *SVI* of Coke at different time periods. It also shows interest by subregion and related top searches.



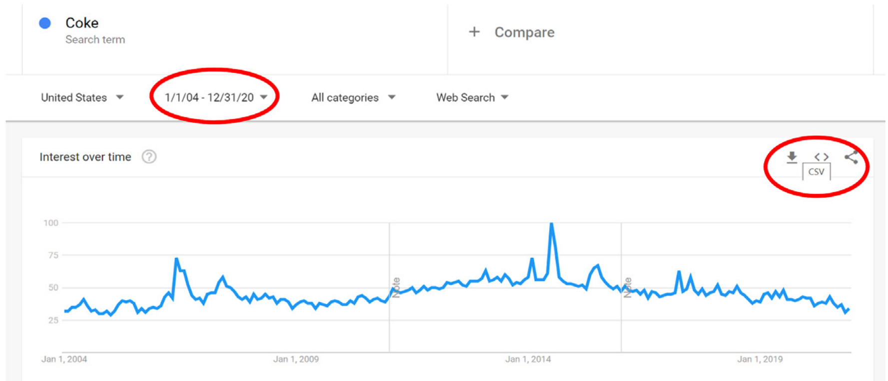
The region can be a whole country, state or a city. We focus only on the United States.



We can also choose the time range for the search.



In our sample, we choose a specific time period from January 2004 to December 2020. Google trends returns the following screenshot. To download the *SVI*, click the arrow in the upper right corner, and it will show the "CSV". Clicking on this downloads the CSV format file of Coke *SVI* from 1/1/2004 to 12/31/2020.



Appendix B. Definitions of Variables

Variable	Definition
ΔSales$_t$	Seasonal change in firm sales divided by previous four-quarter sales, ((*[SALEQ]*$_t$-*[SALEQ]*$_{t-4}$)/*[SALEQ]*$_{t-4}$).
ΔSales$_{t-1}$	One-quarter-lagged scaled seasonal change in sales, (*[SALEQ]*$_{t-1}$-*[SALEQ]*$_{t-5}$)/*[SALEQ]*$_{t-5}$).
ΔSVI	Seasonal change in firm Google search volume index (*SVI*) divided by the previous four-quarter, *SVI* ((*SVI*$_t$-*SVI*$_{t-4}$)/*SVI*$_{t-4}$).
AccRev	Seasonal change in accounts receivable divided by previous four-quarter total assets, ((*[RECTQ]*$_t$ -*[RECTQ]*$_{t-4}$)/*[ATQ]*$_{t-4}$).
Age	The number of years since listing in the CRSP database.
AgeSqr	Square of *Age*.
Allowance	Allowance for uncollectible accounts scaled by gross accounts receivable, (*[RECDQ]*$_t$/(*[RECTQ]*$_t$+*[RECDQ]*$_t$)).
AMUP	An indicator variable set to 1 if a firm is in the bottom *ΔSales* quartile four quarters ago but in the highest *ΔSales* quartile currently.
BIG4	An indicator variable which equals 1 if a firm is audited by a Big Four auditor and 0 otherwise.
BTM	The book value of equity *[CEQQ]*$_t$ divided by the market value of equity (*[PRCCQ]*$_t$**[CSHOQ]*$_t$) at the end of the most recent fiscal quarter for which the data are available.
DefRev	Seasonal change in deferred revenue divided by the previous four-quarter total assets, ((*[DRQ]*$_t$ -*[DRQ]*$_{t-4}$)/*[ATQ]*$_{t-4}$).
DiscretionaryRev	The regression residual based on the conditional revenue model of Stubben (2010), as shown in our Equation (4).
F_Score	100 times the fraud detection score estimated from Dechow et al. (2011), adapted for quarterly frequency.
GM	Gross margin, sales minus cost of goods sold scaled by sales, (*[SALEQ]*$_t$-*[COGSQ]*$_t$)/*[SALEQ]*$_t$).
GMSqr	Square of *GM*.
GrossRec	Gross receivables divided by lagged total assets, (*[RECTQ]*$_t$+*[RECDQ]*$_t$)/*[ATQ]*$_{t-1}$).
GrowNeg	An indicator variable which equals 1 if industry-median- adjusted revenue growth is negative and 0 otherwise.
GrowPos	An indicator variable which equals 1 if industry-median- adjusted revenue growth is positive and 0 otherwise.
HMUP	An indicator variable set to 1 if a firm is in the bottom change in headcount quartile but in the highest *ΔSales* quartile currently.
IO_Own	The fraction of shares owned by institutions.
Lev	Total liability divided by the book value of equity, (*[ATQ]*$_t$-*[CEQQ]*$_t$)/*[CEQQ]*$_t$).
Loss	An indicator variable which equals 1 if the firms has negative income before extraordinary items *[IBQ]*$_t$ and 0 otherwise.
Misstate_Rev	An indicator variable which equals 1 for quarters in which firms are identified as making an upward misstatement of revenues and 0 otherwise.
MUP	An indicator variable. For each calendar quarter and industry, all firms are sorted into quartiles based on the magnitude of *ΔSVI* and separately into quartiles based on the magnitude of *ΔSales*. Firms in the bottom *ΔSVI* quartile but the top *ΔSales* quartile are coded 1; 0 otherwise.
OPCycle	The length of operating cycle, measured using COMPUSTAT annual file, (360/(*[SALE]*$_t$/.5*(*[RECT]*$_t$+*[RECT]*$_{t-1}$))+ 360/(*[COGS]*$_t$/.5*(*[INVT]*$_t$+*[INVT]*$_{t-1}$)).
Past Ret	The returns over the past12 months.
RecTurnover	Gross receivable turnover, (*[SALEQ]*$_t$/(*[RECTQ]*$_t$+*[RECDQ]*$_t$)).
Ret_Vol	The standard deviation of the monthly stock returns in the prior year.
ROA	Return on total assets, the net income before extraordinary items [IBQ]$_t$ divided by previous quarter total assets, (*[ATQ]*$_{t-1}$ (*[IBQ]*$_t$/*[ATQ]*$_{t-1}$).
Sale_Vol	The standard deviation of annual sales (*[SALE]*/average assets*[AT]*) over the last eight quarters, at least three years data are required.
Special	The amount of special items scaled by book value of assets.
StdSales	Standard deviation of the firm's sales scaled by total assets, (*[SALEQ]*$_t$/*[ATQ]*$_t$) over the previous 8 quarters.
Size	The natural log of the market capitalization, (*[PRCCQ]*$_t$**[CSHOQ]*$_t$) (in millions) at quarter end.
#Analysts	Number of analysts following of the company.

* Brackets contain COMPUSTAT item code.

Appendix C. Regressions of Sales Changes on Current and Lagged Changes in SVI Index

Dependent Variable	(1)		(2)		(3)	
	$\Delta Sales_t$		$\Delta Sales_t$		$\Delta Sales_t$	
	Coefficients	t-stats	Coefficients	t-stats	Coefficients	t-stats
ΔSVI_t	0.080***	(8.22)	0.029***	(4.55)	0.022***	(3.41)
ΔSVI_{t-1}			0.005	(1.18)	0.004	(0.81)
ΔSVI_{t-2}			0.008*	(1.78)	−0.003	(-0.61)
ΔSVI_{t-3}			0.006*	(1.70)	0.003	(0.71)
$\Delta Sales_{t-1}$			0.566***	(24.08)	0.609***	(23.15)
$\Delta Sales_{t-2}$			0.132***	(7.87)	0.098***	(4.04)
$\Delta Sales_{t-3}$			0.088***	(4.33)	0.077***	(3.65)
$\Delta Sales_{t-4}$			−0.205***	(-10.89)	−0.214***	(-11.23)
Size					0.004***	(3.20)
BTM					−0.010***	(-3.60)
Lev					0.000	(0.23)
Loss					−0.036***	(-6.95)
BIG4					−0.016***	(-5.18)
OPCycle					−0.000	(-1.45)
Age					−0.001***	(-8.68)
Sale_Vol					0.104**	(2.42)
IO_Own					0.004	(1.28)
Special					0.003	(1.38)
Ret_Vol					0.066	(0.44)
Past Ret					0.037***	(9.82)
#Analysts					−0.000	(-0.76)
# of Obs.	46,739		39,685		30,863	
Adj. R^2	0.06		0.46		0.54	

Note: This table reports estimation results of regressions of $\Delta Sales_t$ on ΔSVI_t, $\Delta Sales_{t-1}$, and control variables. The sample period is from 2004 to 2020. The dependent variable, $\Delta Sales_t$, is the percentage seasonal changes in a firm's quarterly sales. The key independent variable, ΔSVI_t, is the percentage seasonal changes in a firm's *SVI*. $\Delta Sales_{t-1}$ to $\Delta Sales_{t-4}$ are one-quarter- to four-quarter-lagged seasonal changes in sales, respectively, and lagged ΔSVI_{t-1} *to* ΔSVI_{t-3} are one-quarter- to three-quarter-lagged seasonal changes in *SVI*. Control variables including size (*Size*), book-to-market ratio (*BTM*), leverage (*Lev*), a loss indicator (*Loss*), a Big Four indicator (*BIG4*), operating cycle (*OPCycle*), firm age (*Age*), the standard deviation of sales over at least three of the last eight quarters (*Sale_Vol*), institutional ownership (*IO_Own*), special items (*Special*), the standard deviation of the monthly stock returns in the prior year (*Ret_Vol*), the return over the fiscal quarter (*Past Ret*) and the number of analysts following (*#Analysts*) are as defined in Appendix B. Calendar quarter fixed effects are included, and standard errors are clustered by firm and by calendar quarter. *t*-statistics are reported in parentheses. The intercept is included but not tabulated for brevity. *** indicates $p < 0.01$; **$p < 0.05$; *$p < 0.1$.

References

Albrecht, A., Kim, K., & Lee, K. (2020). *Mandatory disclosure and management discretion: On the case of changes in accounting estimates*. Working paper. Texas Christian University.

Allee, K. D., Baik, B., & Roh, Y. (2021). Detecting financial misreporting with real production activity. *Contemporary Accounting Research, 38*, 1581–1615.

American Institute of Certified Public Accountants (AICPA). (2002). *Consideration of fraud in a financial statement audit, statement on auditing standards no. 99*. New York, NY: AICPA.

Amir, E., & Lev, B. (1996). Value-relevance of non-financial information: The wireless communications industry. *Journal of Accounting and Economics, 22*, 3–30.

Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017). Big data and analytics in the modern audit engagement: Research needs. Auditing. *A Journal of Practice & Theory, 36*, 1–27.

Askitas, N., & Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly, 55*, 107–120.

Atwood, T. J., Drake, M. S., & Myers, L. A. (2010). Book-tax conformity, earnings persistence and the association between earnings and future cash flows. *Journal of Accounting and Economics, 50*, 111–125.

Baber, W. R., Kang, S. H., & Kumar, K. R. (1998). Accounting earnings and executive compensation: The role of earnings persistence. *Journal of Accounting and Economics, 25*, 169–193.

Bartov, E., Marra, A., & Momenté, F. (2021). Corporate social responsibility and the market reaction to negative events: Evidence from inadvertent and fraudulent restatement announcements. *The Accounting Review, 96*, 81–106.

Bartov, E., Mohanram, P., & Seethamraju, C. (2002). Valuation of internet stocks-an IPO Perspective. *Journal of Accounting Research, 40*, 321–346.

Beasley, M. S., Hermanson, D. R., Carcello, J. V., & Neal, T. L. (2010). *Fraudulent financial reporting: 1998-2007: An analysis of US public companies*. COSO, Committee of Sponsoring Organizations of the Treadway Commission.

Bell, T., Peecher, M. E., & Solomon, I. (2005). *The 21st century public-company audit: Conceptual elements of KPMG's global audit methodology*. Montvale, New Jersey: KPMG.

Boyle, E. S., Lewis–Western, M. F., & Seidel, T. A. (2021). Do quarterly and annual financial statements reflect similar financial statement error in the post–SOX era? *Journal of Financial Reporting, 6*, 1–31.

Brazel, J. F., Jones, K. L., & Zimbelman, M. F. (2009). Using non-financial measures to assess fraud risk. *Journal of Accounting Research, 47*, 1135–1166.

Budescu, D. V., Peecher, M. E., & Solomon, I. (2012). The joint influence of the extent and nature of audit evidence, materiality thresholds, and misstatement type on achieved audit risk. *Auditing: A Journal of Practice & Theory, 31*, 19–41.

Bulut, L. (2018). Google Trends and the forecasting performance of exchange rate models. *Journal of Forecasting, 37*, 303–315.

Bushee, B. J., Core, J. E., Guay, W., & Hamm, S. J. (2010). The role of the business press as an information intermediary. *Journal of Accounting Research, 48*, 1–19.

Callen, J. L., Robb, S. W., & Segal, D. (2008). Revenue manipulation and restatements by loss firms. *Auditing: A Journal of Practice & Theory, 27*, 1-29.

Caylor, M. L. (2010). Strategic revenue recognition to achieve earnings benchmarks. *Journal of Accounting and Public Policy, 29*, 82–95.

Chakravarthy, J., DeHaan, E., & Rajgopal, S. (2014). Reputation repair after a serious restatement. *The Accounting Review, 89*, 1329–1363.

Cheng, H., Liu, T., & Wei, K. C. (2020). *The effects of financial markets on managerial opportunistic behavior: Evidence from CEOs' tone manipulation* (Working paper).

Chi, S., & Shanthikumar, D. (2016). Local bias in Google search and the market response around earnings announcements. *The Accounting Review, 92*, 115–143.

Choi, H., & Varian, H. (2012). Predicting the present with Google trends. *The Economic Record, 88*, 2–9.

Da, Z., Engelberg, J., & Gao, P. (2011a). *In search of fundamentals* (Working paper).

Da, Z., Engelberg, J., & Gao, P. (2011b). In search of attention. *The Journal of Finance, 66*, 1461–1499.

Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all fears investor sentiment and asset prices. *Review of Financial Studies, 28*, 1–32.

D'Amuri, F., & Marcucci, J. (2017). The power of Google searches in forecasting unemployment. *International Journal of Forecasting, 33*, 801–816.

Dechow, P., Ge, W., Larson, C., & Sloan, R. (2011). Predicting material accounting misstatements. *Contemporary Accounting Research, 28*, 17–82.

Dechow, P., Ge, W., & Schrand, C. (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal Of Accounting And Economics, 50*, 344–401.

Della Penna, N., & Huang, H. (2010). *Constructing consumer sentiment index for US*

using Google searches (Working Paper).
Demerjian, P. R., Lev, B., Lewis, M. F., & McVay, S. E. (2013). Managerial ability and earnings quality. *The Accounting Review, 88*, 463–498.
Drake, M. S., Guest, N. M., & Twedt, B. J. (2014). The media and mispricing: The role of the business press in the pricing of accounting information. *The Accounting Review, 89*, 1673–1701.
Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2012). Investor information demand: Evidence from Google searches around earnings announcements. *Journal of Accounting Research, 50*, 1001–1040.
Dyck, I. J., Morse, A., & Zingales, L. (2023). How pervasive is corporate fraud? *Review of Accounting Studies*, 1–34.
Earley, C. (2015). Data analytics in auditing: Opportunities and challenges. *Business Horizons, 58*, 493–500.
Fang, L., & Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance, 64*, 2023–2052.
n.d. FASB (Financial Accounting Standards Board). Accounting standards codification (ASC). Retrieved January 16, 2013, from https://fasb.org/page/PageContent?pageId=/standards/implementing/revrec/fasb-iasb-resource-group/revenue-recognition-bridge-page.html&bcpath=tfft
Files, R. (2012). SEC Enforcement: Does forthright disclosure and cooperation really matter? *Journal of Accounting and Economics, 53*, 353–374.
Froot, K., Kang, N., Ozik, G., & Sadka, R. (2017). What do measures of real-time corporate sales tell us about earnings surprises and post-announcement returns? *Journal of Financial Economics, 125*, 143–162.
Givoly, D., Hayn, C. K., & Katz, S. P. (2010). Does public ownership of equity improve earnings quality? *The Accounting Review, 85*, 195–225.
Glover, S. M., Prawitt, D. F., & Drake, M. S. (2015). Between a rock and a hard place: A path forward for using substantive analytical procedures in auditing large P&L accounts: Commentary and analysis. *Auditing: A Journal of Practice & Theory, 34*, 161–179.
Guest, N. M. (2021). The information role of the media in earnings news. *Journal of Accounting Research, 59*, 1021–1076.
Hales, J., Moon, J., & Swenson, L. (2018). A new era of voluntary disclosure? Empirical evidence on how employee postings on social media relate to future corporate disclosures. *Accounting, Organizations and Society, 68–69*, 88–108.
Huang, J. (2018). The customer knows best: The investment value of consumer opinions. *Journal of Financial Economics, 128*, 164–182.
Huang, T. C., & Hairston, S. (2021). Analyst revenue forecasts and firm revenue misstatements. *European Accounting Review*, 1–36.
Huang, K., Li, M., & Markov, S. (2020). What do employees know? Evidence from a social media platform. *The Accounting Review, 95*, 199–226.
Hughes, K. E. (2000). The value relevance of non-financial measures of air pollution in the electric utility industry. *The Accounting Review, 75*, 209–228.
Jackson, S. B., & Liu, X. (2010). The allowance for uncollectible accounts, conservatism, and earnings management. *Journal of Accounting Research, 48*, 565–601.
Kang, J., Stice-Lawrence, L., & Wong, Y. (2021). The firm next door: Using satellite images to tease out information acquisition costs. *Journal of Accounting Research, 59*, 713–750.
Lennox, C., Lisowsky, P., & Pittman, J. (2013). Tax aggressiveness and accounting fraud. *Journal of Accounting Research, 51*, 739–778.
Nelson, M., Elliot, J., & Tarpley, R. (2002). Evidence from auditors about managers' and auditors' earnings management decisions. *The Accounting Review, 77*, 175–202.
Nelson, M., Elliot, J., & Tarpley, R. (2003). How are earnings managed? Examples from auditors. *Accounting Horizons, 17*(Supplement), 17–35.
Palmrose, Z., Richardson, V., & Scholz, S. (2004). Determinants of market reactions to restatement announcements. *Journal of Accounting and Economics, 37*, 1–32.
Peecher, M. E., Schwartz, R., & Solomon, I. (2007). It's all about audit quality: Perspectives on strategic-systems auditing. *Accounting, Organizations and Society, 32*, 463–485.
Piotroski, J. D., & Roulstone, D. T. (2004). The influence of analysts, institutional investors, and insiders on the incorporation of market, industry, and firm-specific information into stock prices. *The Accounting Review, 79*, 1119–1151.
Public Company Accounting Oversight Board (PCAOB). (2007). *Observations on auditors' implementation of PCAOB standards relating to auditors' responsibilities with respect to fraud*. Release No. 2001-001, January 22, Washington, DC.
Riley, R., Pearson, T., & Trompeter, G. (2003). The value relevance of non-financial performance variables and accounting information: The case of the airline industry. *Journal of Accounting and Public Policy, 22*, 231–254.
Sharma, R. (2015). *Big data and analytics in the audit process*. Harvard Law School Forum on Corporate Governance.
Srivastava, A. (2014). Selling-price estimates in revenue recognition and the usefulness of financial statements. *Review of Accounting Studies, 19*, 661–697.
Stubben, S. R. (2010). Discretionary revenues as a measure of earnings management. *The Accounting Review, 85*, 695–717.
Suhoy, T. (2009). *Query indices and a 2008 downturn: Israeli data* (Working paper).
Tang, V. W. (2018). Wisdom of crowds: Cross-sectional variation in the informativeness of third-party-generated product information on Twitter. *Journal of Accounting Research, 56*, 989–1034.
Teoh, S. H. (2018). The promise and challenges of new datasets for accounting research. *Accounting, Organizations and Society, 68*, 109–117.
Teoh, S. H., Wong, T., & Rao, G. (1998). Are accruals during initial public offerings opportunistic? *Review of Accounting Studies, 3*, 175–208.
Trotman, K. T., & Wright, W. F. (2012). Triangulation of audit evidence in fraud risk assessments. *Accounting, Organizations and Society, 37*, 41–53.
Trueman, B., Wong, F., & Zhang, X. (2001). Back to basics: Forecasting the revenues of internet firms. *Review of Accounting Studies, 6*, 305–329.
Turner, L., Dietrich, J. R., Anderson, K., & Bailey, A. (2001). *Accounting restatements* (Working paper).
Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: Survey-based indicators vs. google trends. *Journal of Forecasting30*, 565–578.
Zhu, C. (2019). Big data as a governance mechanism. *Review of Financial Studies, 32*, 2021–2061.