

Measuring Business Social Irresponsibility: The Case of Sin Stocks

March 2023

Abstract

Negative screening (of "sin" stocks) is the most common strategy used by socially responsible investors. There is no consensus in the literature whether these exclusions result in higher cost of capital (and hence higher expected returns) for targeted firms. The existing literature identifies sin companies using industry classification codes (IC). We propose an alternative measure of firms' exposure to sin activities (*sinfulness*) based on textual analysis (TA). Sinfulness captures both cross-sectional and time-series variation in firms' exposure to sin activities. The correlation between the IC and TA sin indicators is only 0.69, with twice as many sin stocks in TA than in IC. TA reveals several important false positive and numerous false negative sin stocks in IC. While the number of publicly listed sin-related stocks has declined by 43% between 1997 and 2021, their total market capitalization has increased almost threefold from about \$200bn to \$600bn during the same period. A sin-weighted portfolio of sin stocks earns an annualized Fama-French 6-factor alpha of 4%. Overall, our study highlights important shortcomings of using IC to identify sinful firms and resurrects the sin premium, that is, more sinful stocks have higher expected returns.

Keywords: Business social responsibility, Sin stocks, Textual analysis.

JEL Classification Codes: G11, G12, G14, D71.

1 Introduction

Negative exclusionary screening refers to an investment strategy in which socially controversial firms in particular sectors are excluded from the portfolio. Global Sustainable Investment Review reports that, in 2020, more than \$15 trillion (43% of total sustainable investments) were invested using negative screening. The most common negative screen employed is excluding "sin" stocks (Renneboog et al., 2011; Luo and Balvers, 2017), a term used for companies that make money from exploiting human weaknesses and vices and generally include firms in the tobacco, alcohol, gambling, and adult entertainment industries.¹ Consequently, the performance of sin stocks has received significant attention from both practitioners (Ahrens, 2004; Waxler, 2004) and academics (Fabozzi et al., 2008; Statman and Glushkov, 2009; Hong and Kacperczyk, 2009; Blitz and Fabozzi, 2017; Zerbib, 2022).

The prior literature identifies sin stocks based on industry classification codes (IC) assigned to companies by data providers (e.g., Compustat or CRSP). Several studies have documented lower current valuations and higher abnormal positive returns for sin stocks with different proposed mechanisms, including being underpriced because they are shunned or having higher litigation risk (Fabozzi et al., 2008; Statman and Glushkov, 2009; Hong and Kacperczyk, 2009). Some other studies question the existence of the sin premium. For example, Blitz and Fabozzi (2017) show that the excess return of sin stocks is fully explained by their exposure to profitability and investment factors, and Eccles et al. (2022) find that sin stocks do not have lower (current) valuations after controlling for differences in fundamentals. Similarly, Berk and van Binsbergen (2021) find no effect of divestment by socially responsible investors on the cost of capital of targeted firms and argue that "the sin premium could well be attributable to risk not correctly adjusted for" in earlier papers in the literature. In addition to studies that focus directly on the performance of sin stocks, many other studies closely follow Hong and Kacperczyk (2009), hereafter HK (2009), to identify, study, or control for socially irresponsible (sin) stocks using industry classification codes (Kim and Venkatachalam, 2011; Luo and Balvers, 2017; Pedersen et al., 2021; Zerbib, 2022).

Although simple and straightforward, identifying sin stocks based on the IC method has several important shortcomings. First, industry classifications are somewhat arbitrary. For example, the four-digit SIC codes assigned by CRSP and Compustat match only for 32% of companies during our sample from 1996 to 2021. Consistently, sin indicators created based on industry codes from CRSP and Compustat have a correlation of only 0.80. Importantly, as we will discuss, none of these two classifications is strictly better in identifying true sin

¹For example, Renneboog et al. (2011) report that more than 90% of socially responsible investment funds in the U.S. use sin screens.

stocks and many sin stocks would not be counted sin in either CRSP or Compustat. Second, the IC method yields a binary outcome, i.e., a firm would be classified as either sinful or not sinful. Consequently, given the above-mentioned differences in the codes assigned across data providers, the results might be very sensitive to the choice of the classification source. Moreover, the binary nature of sinfulness in the IC method does not allow for studying whether "more sinful" firms have higher stock returns, something one might expect regardless of the underlying mechanism (e.g., litigation risk, having lower investor base, etc.) behind the sin premium. Third, precisely because of the binary measure of sin stocks based on the IC method, both equal-weighted and especially value-weighted portfolios of sin stocks could be misleading. Equal weighting assumes that all sin stocks are equally sinful, measuring the sin premium as an average across all sin stocks regardless of their exposure to sinful business activities. In value weighting, a large company is assigned a significant weight, even if it is falsely identified as a sin stock or has a small sin-related business segment. Lastly, due to the lack of a clear industry classification, sin stocks from the adult entertainment industry are excluded from previous studies (HK, 2009; Blitz and Fabozzi, 2017).

In this paper, we propose an alternative approach based on textual analysis (TA method) of corporate annual reports that overcomes the above-mentioned shortcomings. Specifically, we use web crawling to download the universe of annual reports available from the US Securities and Exchange Commission's (SEC) Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system. To identify sin stocks, we focus on tobacco, alcohol, gambling, and adult entertainment industries. The first three are always considered sin in the literature and more than 80% of socially responsible investment (SRI) funds use these screens (Luo and Balvers, 2017). We add adult entertainment to this list as this is the fourth commonly accepted sin industry (and is a common sin screen used by SRI funds as reported in Luo and Balvers, 2017), but it is excluded from previous studies only because there is no way to identify firms in this sector using industry classifications codes.² We create a sin dictionary including words and phrases related to each sin industry that are likely to appear in annual reports of firms engaged or exposed to sinful activities. Our dictionary includes not only unigrams (single words), but also n-grams (a sequence of n words). Examples from our sin dictionary are *tobacco* and *cigarette* for tobacco, *alcohol* and *alcoholic beverages* for alcohol, *wagering*, *casino*, and *gaming industry* for gambling, and *adult business* and *sexually explicit* for adult entertainment. Our method is based on the simple premise that the higher the

²Most studies drop weapons as a sin industry, particularly when studying U.S. corporations (HK, 2009; Luo and Balvers, 2017; Blitz and Fabozzi, 2017). Furthermore, exclusionary restrictions are increasingly applied to some other companies (e.g., fossil fuel firms). Although such companies may not be beneficial to the environment, they are not typically considered sin firms in the sense that they do not make money from human vices.

exposure of a firm to a particular sin industry, the higher the frequency of sin-related words in its filing. To properly identify sinfulness, it is important to search throughout the report as opposed to just the business section as companies often discuss their exposure to sinful activities in other sections such as the one related to risk factors and litigation.

Our analysis is based on 117,250 annual filings for the fiscal years 1996-2021. The number of sin stocks in our TA method is almost twice as large as the one obtained using the IC method. Analyzing the differences, we document significant issues in identifying sin companies using the IC method. We find six major false positive sin stocks based on the IC method and numerous cases of false negative stocks where a company's business is clearly in a sin industry, but the industry code assigned to the company does not identify it as a sin stock. Even for some well-known sin firms (e.g., Churchill Downs), the IC method identifies stocks as sin in some years and not in other years, without any justifiable reason and even though the company's business has not changed at all. The TA method also measures reasonably well changes in sinfulness of companies over time within a sin industry (e.g., a tobacco firm acquires more tobacco related businesses) and captures expansions to other sin industries (e.g., an alcohol company expands to the tobacco industry). Furthermore, unlike the IC method, the TA approach identifies nine unique companies in the adult entertainment industry. Overall, these results highlight significant weaknesses in relying on industry codes to identify sin stocks.

We use both the IC and the TA method to investigate the existence of the sin premium using equal- and value-weighted portfolios (as is common in the literature) as well as a sin-weighted portfolio. Sin weights are computed using the sinfulness measure we obtain from the frequency of sin-related words in the filings. In a sin-weighted portfolio, companies that are more exposed to sinful business activities receive a higher weight in the portfolio construction. We do not find a sin premium in the equal- and value-weighted portfolios of sin stocks identified by our TA method. However, in the sin-weighted portfolio, we observe a positive and significant annual excess return of 4% (over comparables and after controlling for common risk factors), almost twice as large as previously estimated by [HK \(2009\)](#). When we use the IC method to identify sin stocks, we also find no sin premium in the value-weighted portfolio. We do, however, observe a positive excess return for the IC method in equally-weighted portfolios; that excess return increases by about 25% when using a sin-weighted portfolio. Overall, these results resurrect the sin premium and highlight that the premium is mainly driven by companies that are more involved in sinful activities.

Our study contributes to the active debate on the performance of socially irresponsible firms and whether exclusionary screens lead to higher cost of capital (and hence higher expected returns) for the targeted firms ([Heinkel et al., 2001](#); [HK, 2009](#); [Chava, 2014](#); [Blitz](#)

and Fabozzi, 2017; Berk and van Binsbergen, 2021; Zerbib, 2022; Eccles et al., 2022; Broccardo et al., 2022; Edmans et al., 2022). Although several studies have documented a sin premium, i.e., that socially irresponsible firms have higher (expected) returns (HK, 2009; Zerbib, 2022), other works suggest that such a finding is due to other factors such as uncontrolled risk factors or differences in fundamentals (Blitz and Fabozzi, 2017; Berk and van Binsbergen, 2021; Eccles et al., 2022).³ Our study highlights important issues with using industry codes to identify irresponsible firms, proposes a novel approach to measuring sinfulness, and its findings resurrect the sin premium.

Our work is also related to a rapidly growing literature that uses textual analysis to measure (and study) difficult-to-measure firm-level metrics such as financial constraints (Hoberg and Maksimovic, 2014; Bodnaruk et al., 2015), geographic dispersion (García and Norli, 2012), corporate culture (Li et al., 2020), corporate innovation (Bellstam et al., 2021), competitors and competitive threats (Hoberg et al., 2014; Hoberg and Phillips, 2016), or the information content of the IPO prospectus (Loughran and McDonald, 2013; Hanley and Hoberg, 2010). In terms of methodology, our work is similar to studies that use a dictionary-based method (Tetlock, 2007; Tetlock et al., 2008; Feldman et al., 2009; Hanley and Hoberg, 2010; Loughran and McDonald, 2011; García and Norli, 2012; Huang et al., 2013; Muslu et al., 2015; Bochkay et al., 2019; Loughran and McDonald, 2020) rather than studies that use machine learning techniques (Buehlmaier and Whited, 2018; Li et al., 2020; Bertomeu et al., 2020; Azimi and Agrawal, 2021; Bellstam et al., 2021; Huang et al., 2023).⁴

Several previous studies have noted the somewhat arbitrary and inflexible nature of industry classifications (Kahle and Walkling, 1996; Hoberg and Phillips, 2016). Hoberg and Phillips (2016) propose an alternative classification in which companies and their competitors are grouped according to the similarity of product descriptions in their annual reports. Rather than grouping competitors, our work aims at identifying socially irresponsible firms and measuring their exposure to sin industries.

The remainder of this article is structured as follows. Section 2 describes our data and sample construction, Section 3 discusses our methodology, Section 4 presents the empirical results, and Section 5 concludes.

³Somewhat similarly but in the context of carbon emissions, Bolton and Kacperczyk (2021) find a "carbon premium", whereas Aswani et al. (forthcoming) show that this association between carbon emissions and stock returns disappears if one uses *actual* disclosed emissions rather than those *estimated* by data vendors. Hsu et al. (2023) report a significant pollution premium for firms with high toxic emissions.

⁴Given our context in which the creation of an appropriate dictionary is relatively easy, we do not see any value-added by going towards more complex and less transparent methods. Henry and Leone (2016) evaluate different methodologies and conclude that domain-specific dictionary-based approaches often perform as well as (or better than) more complex machine learning techniques.

2 Data and sample construction

We use web crawling to gather all annual reports from the SEC EDGAR platform. Monthly stock returns and market capitalizations are obtained from the Center for Research in Security Prices (CRSP). Furthermore, we obtain industry classification codes (SIC/NAICS) on the firm and segment level from Compustat. In order to conduct our sin identification approach, we require for each firm year to have a "valid" 10-K filing (described below), 10 valid returns, a non-missing market capitalization for June, and at least basic accounting data including an industry classification code in Compustat. Following the common practice in the asset pricing literature (e.g., [Bali et al., 2016](#)), we consider only common stocks (share codes 10 and 11) traded on NYSE, AMEX, and NASDAQ (exchange codes 1, 2, and 3). We adjust returns for delisting following [Shumway \(1997\)](#) and [Shumway and Warther \(1999\)](#). Our sample period ranges from 1996 to 2021, since it was not mandatory to file annual reports electronically before May 6, 1996.

A valid 10-K filing is any kind of annual report with at least 2,000 words (after removing the stopwords such as "a", "and", etc.) in the respective filing. We discard filings with less than 2,000 words following [Loughran and McDonald \(2011\)](#). We consider regular annual reports (10-K), annual reports with checked item 405 (10-K405), annual reports for smaller businesses (10-KSB), and transitional annual reports (10-KT) for companies that change their fiscal year. When companies have multiple annual reports in the same year, we keep the main filings and only use amendments (e.g., 10-K/A) when there is no valid main filing. When there are multiple main filings (i.e., a 10-K and a 10-KT or a 10-KSB), we prioritize the regular annual reports (10-K). If there are still years with multiple main filings, we keep the file that was filed later that year. Any remaining duplicates are treated as identical duplicates and are removed accordingly. For portfolio analysis, we also require that the filing would have been available in real time when constructing portfolios in June each year. We use a linking table to improve the merge of the 10-K filings with the CRSP/Compustat merged file (which is typically done using Central Index Key (CIK)). This mitigates mapping problems that occur when a company's CIK has changed over time, as Compustat only maintains the latest CIK for each company and replaces the historical one with the most recent one. This procedure leads to a final data set of 13,740 companies with 117,250 annual reports from 1996 to 2021.

3 Methodology

3.1 A text-based measure of sinfulness

We create a dictionary of words and phrases related to each sin industry, i.e., tobacco, alcohol, gambling, and adult entertainment. To create the dictionary, we follow [Loughran and McDonald \(2011\)](#) and carefully analyze unigrams or single words, but also bigrams and trigrams (a sequence of two or three words) that appear in annual reports, are related to sin activities, and are unlikely to appear in other contexts than what we aim to capture. For example, while the word *gaming* is often used in gambling companies' reports, it is not part of our dictionary as it would also appear in any company that designs or produces video games. Instead, we have the bigram of *gaming industry*, which is a phrase often used to refer to the gambling industry.⁵ We also categorize sin words within each industry into three to four clusters to ensure that we capture different aspects of each business, something that is also used (as will be discussed later) in our algorithm to define the list of sin-related stocks. For the gambling industry, the clusters and some examples of words that belong to each cluster (shown in parentheses) include: general gambling (*gambling, betting*), gambling equipment (*slot machines, gaming machines*), gambling establishments (*casino, horse racetrack*), and types of gambling (*sports betting, lottery*).

After creating the dictionary, we searched through all 117,250 annual reports to count the frequency of sinful words in each filing. Our idea is based on the simple premise that companies that are more engaged or whose businesses are more related to sin activities will have a higher count of sin-related words in their reports. We note that this idea would almost surely generate absurd outcomes if applied to some particular aspects of business social responsibility, such as measuring diversity and inclusion in the sense that company reports with high counts of the word *diversity* are uninformative of how much the company truly cares about diversity. However, in our particular context of identifying companies that have exposure to sin activities, due to significant regulations, disclosure requirements, and litigation risks, companies do reveal their involvement in these activities. However, such information could appear anywhere in the filing, including the business section, management discussion and analysis (MD&A), risk factors, legal proceedings and litigations, etc. This is why it is important to search for sin words throughout annual reports rather than in a particular section of it.

To include a company as sin-related, we set the following (rather strict) three requirements to make sure that every sin company has a non-negligible exposure to sinful activities:

⁵From now on, we simply use "words" to refer to both words and phrases in our dictionary.

(1) A sin word would be counted in a filing only if it appears at least five times in the filing. Otherwise, we pretend that it did not exist. (2) The filing must include at least two unique words from two different clusters of words within each industry, as explained above. (3) The frequency of sin-related words in the filing has to be at least 20. For example, consider a company whose filing includes 2 counts of *gambling*, 20 counts of *slot machines* and 15 counts of *gaming machines*. Since *gambling* has appeared only three times, we ignore it following condition (1). Although the two other words (*slot machines* and *gaming machines*) have appeared 35 times in total (hence satisfying condition 3), they are both from the same cluster of words within gambling. Consequently, condition (2) above does not hold and hence the company would not be considered sin. All information about the dictionary, the sin industry-specific word clusters, descriptive statistics related to each word, and further detailed information on our algorithm are provided in Table A.1 of the Appendix.

Similarly to most related studies (Tetlock, 2007; Tetlock et al., 2008; Feldman et al., 2009; Huang et al., 2013), we use word frequency with equal weighting instead of term frequency-inverse document frequency (*tf-idf*). In the latter, which is mainly used for information retrieval tasks (e.g., Google search engine), more common words in the sample get relatively lower weights and vice versa. Although in contexts with large dictionaries and a high likelihood of ambiguous and misclassified words (e.g., sentiment analysis), *tf-idf* might be preferred, using it in our study makes little sense. Moreover, Henry and Leone (2016) compare alternative methodologies and conclude that although more complex techniques are potentially advantageous in certain contexts, equal-weighted, domain-specific, word-frequency measures are often as powerful (or better). Furthermore, as argued in Henry and Leone (2016), instead of *tf-idf* weighting a better approach is to exclude potentially problematic words from the list. This is indeed something we have carefully thought about and did not, for example, include words such as "smoke" or "gaming" in our dictionary, as these words might be used in many different non-sin-related contexts.

3.2 Identifying sin stocks using industry classification

For the sake of comparison, we closely follow HK (2009) to create sin indicators based on industry classification (IC) codes. In doing so, we also highlight important practical points for researchers who would like to use such an approach in their work.

We follow HK (2009) in identifying the *triumvirate* of sin stocks, i.e., alcohol, tobacco, and gambling, mainly based on the stock's Fama and French (1997) 48 industry classification code (hereafter FF48). Thus, we consider a firm to be an alcohol company if its Standard Industrial Classification (SIC) code belongs to the FF48's Beer group, i.e., the firm has a

SIC code of 2082 (malt beverages), 2083 (malt), 2084 (wine), or 2085 (distilled and blended liquors). However, contrary to the established literature, we do not consider the SIC code 2080 (beverages) when defining our alcohol industry. Our TA approach revealed that most companies with a SIC of 2080 are false positives (e.g., Coca-Cola and PepsiCo with Compustat SIC codes of 2080 from 1996-2010 and 1996-1998; 2004-2021, respectively), in the sense that they do not contain any or very little alcohol related words in their annual reports (10-K filings), suggesting that these companies are just sellers/producers of non-alcoholic beverages.

In addition, we consider a firm to be a tobacco company if the SIC code belongs to the [FF48](#)'s Smoke group, i.e., the firm has a SIC code within the range of 2100-2199 (tobacco products). Following [HK \(2009\)](#), we use the North American Industry Classification System (NAICS) to identify gambling firms, as the [FF48](#) industries do not distinguish gambling firms from those of the entertainment and hotel industry. Therefore, we consider a firm to be a gambling company if the firm has a NAICS code of 7132, 71312, 713210, 71329, 713290, 72112, or 721120. Finally, we identify a firm as an alcohol, tobacco, or gambling company if any of its business segments (BUSSEG) in the Compustat segment file has a SIC or NAICS code as defined above.

While executing the IC method to identify sin companies, we noted important crucial decisions a researcher faces (which is rarely discussed, if any, in the literature) and highlight the best practices below. First, we note that a companies' business often changes, i.e., a company goes into sin activities or divests from them, and therefore, although sometimes done by researchers, it is not generally advisable to forward- or back-fill a SIC- or NAICS-based sin industry classification. Back-filling based on the latest SIC/NAICS codes (header variables) causes issues for companies that have no more sin activities at the time of the classification, but used to be a sin company in the past. These earlier sin years would be neglected, even though historical SIC or NAICS codes would have been able to identify the company as a sin stock for only certain years in the past. A similar problem occurs when researchers forward-fill, i.e., they consider a company to be sinful if at any point in time a historical SIC or NAICS code identified a firm as a sin stock. As both practices may distort the assessment of the pricing of sin stocks, we advise using historical SIC and NAICS codes when possible. We note that back-filling can be helpful, especially when dealing with missing segment data (see, e.g., [HK, 2009](#)). However, researchers should keep in mind that doing so creates a trade-off between false negatives and false positives, which should be evaluated in the context of the respective research design before deciding to back-fill an industry classification. Since our study does not suffer from missing data issues, we always use historical SIC and NAICS codes (missings are replaced with the current code) and do

not forward- or back-fill sin industry classifications.

Furthermore, we note that a company’s SIC code might be different depending on the source. Reviewing our IC approach based on historical SIC codes from SEC-filings, Compustat, and CRSP revealed that the choice of the source of SIC codes has a non-negligible influence on the sin industry classification. The moderate correlations within the tobacco classifications (Compustat to CRSP: 0.86; Compustat to SEC: 0.81; SEC to CRSP: 0.79) as well as within the alcohol classifications (Compustat to CRSP: 0.70; Compustat to SEC: 0.67; SEC to CRSP: 0.56) suggest the importance of the SIC code source choice. Interestingly, the correlation within the gambling classification (Compustat to CRSP: 0.84) also suggests a similar relevance for the NAICS source choice.⁶

Two notable exemplary classification differences with respect to the SIC code choice are Coca-Cola and Pyxus International. Coca-Cola would falsely be identified as an alcohol company only when using Compustat SIC codes with the original [HK \(2009\)](#) approach, i.e., including the SIC code 2080 in the alcohol industry. On the other hand, Pyxus International (formerly Alliance One International Inc. and Dimon Inc.) would correctly be identified as a tobacco company for its entire history (1996-2021) in our sample according to CRSP SIC codes, whereas it would never be a tobacco company according to Compustat SIC codes.

Overall, we find that the practice of backward- or forward-filling industry codes for a firm might be quite problematic and that the choice of industry code source has a significant impact on which companies would be considered sin. Although researchers rarely reveal their source of industry codes, we always use Compustat as our source of historical SIC (after correcting for the issue with the code 2080 mentioned above) and NAICS codes both at the company and segment level.⁷

4 Empirical analysis and results

4.1 Descriptive analysis

4.1.1 The evolution of sin stocks

Table 1 presents a summary statistics of the sin firms identified in the TA (Panel A) and the IC (Panel B) methods. For each panel, we report the summary statistics for the frequencies of sin-related words in each industry, as well as the unique number of firms and the number of firm-year observations per industry. As shown, the number of unique firms identified as

⁶The reported correlations do not include any segment-based sin classifications.

⁷Researchers often refer to [FF48](#), implying that they use historical Compustat SIC codes and perhaps turning to historical CRSP SIC codes when Compustat SIC codes are missing.

sin in the TA method is almost twice as large as the one in the IC method (327 versus 153). There are 2,406 sin firm-year observations in the TA compared to 1,258 in the IC method. The difference is not driven by one particular sin industry. The number of unique firms in tobacco, alcohol, and gambling according to TA (IC) are 56 (22), 134 (37) and 145 (97), respectively. Additionally, the TA method identifies nine unique firms in the adult entertainment business, a sector that could not be identified at all based on the IC method.

The fact that more firms are identified as sin in the TA method (compared to IC) should not be surprising. The IC method presumably picks up very sinful companies (as it has to return a binary output), whereas the TA method could include not only highly sinful companies but also those that are relatively less involved in sin-related activities. Indeed, comparing the average counts of sin-related words in Table 1 Panels A and B for each industry confirms that sin companies in the IC method tend to more sinful companies. For example, the average frequency of tobacco-related words for sin TA firms in Panel A is 348 versus 598 for sin IC firms in Panel B. Similar patterns exist for the alcohol (174 versus 348) and the gambling industry (300 versus 339), respectively.⁸

However, as will be shown in detail later on, the difference between sin TA and sin IC firms is not only due to the fact that there are less sinful companies included in the TA method. We will show that a significant amount of firms that are identified as sin stocks by the TA method are indeed true sin companies (that have been assigned other non-sin related industry codes by Compustat). However, sin companies identified by the IC method are not all sin companies, that is, some are false positive sin stocks. This could already be inferred from the statistics provided in Panel B of Table 1, as there are sin companies with 0 counts of sin-related words in all three industries. In other words, a company with zero counts of sin-related words is unlikely to be a sin company, which we will confirm case by case in Section 4.1.3.

[Please insert Table 1 near here]

To better understand the evolution of sin stocks over time, Figure 1 plots the number of sin stocks within each industry, as well as their total number over time. Panel A shows these statistics for sin TA companies. As shown, the total number of sin-related stocks has declined by more than 40% (from 123 to 70) during our sample. Although we observe a drop in all industries during the sample period (except adult entertainment, which increases and then drops to where it started), the most significant decline is seen for gambling companies. The graph also plots (the gray area) the total market capitalization of sin-related stocks over

⁸We note that one advantage of the TA method is precisely its ability to identify different degrees of sinfulness, as will be discussed in next sections.

time. Perhaps surprisingly, despite a large drop in the number of sin companies, their total market cap increased almost threefold from \$215bn in 1996 to \$572bn in 2021. Panel B of Figure 1 shows similar statistics for sin stocks in the IC method. Overall, other than the lower number of sin stocks in this plot, similar patterns are visible as in Panel A. That is, sin stocks have become much lower in number but much larger in size in the past three decades.

Table 2 presents the underlying data behind Figure 1 (the number of sin stocks per year and industry in the two methods) as well as the statistics showing the overlap between the two methods. In addition to the conclusions we draw from Figure 1, an interesting observation is the number of sin stocks in the adult entertainment business. The number of stocks in adult entertainment for most of years in the 2000s has been in the range of 4-6 stocks, which is comparable with 7-8 stocks in the tobacco business. This clearly suggests that the excess return of an equal-weighted sin portfolio (as in HK, 2009) could be highly sensitive to the inclusion of this sector. Analyzing the differences, while most sin stocks in the IC method are also identified as sin in the TA method, there are, quite surprisingly, 57 firm-year cases during the sample where a company shows up as sin only in the IC method. We will argue later on that these are false positive sin stocks identified in the IC method.

[Please insert Figure 1 near here]

Table A.2 in the Appendix provides a complete list of all companies identified as sin stocks by both the IC and TA method (Panel A), only by the TA method (Panel B) and only by the IC method (Panel C).

4.1.2 Changes in sinfulness within and across sin industries

To highlight the power of our proposed sinfulness measure, we present below some cases in which companies' business changes (by either expanding their activities within a sin business or to a new sin industry). We also show how our measure of sinfulness captures these changes.

(a) *Pyxus International, Inc.*

Pyxus International was formerly known as Alliance One International, Inc. and was founded in 1873. The company has an enterprise value of about \$1bn (by the end of 2022). Its main business has been in the tobacco industry, clearly highlighted in the company's annual report:

Historically, Pyxus' core business has been as a tobacco leaf merchant, purchasing, processing, packing, storing and shipping tobacco to manufacturers of cigarettes and other consumer tobacco products throughout the world. ... We purchase

tobacco in more than 35 countries and ship to approximately 90 countries. ... our primary customers are major consumer tobacco product manufacturers. (Pyxus International, Inc. 10-K filing, 2018)

In fact, historically, the company appears among tobacco companies according to our measure of sinfulness. In 2019, we observe a huge jump in the frequency of tobacco-related words (see, Figure 2 (a)), which is driven by the company’s decision to diversify into new businesses, i.e., vapor and cannabis.

[Please insert Figure 2 near here]

Specifically, Pyxus acquired interests in Purilum, Nicotine River, and Humble Juice, all involved in the development, production, and sale of consumable liquid nicotine and e-liquids to manufacturers and distributors of vapable products. As mentioned in the Pyxus reports, all these companies are registered with the “US Food and Drug Administration (FDA) as manufacturers of tobacco products (nicotine containing vaping products).” Furthermore, Pyxus acquired significant equity positions in Criticality LLC, Island Garden Inc., FIGR East, FIGR Norfolk, all engaged in the cannabis business. In fact, the company discusses in detail the significant risks associated with expanding its traditional tobacco business to these new lines of business, suggesting an increased exposure to sinful business activities. The decline in the counts of tobacco-related words in the following years is also due to the disposal of some of those newly acquired equity interests (e.g., Humble Juice).

Interestingly, this is a company that, as we will explain in more detail later, would not be considered sin in any year over its history according to the IC method.⁹ Our TA approach not only correctly identifies the company as being significantly involved in the tobacco business, but also captures changes in the company’s business in going deeper within this industry.

(b) Loews, Corp.

Loews, Corp., a constituent of the S&P 500, was historically a holding company with subsidiaries spanning diverse sectors (e.g., property, insurance, hotels, offshore oil and gas drilling rigs, distribution and sale of watches and clocks), including the production and sale of cigarettes through its wholly owned subsidiary Lorillard, Inc. In 2008, Loews sold its entire ownership interest in Lorillard. As a result, the company’s frequency of tobacco-related words declined sharply in 2008 (see, Figure 2 (a)), approaching zero. The company states after the separation from Lorillard:

⁹Surprisingly, Capital IQ, which, similar to Compustat, is owned by S&P, assigns a NACE code (the industry classification standard used in the European Union) corresponding to Manufacture of Tobacco Products to the company.

As a result of our ownership of Lorillard prior to the Separation, from time to time we have been named as a defendant in tobacco-related lawsuits. We are currently a defendant in four such lawsuits and could be named as a defendant in additional tobacco-related suits, notwithstanding the completion of the Separation. (Loews, Corp. 10-K filing, 2009)

The company clearly communicates that the sinful business activities have been discontinued, but that there is still an exposure to tobacco litigation risk due to its previous involvement. Our sinfulness measure is able to reflect this fine-grained development.

(c) Beam, Inc.

Beam, Inc., in 1996 still known as American Brands, Inc., was a holding company with subsidiaries engaged in various businesses including the manufacturing and sale of cigarettes, cigars and smoking tobaccos, distilled spirits, various types of hardware and home improvement products, golf and leisure products and office products, supplies and accessories. The company made strategic choices to focus more on alcohol. Specifically, the company sold its domestic tobacco subsidiary, the American Tobacco Company, in 1994. As a result, the company's tobacco activities were significantly reduced in the following years, as can be seen in a sharp drop in the frequency of tobacco-related words (see, Figure 2 (b)). In 2011, the company sold and spun off all other segments, resulting in the continuation of Beam as an independent publicly traded pure-play spirits company. Interestingly, according to the TA method, the company continued to be exposed to tobacco-related activities throughout its history as the frequency did not reach zero. Investigating the company's filing is consistent with the continued exposure to tobacco-related activities:

Numerous legal actions are pending in various jurisdictions against leading tobacco manufacturers based upon allegations that cancer and other ailments have resulted from tobacco use. The Company has been named in some of these actions relating to tobacco products made and sold by former subsidiaries. It is not possible to predict the outcome of pending tobacco-related litigation, and it is possible that some of these actions could be decided unfavorably. (Beam, Inc. 10-K filing, 2013)

This case clearly demonstrates how our sinfulness measure captures the reduction in sinfulness of a company that exists from one (out of two) sinful businesses it is engaged in.

(d) Constellation Brands

While the previous case (Beam Inc.) was a situation in which a company exits from one of its sinful industries, there are situations where a firm increases its engagement from one sin

sector to two sectors. Such cases are also easily picked up by our approach. One such example is Constellation Brands, a well-known alcohol company for which we observe a significant increase in tobacco-related words in recent years (see, Figure 2 (b)). When investigating the reports, the company has indeed started to expand to the cannabis business.

Specifically, in 2017, Constellation Brands acquired significant ownership in Canopy Growth Corp. a leading provider of cannabis products. The company does highlight significant operational, strategic, and financial risks related to this investment (in a "sinful" business):

This investment could affect consumer perception of our existing brands and our reputation with various constituencies. ..The restrictions on advertising, marketing, and the use of logos and brand names could have a material adverse effect on Canopy's business, liquidity, financial condition, and/or results of operations, and our investment in Canopy.. Although the Agriculture Improvement Act of 2018 has taken hemp and hemp derived cannabinoids out of the most restrictive class of controlled substances, marijuana is a schedule-1 controlled substance in the U.S. and is currently illegal under U.S. federal law. Even in those U.S. states in which the recreational use of marijuana has been legalized, its use remains a violation of U.S. federal law. Since U.S. federal laws criminalizing the use of marijuana preempt state laws that legalize its use, continuation of U.S. federal law in its current state regarding marijuana would likely limit the expansion of Canopy's business into the U.S. (Constellation Brands 10-K filing, 2017)

(e) *Molson Coors Beverage, Co.*

Another similar case as Constellation Brands, is Molson Coors Beverage Company, which is another well-known alcohol company. We also observe an increase in the counts of tobacco-related words since 2017 for Molson Coors (see, Figure 2 (b)). According to the company's 2018 filing, a wholly-owned subsidiary within their Canadian business completed the formation of an independent Canadian joint venture with HEXO, which is a cannabis company. As a result, in addition to Molson Coors' alcohol business, the company started to be subject to the litigation and reputational risk of being in the cannabis business as well, which is described in detail in the company filing:

Our Americas business faces numerous risks relating to its joint ventures in the Canadian cannabis industry.. In recent years, there has been an increase in public and political attention on health and well-being as it relates to ...cannabis. Negative publicity and changes in consumer perceptions in relation to CBD or other cannabinoid beverages could adversely affect the sale and consumption of our products which could, in turn, adversely affect our business and financial conditions. Additionally, the concerns around CBD and cannabis as well as health

and well-being could result in unfavorable regulations or other legal requirements in certain of our markets, such as advertising, selling and other restrictions, increased taxes associated with our sales, or the establishment of minimum unit pricing. Any such regulations or requirements could change consumer and customer purchasing patterns, which could negatively impact our business, results of operations, cash flows or financial condition. (Molson Coors Beverage, Co. 10-K filing, 2018)

Overall, the cases shown in Figure 2 highlight the power of textual analysis to detect not only the involvement of a firm in sinful activities, but also changes in the sinfulness of firms within a particular industry (e.g., a tobacco company expanding or divesting its tobacco-related activities) or across industries (e.g., an alcohol company expanding into tobacco-related activities).

4.1.3 False positives in the IC method

In the previous section, we discussed cases highlighting the overall performance of the TA method in capturing companies' involvement in sinful industries. In this (and the subsequent three sections), we compare the classification results of the IC and TA methods. Specifically, in the current section, we discuss cases identified in the IC method as sin companies that show no (or extremely low) levels of sinfulness according to the TA method. As Panel A of Table 3 shows, there are only six companies (two in alcohol and four in gambling) that are identified as sin in at least one year according to the IC method, but never appear as sin in the TA method. The two alcohol identified companies are Seaboard Corp. and Valero Energy Corp. The gambling companies are Elmers Restaurants Inc., Creative Learning Products Inc., Activision Blizzard Inc., and Santa FE Financial Corp. We argue below that all these are cases of false positives identified based on the IC method.

[Please insert Table 3 near here]

The first false alcohol company is Seaboard Corp., a company that exists throughout the sample but is identified as sin in IC only between and including 1997 and 2001. The company is identified as an alcohol segment code (SIC code of 2084) appears in those years in the Compustat segment files. However, this company is surely not an alcohol company. For example, the company's annual report for the fiscal year 1997 filed on March 26, 1998 has no single mention of alcohol or alcohol-related words. According to the company's filing, Seaboard Corporation "is a diversified international agribusiness and transportation company which is primarily engaged in domestic pork and poultry production and processing, commodity merchandising, baking, flour milling and shipping. Overseas, the Company is primarily engaged in flour and feed milling, shrimp and produce farming and electric power generation."

Interestingly, during the past few years, the company reports an alcohol segment in its annual filing (although not shown in the Compustat segment database), but as highlighted in the company’s 2021 filing, this is only related to industrial alcohol: ”Seaboard, through its subsidiary, Seaboard Energias Renovables y Alimentos S.R.L., operates a vertically integrated sugar and alcohol production facility in Argentina. The alcohol is marketed to industrial users and sold as dehydrated alcohol to certain oil companies under the Argentine governmental bioethanol program, which requires alcohol to be blended with gasoline.”¹⁰

The second alcohol company identified as sin in the IC but not in the TA method is Valero Energy Corp, which also exists throughout our sample period. The company is identified as sin in the IC method from 2008 to 2016, which is due to the alcohol segment reported in the Compustat segment file. This is misleading as the company started reporting a segment on ethanol and that is the reason for being assigned an alcohol segment code by the Compustat segment dataset. Here is a note from the company’s annual filing: ”Our business is organized into three reportable segments: refining, ethanol, and retail. Prior to the second quarter of 2009, we had two reportable segments: refining and retail. As a result of our acquisition of several ethanol plants during the second quarter of 2009, we now present ethanol as a third reportable segment.” And here is the description of their business according to the filing: ”We are a Fortune 500 company based in San Antonio, Texas....We own and operate 16 refineries located in the United States, Canada, and Aruba that produce conventional gasolines, distillates, jet fuel, asphalt, petrochemicals, lubricants, and other refined products as well as a slate of premium products including CBOB and RBOB1, gasoline meeting the specifications of the California Air Resources Board (CARB), CARB diesel fuel, low-sulfur and ultra-low-sulfur diesel fuel, and oxygenates (liquid hydrocarbon compounds containing oxygen).” Consequently, this is another case of false positive identification of sin stocks using the IC method.

The third case is Elmers Restaurants Inc. which shows up in the IC method as a gambling company in three years (2002-2004) out of the nine years (1996-2004) that exist in the sample. The company’s 2002 annual filing describes the business as ”The Company...is a franchisor and operator of full-service, family oriented restaurants” and the business segments as ”The Company primarily operates in two business segments: restaurant operations and franchising the Elmer’s restaurant concept”. The company’s filing shows no sign of sin activities, and indeed there is no single mention of casino, gambling, betting, or wagering in any of its filings.

Creative Learning Products Inc., Activision Blizzards, and Santa Fe Financial Corp. are

¹⁰The company does not show up as sin in the TA method as its filing does not include the minimum requirement of having two unique alcohol-related words.

also companies identified as sin (gambling) at least in one year in the IC but never in the TA method. We checked their corresponding annual reports and confirm that these are all cases of false positive sin stocks in the IC method.

Overall, the six companies that appear at least once in the IC as sin and never sin in the TA method appear to be cases of false positives in which a company is assigned (incorrectly) an industry code related to a sin business.

4.1.4 False negatives in the IC method

There are many companies that show up in the TA method as having at least some exposure to sin activities that are not considered sin in the IC method. This is not surprising as sin companies in the IC method are presumably those that are highly exposed to sin activities. In fact, one of the contributions of the TA method is that it can identify companies that have exposure to sin activities, even though their main business is not directly in the sinful industry.

What is surprising is that there are quite a few companies whose businesses are mainly in or very closely linked to sin activities, but do not show up as sin firms in the IC method. The TA method reveals numerous examples in all sin industries where companies are significantly or predominantly engaged in sin activities, but are not identified as a sin firm in the IC method. The five most revealing cases within each industry are shown in Panel B of Table 3. These companies have several hundred counts of sin-related words, very close and sometimes higher than the average sin-related words of sin companies identified in the IC method. The complete list of sin-related companies in the TA method that are never picked up in the IC method is provided in Panel B of Table A.2.¹¹

We manually checked the annual reports of most of these companies and can confirm that they are indeed significantly engaged in sinful businesses, but the industry codes (SIC and NAICS) assigned to them by Compustat do not identify them as sin. Upon checking several different data providers and industry classifications, we observed that in some cases, other industry codes (e.g., NACE) assigned by other data providers (e.g., Capital IQ) correctly identify some of these sin companies. For example, Pyxus International, previously discussed, is assigned the NACE code (the industry classification standard used in the European Union) that corresponds to the manufacturing of tobacco products. Nevertheless, we believe that Compustat still provides the most reasonable classification.

Examples of tobacco companies we identify are Core Mark Holding and Holts Cigar

¹¹We acknowledge that we would not count many companies in Panel B of Table A.2 as sin based on a binary definition of sin, as we will discuss in later sections. The list is sorted in descending order by the average frequency of sin-related words, so sinfulness tends to decrease as we go down the list.

Holdings, which are in cigarettes and cigar businesses. As highlighted in Core Mark’s annual report: “The distribution of cigarettes is currently a significant portion of our business. In 2020, approximately 66.7% of our net sales were generated from the distribution of cigarettes.” Other examples are Greenlane Holdings and Vapor Corp which are involved in cannabis and vaporizers. As highlighted in their reports (see below), they are typically considered more sinful than traditional tobacco businesses to the extent that even banks are often hesitant to offer them banking services.

Vaporizer products comprise a significant portion of our product portfolio.. Some of the products we sell contain nicotine, which is considered to be a highly-addictive substance, or other chemicals that some jurisdictions have determined to cause cancer and birth defects or other reproductive harm... The State of California has determined that some chemicals found in certain vaporizers, as well as materials frequently consumed by using vaporizers (such as cannabis), cause cancer and birth defects or other reproductive harm. ...We are subject to legislative uncertainty that could slow or halt the legalization and use of cannabis, which could materially and adversely affect our business... We and our customers may have difficulty accessing the service of banks [due to cannabis-related activities], which may make it difficult for us and for them to sell our products.. [G]iven these risks, many banks remain hesitant to offer banking services to cannabis-related businesses. Indeed, we have been asked to close bank accounts due to our activity in the cannabis industry.

Examples from other sin industries are also provided in Panel B of Table 3. For instance, Magna Entertainment (2000-2007) is never classified as sin by the IC method but was a major operator of horse tracks and other betting and wagering activities with an average sin-related word count of 742. Additionally, we identify adult entertainment companies that could not be identified at all in the IC method as there is no clear industry classification for these companies. Overall, we identify nine unique companies in the adult entertainment business. These include some usual suspects, such as RCI Hospitality Holdings and Private Media Group, but also other less well-known companies such as Spice Entertainment, VCG Holdings, and FriendFinder Networks.

It is important to note that companies whose main activities are not in but closely related to sin industries would be picked up consistently in the TA method as sin companies. However, these companies are sometimes assigned industry codes that identify them as sin and sometimes not (in an arbitrary manner). Two examples of such companies in the gaming industry are Everi Holdings Inc. (2005-2021) and Scientific Games Corporation (1996-2021). According to the most recent filing of these two companies:

Everi is a leading supplier of imaginative entertainment and trusted technology

solutions for the casino and digital gaming industry. Focused on player engagement and assisting our casino customers operate more efficiently, the Company develops entertaining game content and gaming machines, gaming systems and services for land-based and iGaming operators.

and

We [Scientific Games Corporation] are a leading developer of technology based products and services and associated content for the worldwide gaming, lottery, social and digital gaming industries. Our portfolio of revenue-generating activities in our continuing operations primarily includes supplying game content and gaming machines, and table game products and services to licensed gaming entities.

It is clear that the main business of both Everi and Scientific Games is being a leading supplier for the gaming industry. Consistently, they are both identified as sin (gambling) companies in the TA method. However, only one shows up as sin in the IC method: Scientific Games. Once again, this highlights the somewhat arbitrary nature of industry classification codes assigned to different companies and why textual analysis of company filings generates outcomes that are more consistent across companies and within companies over time.

4.1.5 The different shades of sinfulness

One important advantage of the TA method is that it also identifies companies whose main business may not be within a sin industry and one may not even be able to understand it by reading their business description. Nevertheless, they are actually either directly engaged in sin businesses and/or they are exposed to the risk of it. Obviously, these companies may not be considered sin with a binary definition, but given our continuous measure of sinfulness, they are shown to be more irresponsible or sinful than companies that are not at all involved in sinful activities. This information is also important for responsible investors that may not want to be exposed to companies that are partially involved in sin activities, or any other investor who would simply want to avoid the risk of being exposed to sin industries. On the other hand, for investors whose aim happens to be to get exposure to sinful industries, identifying these partially sin firms could be of interest.¹²

An example of a company that shows up among sin-related tobacco companies in the TA method (although ranked relatively low in terms of sinfulness) is 7-Eleven Inc., a retail convenience store. While it might be surprising to see such a company in the list of sinful

¹²There are indeed investors who look for getting exposure to sinful businesses due to their apparently strong profitability and resistance to recessions. There are even exchange-traded funds with the sole purpose of investing in sin stocks (e.g., tickers BAD and VICE).

firms, a large part of the company's revenue was indeed generated from selling tobacco products, hence its significant exposure to the tobacco industry. Below is a paragraph from the risk factors section of its 2004 annual report (the last year the company appears in our sample and the year after which it was acquired by a Japanese retailer, Seven & Holdings Co.):

Over the past three fiscal years, tobacco products sales have averaged approximately 29% of our merchandise sales. Future tobacco legislation, national, state and local campaigns to discourage smoking and increases in taxes on cigarettes and other tobacco products could have a material adverse effect on sales of and margins for the tobacco products we sell.

In fact, there are companies that have decided to stop being involved in sinful businesses even if their main business was not within the sin industries. An example is CVS/pharmacy that decided in 2014 to stop selling cigarettes and tobacco products (which was generating more than \$2 billion in sales) in all of its 7,700 stores. The reason, as described by its CEO, was that removing tobacco products from the shelves is "the right thing for us to do for our customers and our company to help people on their path to better health...Put simply, the sale of tobacco products is inconsistent with our purpose".

In summary, one important advantage of the TA method (over the IC method) is its power in identifying companies that are involved in sin-related activities such as selling or distributing cigarettes but whose main business is not actually producing cigarettes and hence would not be assigned the industry code of tobacco.

4.1.6 Sin in both TA and IC but with different coverage

Even in situations where a firm appears correctly (at least in some years) as sin in the IC method, the number of years does not always match that of the TA method. In most of such situations, the number of years being sin in the IC method is less than that of the TA method. We manually check these cases and confirm that the output obtained by the TA method is more accurate.

One significant example from the tobacco industry is Universal Corporation, which is identified as sin in all years that exist in our sample (25 years) by the TA method, but is only considered sin in 13 years according to the IC method. Specifically, between and including 2007 and 2018 there are no tobacco related segments reported for the company in the Compustat segment data. During all these years, the frequency of tobacco-related words actually increased from about 400 to more than 600 for the company. Consistently, for example, in 2018 (which was not identified as sin in the IC method), the company described its business as:

We are the leading global leaf tobacco supplier. We operate in over 30 countries on five continents and procure, finance, process, pack, store and ship leaf tobacco and other agri-products. Tobacco has been our principal focus since our founding in 1918. The largest portion of our business involves procuring and processing flue-cured and burley leaf tobacco for manufacturers of consumer tobacco products.

The company's reported segments and operations in 2019 is identical as above, but suddenly changes status from a non-sin to a sin firm in 2019 according to the IC method. This clearly identifies inconsistency in industry classification codes assigned to the same company across years.

Another significant case from the alcohol industry is that of Leucadia National Corp. This company shows up only one year (in 1998) as a sin company in the IC method but is identified as sin in the TA method up until 2012. The company indeed had winery operations in 1998 but these operations continued (and even increased) in the next years. Consistently, we observe a significant increase in the count of alcohol-related words up to 2010. The company correctly shows up until 2012 in our sinful companies and then the alcohol count goes sharply down to zero over the next years. In fact, in 2013, the company spun off its winery business and reports in its annual filing that "On February 25, 2013, we distributed to our shareholders the common shares of the Crimson Wine Group, Ltd., a holding company through which we historically conducted our winery operations".

In the gambling industry, the largest and most significant difference is for Churchill Downs Inc., one of the most famous players in the gambling industry. Although the count of gambling-related words is very high and the company shows up in every single year as a sin firm in the TA method, it only appears in the IC method since 2006. Manually checking the company's filings around 2006, we confirm that the company's business is identical with even identical wordings. The business descriptions for 2005 and 2006 are identical as follows:

Churchill Downs Incorporated (the "Company") is principally a racing company that conducts pari-mutuel wagering on live Thoroughbred, Quarter Horse and Standardbred horse racing and simulcasts signals of races. Additionally, we offer racing services through our other business interests as well as alternative gaming through video poker operations in Louisiana.

The list of every company that shows up at least once in the TA and IC methods (but not necessarily exactly the same years) is reported in Panel A of Table [A.2](#). The last two columns show the number of years each company has shown up as a sin firm in the IC and TA methods, respectively.

Overall, examining cases where there is a large difference in the number of sin-years shown in the IC versus the TA method, our findings show that using industry codes to identify

sinful companies appears straightforward, but generates outcomes that are not reliable. As discussed, even for significant sinful companies and when there is absolutely no change in the business, the sin indicator based on the IC method could change from one year to another. One might wonder whether a solution to this issue with the IC method could be to back-fill or forward-fill the sin indicator. That is, if a company is identified as sin in one year (e.g., 2004) based on its industry classification, the researcher would count the company as sin in all its history (e.g., 1995 to 2021). Although such an approach would fix some of the issues regarding the randomness of industry codes assigned to companies mentioned above, it creates other serious problems as discussed earlier. Specifically, quite often sin companies spin off their sinful businesses or non-sin firms acquire sin companies. In such cases, back-filling and forward-filling would incorrectly count a company as a sin stock over the entire existence of the company even if the company was actually just exposed to sinful activities in one specific year.

4.2 Sinfulness and asset prices

In the previous sections, we discussed the shortcomings of the IC method in correctly identifying the sin stocks. We also showed the power of the TA method in measuring the exposure of companies to sinful activities. In this section, we investigate the risk-adjusted performance of sin stocks using both the TA and IC methods.

4.2.1 Portfolio Construction

We construct sin portfolios using the two commonly used weighting methods for portfolio construction, i.e., equal- and value-weighted, as well as a novel weighting method, which we call sin-weighted. Regardless of the method (IC or TA), both equal- and value-weighted portfolios are problematic. In the IC method, all false positives would get equal weights as very sinful companies for equal-weighted portfolios and often even higher weights in value-weighted portfolios. For example, Valero Energy Corp. discussed above as a false positive in 9 years in the IC method would get the same weight as important alcohol companies such as Constellation Brands or Boston Beer in the equal weighting. In a value-weighted approach, the company would even get a significantly larger weight than those true sinful companies due to its greater size. Therefore, we argue that an alternative weighting approach would be to construct a weighting scheme depending on sinfulness, i.e, the frequency of sinful words in the company's reports. Such a scheme would be crucial in the TA method, as the universe of sin-related stocks is significantly larger and with a larger heterogeneity in their exposure to sin activities.

We construct portfolios each year in June, thus holding each sin portfolio from July of year t to June of $t + 1$. To be incorporated in the portfolio construction, we require that a company has a valid Compustat SIC/NAICS code for the fiscal year ending in calendar year $t - 1$, a non-missing market capitalization in June, and an annual filing for the fiscal year ending in calendar year $t - 1$ that was filed by the end of June in year t .

4.2.2 Sin weighting

A major advantage of using a textual analysis-based identification approach is that the resulting measure (frequency of sinful words) can potentially serve as a proxy for the sinfulness of individual sin stocks relative to others, beyond a purely binary classification. We acknowledge that the sheer frequency of words within a sinful industry is not an accurate measure of continuous sinfulness. More specifically, we do not claim that a tobacco company with, e.g., 400 sinful tobacco words is four times as sinful as a tobacco company with only 100 sinful tobacco words. Nevertheless, we consider it a suitable measure for a more general assessment of relative sinfulness, in the sense that we consider a stock with 200 sinful words to be relatively more sinful than a stock with 100 or just 20 sinful words.

To investigate the impact of sinfulness on stock returns, we propose a novel approach to portfolio weighting. In addition to equal-weighted or value-weighted portfolios, as is common in the literature, we examine a sin-weighting approach. In a sin-weighted portfolio, stocks with a relatively higher frequency of sinful words receive a larger weight in the portfolio construction. Regardless of the underlying mechanism (e.g., being shunned by responsible investors, having higher litigation risk, etc.), the abnormal returns to sin stocks, if any, should be larger or particularly visible for more sinful ones.

To account for this relative sinfulness, we calculate in each year t for each sin industry j (i.e., alcohol, tobacco, gambling, and adult entertainment) a relative sinfulness ($\tilde{s}_{i,j,t}$) of each sin stock i as follows:

$$\tilde{s}_{i,j,t} = \frac{Freq_{i,j,t-1}}{\frac{\sum_{i=1}^{N_{j,t}} Freq_{i,j,t-1}}{N_{j,t}}}, \quad (1)$$

where $Freq_{i,j,t-1}$ is the absolute frequency of sin-related words for sin stock i in sin industry j in the filing from the fiscal year ending in calendar year $t - 1$ and $N_{j,t}$ is the total number of sin stocks in year t in sin industry j . To limit the effects of outliers, we winsorize $Freq_{i,j,t-1}$ in each year t for each sin industry j at the 90th percentile prior to calculating $\tilde{s}_{i,j,t}$.

We then use the relative sinfulness ($\tilde{s}_{i,j,t}$) to calculate a portfolio weight ($\tilde{w}_{i,t}$) of sin stock

i in the overall sin portfolio as follows:

$$\tilde{w}_{i,t} = \frac{\sum_{j=1}^{I_t} \tilde{s}_{i,j,t}}{\sum_{j=1}^{I_t} \sum_{i=1}^{N_{j,t}} \tilde{s}_{i,j,t}}, \quad (2)$$

where $\tilde{s}_{i,j,t}$ is the relative sinfulness measure of sin stock i in sin industry j in year t , $N_{j,t}$ is the total number of sin stocks in year t in sin industry j , and I_t is the total number of sin industries in year t .

We note that the summations with the subscript j (in both the numerator and denominator) exist because some stocks might be considered sinful in several sin industries. In other words, for a company that is involved in only one sin industry, Eq. 2 simply divides its relative sinfulness, measured according to Eq. 1, to the sum of relative sinfulness of all other sin stocks in a given year.

By dividing the frequency of, e.g., sinful tobacco words of a tobacco company by the average frequency of sinful tobacco words across all tobacco companies in that year (see, Eq. 1), a tobacco stock with a relatively high frequency of tobacco words will receive a larger weight (see, Eq. 2) in the overall sin portfolio. In addition, this approach controls for a time-varying change in the general occurrence of specific words and, in a second step, allows for a more neutral portfolio construction across the different sin industries. Hence, without an appropriate relative consideration, e.g., the tobacco industry would receive a greater weight due to the larger absolute frequencies of the selected tobacco words (see, Panel A of Table 1), which would imply that this industry is more sinful than, e.g., the alcohol industry. In our approach, we regard all four sin industries as comparably sinful by giving stocks from all industries the opportunity to receive a relatively larger weight regardless of the absolute number of sinful words within a sin industry. The described procedure results in moderate max. weights for the TA method ranging from a minimum of 1.74% to 5.3% and ensures that the inference is not driven by a few stocks dominating the sin portfolio.

4.2.3 Is there a sin premium?

To investigate excess returns to sin stocks, we form portfolios based on three different weighting schemes: equal-weighted, value-weighted, and sin-weighted. Figure 3 (a) shows the performance of \$1 invested in July 1997 in a portfolio of sin stocks (with different methods of identifying sin and different weighting schemes). For the sake of comparison, we also plot the performance of a portfolio of comparable companies. Other than the comparables portfolio, we have four portfolios plotted: Sin TA is a portfolio of sin stocks identified in the TA method, whereas Sin IC includes sin stocks according to the IC method. For each of

these portfolios, we plot equal-weighted and sin-weighted portfolio performances. Returns to portfolios of sin stocks have been quite high, although there are significant differences between methods and weighting schemes. In addition to the generally positive returns, two clear patterns can be identified. First, the (equal- or sin-weighted) IC sin portfolios, shown with dashed lines, have had higher returns than the corresponding TA sin portfolios, presumably since companies in the IC method are more sinful, on average. Second, and more interestingly, within each method, the cumulative returns to the sin-weighted portfolios are higher than those of equal-weighted portfolios. This is consistent with the idea that more sinful firms have higher expected returns than less sinful ones. Figure 3 (b) shows the performance of long (sin) - short (comparables) portfolios, respectively. We observe the same two messages as above.

[Please insert Figure 3 near here]

We proceed to formally estimate the excess returns of sin stocks using time-series factor regressions. We use the return on sin portfolios in excess of comparables as the dependent variable and estimate alphas using the [Fama and French \(1993\)](#) 3 factor model [FF3], the [Carhart \(1997\)](#) 4 factor model [CH4], and the [Fama and French \(2018\)](#) 6 factor model [FF6], respectively. We report Newey-West standard errors.

Table 4 reports the results. Panel A shows our baseline results, including all stocks that appear in the TA method as sinful with a minimum count of 20 sin-related words. There are 9 columns in the table that report the results of three different weighting schemes and three different factor models (FF3, CH4, and FF6) within each scheme. As shown, excess return estimates (α) in equal- and value-weighted portfolios of sin stocks are very small in magnitude and are not statistically different from zero (Columns 1-6). However, a sin-weighted portfolio generates a significant monthly excess return of 34 basis points (t-stat of 2.74) in the FF6 factor model (Column 9). These results are in contrast to the results of recent studies (using the IC method, e.g., [Blitz and Fabozzi, 2017](#)) which argue that sinful stocks do not earn higher returns after controlling for common determinants of stock returns. Our results show that the sin premium does exist, is almost twice as large as previously estimated, and is a function of sinfulness (which is why it is absent from equal- and value-weighted portfolios). We note that there are 84 stocks, on average, in the long leg and 227 stocks in the short leg of these portfolios.

Panels B and C of Table 4 report similar results as in Panel A but using a stricter measure of sinfulness. Specifically, a firm in Panel B (C) is considered sin if its filing has a minimum of 100 (200) sin-related firms (referred to as TA 100 or TA 200). We note that requiring such a high frequency of sin-related words effectively means removing quite

a few sinful companies from the sin portfolio that are counted as sin even according to the IC method. Consequently, we do expect that the results of the equal-weighted (but not necessarily value-weighted) portfolios converge to those of the sin-weighted portfolios. The reason is that, by increasing the minimum required sin-related words, we are focusing on a subset of highly sinful companies. However, because two companies that are highly exposed to a sin industry could have very different market values, the value-weighted results may not necessarily converge to equal- and sin-weighted results as we focus on more sinful stocks. The excess returns reported in Panel B Column 3 show that the excess return of TA 100 is 26 basis points and marginally significant (t-stat of 1.92) in an equal-weighted portfolio. The corresponding number for a value-weighted portfolio is an insignificant 9 basis points (t-stat of 0.46). However, as shown in Panel B Column 9, the excess return to the sin-weighted portfolio stays significant and increases slightly in magnitude to 40 basis points (t-stat of 2.88). Panel C of Table 4 shows that when we focus on TA 200, the results of all weighting schemes begin to converge and become statistically significant.

[Please insert Table 4 near here]

Despite the issues with the IC method that we highlighted, for the sake of comparison, we repeat the same regressions as above, but this time with sin firms as identified by the IC method. The results are reported in Table 5. As shown in Column 3, we estimate a positive and significant alpha of 46 basis points (t-stat 2.78) in the FF6 factor model. However, the value-weighted portfolio of these sin stocks does not earn a significant excess return. Specifically, we estimate a CH4 alpha (Column 5) of 23 basis points (t-stat of 1.31), which reduces to 9 basis points (t-stat of 0.52) when controlling for investment and profitability factors (Column 6). Interestingly, even in this method, where we identify companies based on the IC method, using a sin-weighting approach leads to estimated excess returns that are larger in magnitude and statistical significance. As shown in Column 9, a sin-weighted portfolio yields a significant alpha of 56 basis points (t-stat 3.71) in the FF6 factor model. It should be noted that in the sin-weighting approach, false positive sin stocks in the IC method are effectively excluded from the portfolio as they get zero weights in the portfolio construction. This is the reason why we have lower number of stocks in the sin-weighted portfolio (Columns 7-9) compared to the equal- or value-weighted portfolios (Columns 1-6).

[Please insert Table 5 near here]

We perform two further analyses to check the robustness of these results. First, although the excess returns in Tables 4 and 5 are computed using comparables firms (which is arguably

the most appropriate method), one may wonder if these results are robust to using the risk-free rate as a benchmark in computing the excess returns. Tables [A.3](#) and [A.4](#) show the results using the TA method and the IC method, respectively. Overall, the results for sin-weighted portfolios are similar as we reported before, but the magnitudes of alphas increase (e.g., 51 versus 34 basis points in Panel A Column 9). However, the value-weighted results are no longer even marginally significant when we use the risk-free rate as the benchmark. Second, as shown in [Figure 3](#), both the equal- and sin-weighted portfolios performed quite well during and after the Covid pandemic. To analyze the sensitivity of our results to this particular period, we report our results excluding the Covid period for the IC and TA methods in Tables [A.5](#) and [A.6](#), respectively. As shown, all the main results discussed above remain qualitatively intact.

Overall, these results show that the sin premium does exist, i.e., sin stocks have higher expected returns. We document a premium that is substantially larger in magnitude than what was previously estimated and we demonstrate that it is driven, as it should be, by more sinful companies, not necessarily larger ones. Therefore, focusing on value-weighted portfolios could generate misleading results.

5 Conclusion

We propose a novel method based on the textual analysis of corporate annual reports to identify sin stocks and to measure their sinfulness. Our method performs much better than the procedure used in the prior literature, which relies on using industry classification codes. Using industry classification codes generates several false positive, numerous false negative sin stocks, and sometimes random time-series variation within a stock (e.g., inconsistent and random jumps back and forth between sin and non-sin classification). Furthermore, the outcome of sinfulness using industry classifications is binary, i.e., a company is either sinful or not. Furthermore, researchers have been unable to identify companies in sinful sectors that do not have a specific industry code (e.g., adult entertainment or pornography). Our method overcomes and/or improves upon all the above-mentioned shortcomings. Contrary to the findings of several recent studies, we find strong evidence consistent with the existence of a sin premium. A sin-weighted (but not necessarily an equal- or value-weighted) portfolio of sin stocks generates a [FF6](#) alpha of 4% per year from 1997 to 2021. This suggests that investors require higher expected returns to hold more sinful stocks.

Although our focus is on particular sinful businesses that have always been considered sin in the prior literature (which is also important for the sake of comparison), the method we introduce could potentially be customized to capture firms' exposure to any other sinful

activities such as being involved in the weapon industry. However, it is important to note that relying on companies' disclosure to measure many other aspects of corporate social (ir)responsibility (e.g., diversity and inclusion) would generate misleading results. Companies that write a lot about diversity and inclusion may actually care very little about it. Even in these cases, researchers could use textual analysis but apply it to anonymous employee reviews (e.g., from Glassdoor's website) or newspaper articles. Finally, reliably measuring many aspects of corporate social responsibility, such as efficient use of resources or carbon emissions, is impossible in an algorithmic way and would require a company-by-company assessment by experts.

References

- Ahrens, D., 2004. Investing in Vice: The Recession-Proof Portfolio of Booze, Bets, Bombs & Butts. St. Martin's Press.
- Aswani, J., Raghunandan, A., Rajgopal, S., forthcoming. Are carbon emissions associated with stock returns? *Review of Finance* .
- Azimi, M., Agrawal, A., 2021. Is positive sentiment in corporate annual reports informative? evidence from deep learning. *The Review of Asset Pricing Studies* 11, 762–805.
- Bali, T. G., Engle, R. F., Murray, S., 2016. Empirical asset pricing: The cross section of stock returns. Wiley, Hoboken, New Jersey.
- Bellstam, G., Bhagat, S., Cookson, J. A., 2021. A text-based analysis of corporate innovation. *Management Science* 67, 4004–4031.
- Berk, J. B., van Binsbergen, J. H., 2021. The impact of impact investing. Stanford University Graduate School of Business Research Paper .
- Bertomeu, J., Cheynel, E., Floyd, E., Pan, W., 2020. Using machine learning to detect misstatements. *Review of Accounting Studies* 26, 468–519.
- Blitz, D., Fabozzi, F. J., 2017. Sin stocks revisited: Resolving the sin stock anomaly. *Journal of Portfolio Management* 44, 105–111.
- Bochkay, K., Hales, J., Chava, S., 2019. Hyperbole or reality? investor response to extreme language in earnings conference calls. *The Accounting Review* 95, 31–60.
- Bodnaruk, A., Loughran, T., McDonald, B., 2015. Using 10-k text to gauge financial constraints. *Journal of Financial and Quantitative Analysis* 50, 623–646.
- Bolton, P., Kacperczyk, M., 2021. Do investors care about carbon risk? *Journal of Financial Economics* 142, 517–549.
- Broccardo, E., Hart, O., Zingales, L., 2022. Exit versus voice. *Journal of Political Economy* 130, 3101–3145.
- Buehlmaier, M. M. M., Whited, T. M., 2018. Are financial constraints priced? evidence from textual analysis. *The Review of Financial Studies* 31, 2693–2728.
- Carhart, M. M., 1997. On persistence in mutual fund performance. *Journal of Finance* 52, 57–82.

- Chava, S., 2014. Environmental externalities and cost of capital. *Management Science* 60, 2223–2247.
- Eccles, R. G., Rajgopal, S., Xie, J., 2022. Does esg negative screening work? SSRN Working paper .
- Edmans, A., Levit, D., Schneemeier, J., 2022. Socially responsible divestment. European Corporate Governance Institute Working Paper No. 823/2022 .
- Fabozzi, F. J., Ma, K. C., Oliphant, B. J., 2008. Sin stock returns. *Journal of Portfolio Management* 35, 82–94.
- Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3–56.
- Fama, E. F., French, K. R., 1997. Industry costs of equity. *Journal of Financial Economics* 43, 153–193.
- Fama, E. F., French, K. R., 2018. Choosing factors. *Journal of Financial Economics* 128, 234–252.
- Feldman, R., Govindaraj, S., Livnat, J., Segal, B., 2009. Management’s tone change, post earnings announcement drift and accruals. *Review of Accounting Studies* 15, 915–953.
- García, D., Norli, Ø., 2012. Geographic dispersion and stock returns. *Journal of Financial Economics* 106, 547–565.
- Hanley, K. W., Hoberg, G., 2010. The information content of IPO prospectuses. *Review of Financial Studies* 23, 2821–2864.
- Heinkel, R., Kraus, A., Zechner, J., 2001. The effect of green investment on corporate behavior. *The Journal of Financial and Quantitative Analysis* 36, 431.
- Henry, E., Leone, A. J., 2016. Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. *The Accounting Review* 91, 153–178.
- Hoberg, G., Maksimovic, V., 2014. Redefining financial constraints: A text-based analysis. *The Review of Financial Studies* 28, 1312–1352.
- Hoberg, G., Phillips, G., 2016. Text-based network industries and endogenous product differentiation. *Journal of Political Economy* 124, 1423–1465.

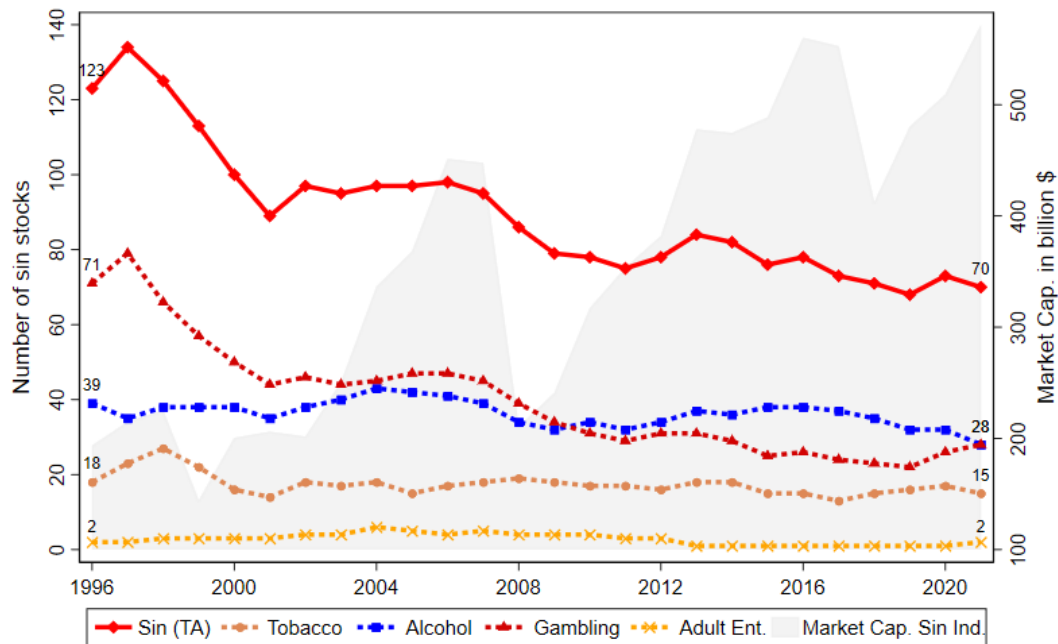
- Hoberg, G., Phillips, G., Prabhala, N., 2014. Product market threats, payouts, and financial flexibility. *The Journal of Finance* 69, 293–324.
- Hong, H., Kacperczyk, M., 2009. The price of sin: The effects of social norms on markets. *Journal of Financial Economics* 93, 15–36.
- Hsu, P.-H., Li, K., Tsou, C.-Y., 2023. The pollution premium. *The Journal of Finance* .
- Huang, A. H., Wang, H., Yang, Y., 2023. Finbert: A large language model for extracting information from financial text. *Contemporary Accounting Research* .
- Huang, X., Teoh, S. H., Zhang, Y., 2013. Tone management. *The Accounting Review* 89, 1083–1113.
- Kahle, K. M., Walkling, R. A., 1996. The impact of industry classifications on financial research. *Journal of Financial and Quantitative Analysis* 31, 309.
- Kim, I., Venkatachalam, M., 2011. Are sin stocks paying the price for accounting sins? *Journal of Accounting, Auditing and Finance* 26, 415–442.
- Li, K., Mai, F., Shen, R., Yan, X., 2020. Measuring corporate culture using machine learning. *The Review of Financial Studies* 34, 3265–3315.
- Loughran, T., McDonald, B., 2013. IPO first-day returns, offer price revisions, volatility, and form s-1 language. *Journal of Financial Economics* 109, 307–326.
- Loughran, T., McDonald, B., 2020. Measuring firm complexity. *SSRN Electronic Journal* .
- Loughran, T. I., McDonald, B., 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *Journal of Finance* 66, 35–65.
- Luo, H. A., Balvers, R. J., 2017. Social screens and systematic investor boycott risk. *Journal of Financial and Quantitative Analysis* 52, 365–399.
- Muslu, V., Radhakrishnan, S., Subramanyam, K. R., Lim, D., 2015. Forward-looking md&a disclosures and the information environment. *Management Science* 61, 931–948.
- Pedersen, L. H., Fitzgibbons, S., Pomorski, L., 2021. Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics* 142, 572–597.
- Renneboog, L., Horst, J. T., Zhang, C., 2011. Is ethical money financially smart? nonfinancial attributes and money flows of socially responsible investment funds. *Journal of Financial Intermediation* 20, 562–588.

- Shumway, T., 1997. The delisting bias in crsp data. *The Journal of Finance* 52, 327–340.
- Shumway, T., Warther, V. A., 1999. The delisting bias in CRSP's nasdaq data and its implications for the size effect. *The Journal of Finance* 54, 2361–2379.
- Statman, M., Glushkov, D., 2009. The wages of social responsibility. *Financial Analysts Journal* 65, 33–46.
- Tetlock, P. C., 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance* 62, 1139–1168.
- Tetlock, P. C., Saar-Tsechansky, M., Macskassy, S., 2008. More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance* 63, 1437–1467.
- Waxler, C., 2004. *Stocking Up on Sin: How to Crush the Market with Vice-Based Investing*. Wiley.
- Zerbib, O. D., 2022. A sustainable capital asset pricing model (s-CAPM): Evidence from environmental integration and sin stock exclusion. *Review of Finance* 26, 1345–1388.

Figure 1: Number of sin stocks over time by classification approach

This figure shows the number of Sin (adjusted for double-counting), Tobacco, Alcohol, Gambling and Adult Entertainment stocks for our baseline textual analysis (TA) classification (frequency ≥ 20 & unique ≥ 2) [Subfigure (a)] and the industry classification (IC) approach [Subfigure (b)], respectively. The second y-axis shows the aggregate market capitalization of all sin stocks.

(a) TA classification



(b) IC classification

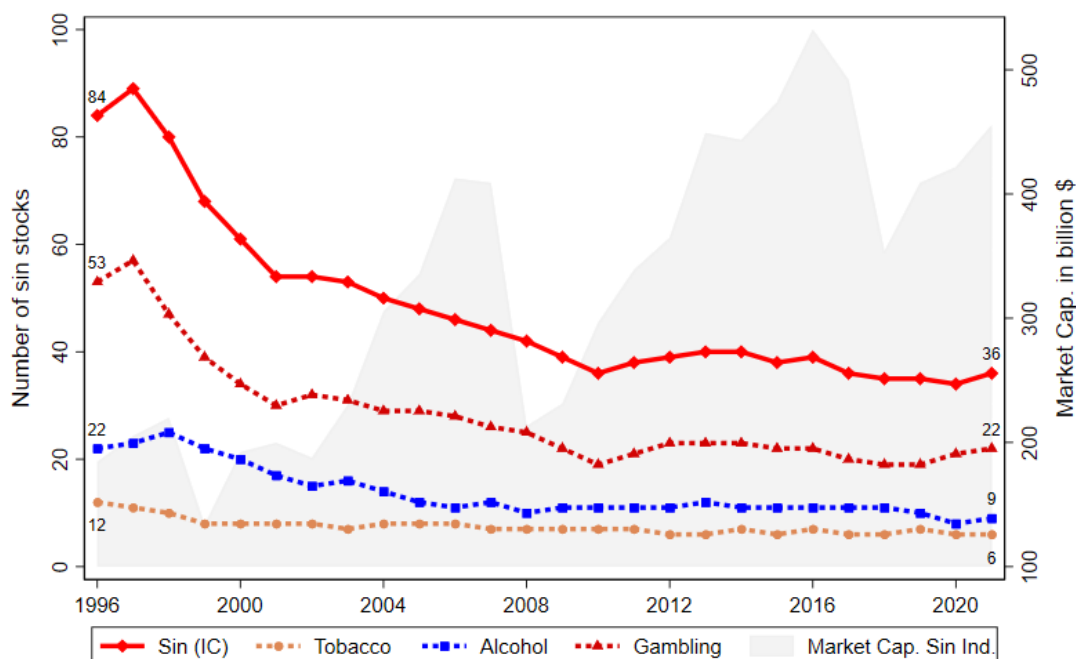
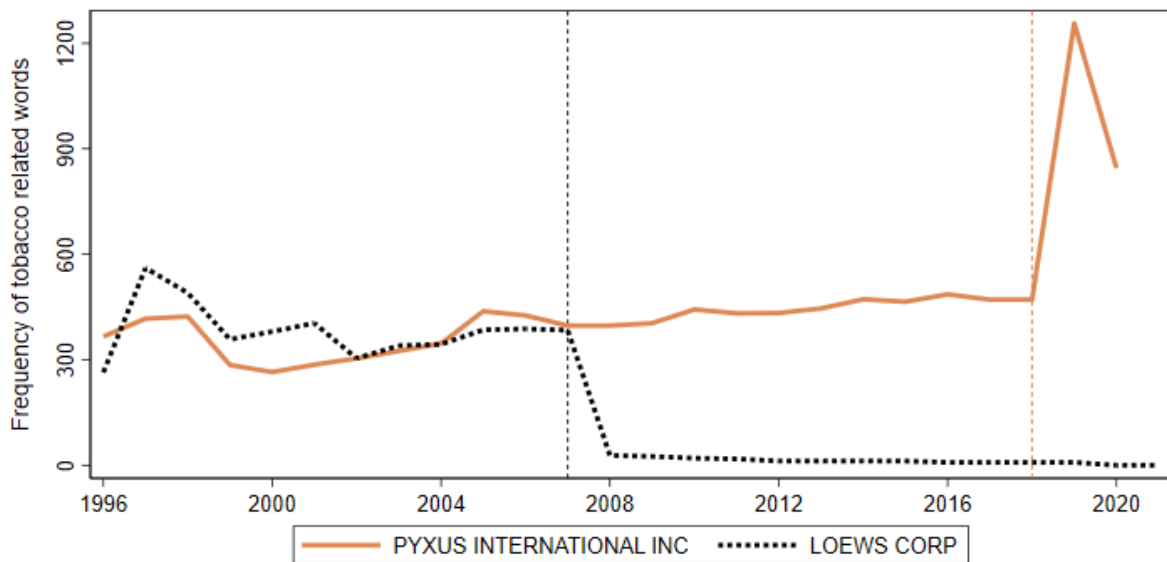


Figure 2: Motivational company examples

This figure illustrates particular cases to highlight the power of textual analysis in identifying within industry changes in sinfulness of a firm [Subfigure (a)] or across industry changes [Subfigure (b)]. Subfigure (a) plots the timeseries (frequency) of tobacco-related words in 10-K filings for two tobacco companies, i.e., Pyxus International, Inc. and Loews, Corp. The vertical dashed lines mark particular events (divestment in case of Loews and acquisitions in case of Pyxus) that led to sudden changes in these firm's sinfulness. Subfigure (b) plots tobacco-related counts for three alcohol companies, one that divested out of tobacco whereas two others started to invest in tobacco-related businesses. In order to improve the readability of Subfigure (b), we added an additional y-axis for the two firms which are represented by dashed lines.

(a) Pyxus International, Inc. & Loews, Corp.



(b) Constellation Brands, Molson Coors Beverage, Co. & Beam, Inc.

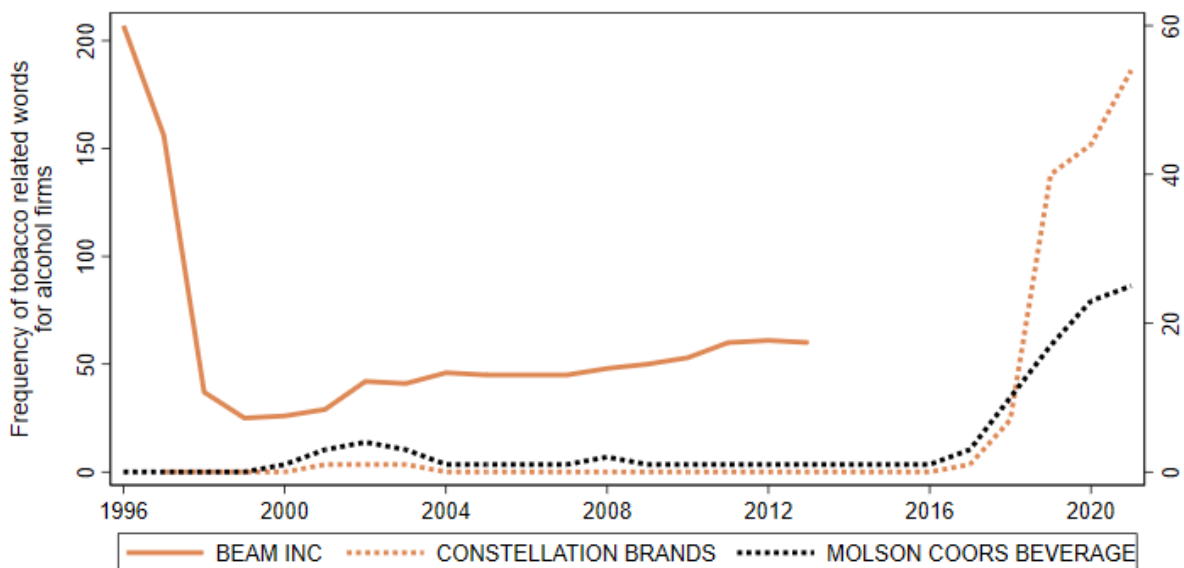
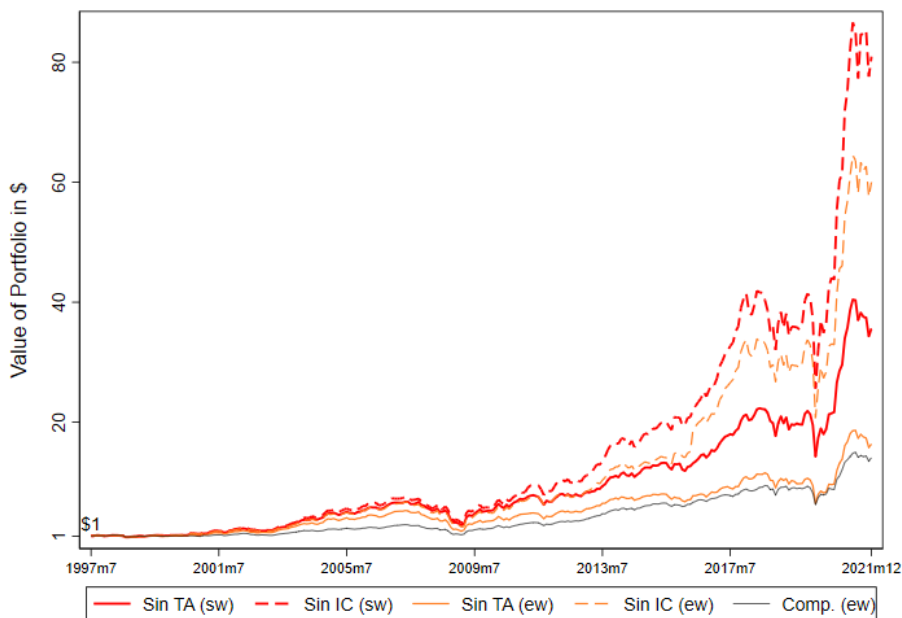


Figure 3: Performance of sin portfolios

This figure shows the performance of portfolios purchased in July 1997 for \$1 and held through December 2021. We show the performance of textual analysis (TA) and industry classification (IC) based sin portfolios with equal- [(ew)], and sin-weighting [(sw)], respectively. Subfigures (a) and (b) show the cumulative returns for long only (sin stocks) and long-short (sin stocks - comparables) portfolios, respectively. For a better comparison we also show the performance of an equally-weighted [Comp.(ew)] portfolio of comparable firms in Subfigure (a).

(a) Long only



(b) Long - short

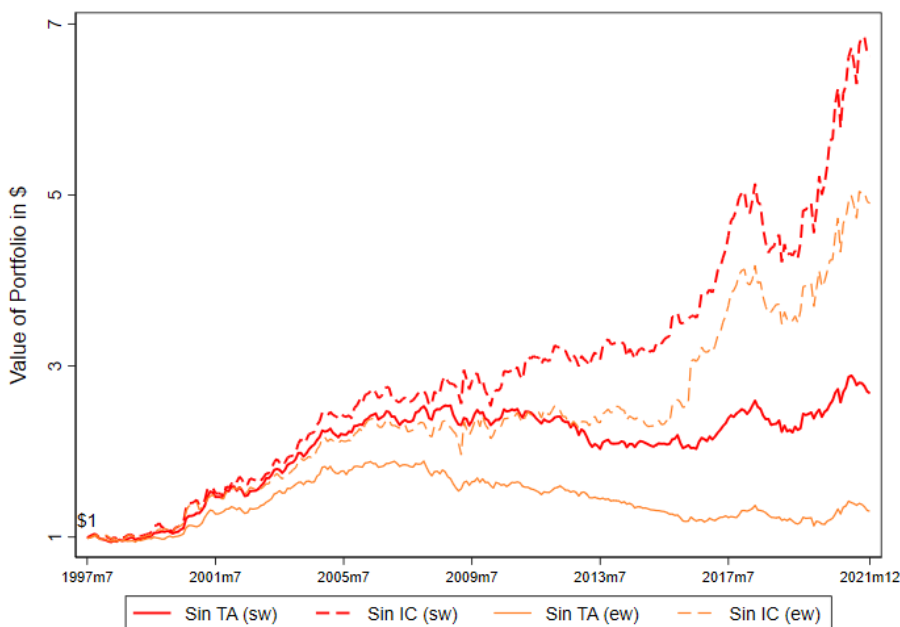


Table 1: Sin word frequencies summary statistics in 10-K filings

This table reports summary statistics of the frequencies of sin-related words in the 10-K filings of all firms that are classified as sin by our textual analysis (TA) approach (Panel A) and by the industry classification (IC) approach (Panel B), respectively. Furthermore, we report the number of sin firm years and unique sin firms, respectively.

Panel A: Sin-related firms based on textual analysis (TA)							
Industry freq.	Mean	SD	Min	p50	Max	# Firm years	# Unique firms
TOBACCO	348	405	20	240	2517	452	56
ALCOHOL	174	210	20	52	1004	945	134
GAMBLING	300	197	20	265	1247	1039	145
ADULT ENT.	113	59	23	107	263	72	9
SIN	270	264	20	208	2578	2406	327

Panel B: Sin related firms based on industry classification (IC)							
Industry freq.	Mean	SD	Min	p50	Max	# Firm years	# Unique firms
TOBACCO	595	484	0	463	2517	194	22
ALCOHOL	348	233	0	313	1004	357	37
GAMBLING	339	197	0	299	1247	736	97
ADULT ENT.	NA	NA	NA	NA	NA	NA	NA
SIN	394	285	0	329	2578	1258	153

Table 2: Sin stock distribution and classification differences

The table reports the number of stocks from each sin industry identified by the industry code based approach (IC) and by our proposed textual analysis-based approach (TA), respectively. Furthermore, we report the number of unique sin stocks over all industries by year (All column adjusted for double-counting) and over the entire history (Total row). Lastly, we report the number of sin stock firm months which are the same between the two approaches, which only show up in the industry code based approach and which only show up in our textual analysis based approach.

Year	Tobacco		Alcohol		Gambling		Adult Ent.		All		Differences			
	IC	TA	IC	TA	IC	TA	IC	TA	IC	TA	No diff.	Only IC	Only TA	# Firm months
1996	12	18	22	39	53	71	NA	2	84	123	78 (59%)	6 (4%)	47 (35%)	131
1997	11	23	23	35	57	79	NA	2	89	134	85 (60%)	4 (2%)	51 (36%)	140
1998	10	27	25	38	47	66	NA	3	80	125	72 (52%)	8 (5%)	56 (41%)	136
1999	8	22	22	38	39	57	NA	3	68	113	64 (53%)	4 (3%)	52 (43%)	120
2000	8	16	20	38	34	50	NA	3	61	100	56 (51%)	5 (4%)	47 (43%)	108
2001	8	14	17	35	30	44	NA	3	54	89	50 (52%)	4 (4%)	42 (43%)	96
2002	8	18	15	38	32	46	NA	4	54	97	51 (49%)	3 (2%)	50 (48%)	104
2003	7	17	16	40	31	44	NA	4	53	95	51 (50%)	2 (1%)	48 (47%)	101
2004	8	18	14	43	29	45	NA	6	50	97	49 (47%)	1 (.%)	54 (51%)	104
2005	8	15	12	42	29	47	NA	5	48	97	48 (47%)	0 (0%)	54 (52%)	102
2006	8	17	11	41	28	47	NA	4	46	98	46 (45%)	0 (0%)	56 (54%)	102
2007	7	18	12	39	26	45	NA	5	44	95	44 (44%)	0 (0%)	56 (56%)	100
2008	7	19	10	34	25	39	NA	4	42	86	40 (43%)	2 (2%)	50 (54%)	92
2009	7	18	11	32	22	34	NA	4	39	79	37 (43%)	2 (2%)	46 (54%)	85
2010	7	17	11	34	19	31	NA	4	36	78	35 (42%)	1 (1%)	47 (56%)	83
2011	7	17	11	32	21	29	NA	3	38	75	36 (45%)	2 (2%)	42 (52%)	80
2012	6	16	11	34	23	31	NA	3	39	78	37 (44%)	2 (2%)	44 (53%)	83
2013	6	18	12	37	23	31	NA	1	40	84	38 (43%)	2 (2%)	47 (54%)	87
2014	7	18	11	36	23	29	NA	1	40	82	38 (44%)	2 (2%)	45 (52%)	85
2015	6	15	11	38	22	25	NA	1	38	76	36 (45%)	2 (2%)	41 (51%)	79
2016	7	15	11	38	22	26	NA	1	39	78	37 (45%)	2 (2%)	42 (51%)	81
2017	6	13	11	37	20	24	NA	1	36	73	35 (46%)	1 (1%)	39 (52%)	75
2018	6	15	11	35	19	23	NA	1	35	71	35 (48%)	0 (0%)	37 (51%)	72
2019	7	16	10	32	19	22	NA	1	35	68	35 (50%)	0 (0%)	34 (49%)	69
2020	6	17	8	32	21	26	NA	1	34	73	33 (44%)	1 (1%)	41 (54%)	75
2021	6	15	9	28	22	28	NA	2	36	70	35 (47%)	1 (1%)	37 (50%)	73
Total	22	56	37	134	97	145	NA	9	153	327	1201 (49%)	57 (2%)	1205 (49%)	2463

Table 3: List of false positive and false negative sin stocks in IC approach

Panel A reports the top 5 false negative sin firms by industry, i.e., firms which are not considered sin in IC, but show up as strong sin firms in the TA approach. Panel B reports all false positive sin firms by industry, i.e., firms which are considered sin in IC, but are not classified as sin firms in the TA approach. Furthermore, we report the permno, the sin industry, the first and last occurrence, the number of sin years in IC and TA, the mean frequency of sin words ($\overline{\text{Freq.}}$), and the mean market capitalization ($\overline{\text{ME}}$) over the entire history of the sin firm. Within each panel, the firms are sorted in descending order of $\overline{\text{Freq.}}$ within each sin industry. The order of the industries (Industry) implies that a firm has more industry related words of the first mentioned sin industry. A comprehensive list of all sin firms is provided in Table A.2 of the Appendix.

Panel A: False positives - Sin firms only in IC approach								
Permno	Company name	Industry	$\overline{\text{Freq.}}$	$\overline{\text{ME}}$	First	Last	IC	TA
36281	SEABOARD CORP	Alcohol	3	452	1997	2001	5	0
85269	VALERO ENERGY CORP NEW	Alcohol	2	20261	2008	2016	9	0
11262	ELMERS RESTAURANTS INC	Gambling	18	11	2002	2004	3	0
75748	CREATIVE LEARNING PRODUCTS INC	Gambling	15	14	1996	1997	2	0
79678	ACTIVISION BLIZZARD INC	Gambling	4	11226	2008	2008	1	0
79867	SANTA FE FINANCIAL CORP	Gambling	0	13	1998	2003	6	0
Panel B: False negatives - Top 5 sin firms only in TA approach								
18556	GREENLANE HOLDINGS INC	Tobacco	472	59	2019	2021	0	3
29867	PYXUS INTL INC	Tobacco	457	375	1996	2020	0	24
76950	VAPOR CORP	Tobacco	419	13	2014	2015	0	2
91018	CORE MARK HOLDING CO INC	Tobacco	295	886	2005	2020	0	16
85553	HOLTS CIGAR HOLDINGS INC	Tobacco	279	31	1998	2000	0	3
85213	SCHEID VINEYARDS INC	Alcohol	458	11	1997	2005	0	9
86249	CENTRAL EUROPEAN DISTRIBUTN CORP	Alcohol	342	702	1998	2012	0	15
80628	GEERLINGS AND WADE INC	Alcohol	295	15	1996	2002	0	7
80736	ROCK BOTTOM RESTAURANTS INC	Alcohol	228	59	1996	1998	0	3
84062	BJS RESTAURANTS INC	Alcohol	211	563	1996	2021	0	26
87619	MAGNA ENTERTAINMENT CORP	Gambling	742	212	2000	2007	0	8
77803	GTECH HOLDINGS CORP	Gambling	580	1893	1997	2006	0	10
84742	INTERACTIVE SYSTEMS WORLDWIDE IN	Gambling	378	23	1997	2006	0	10
64135	W M S INDUSTRIES INC	Gambling	365	1002	1996	2013	0	18
91079	FORTUNET INC	Gambling	351	75	2006	2008	0	3
82518	RCI HOSPITALITY HOLDINGS INC	Adult Ent./Alcohol	205	98	1996	2021	0	26
86744	PRIVATE MEDIA GROUP INC	Adult Ent.	146	138	1999	2010	0	12
79974	SPICE ENTERTAINMENT COS INC	Adult Ent.	92	31	1996	1997	0	2
78871	INTERACTIVE BRAND DVLPMNT INC	Adult Ent.	89	8	2004	2004	0	1
90218	V C G HOLDING CORP	Adult Ent./Alcohol	84	61	2004	2010	0	7

Table 4: Textual analysis sin stock excess returns over comparables

The table reports monthly excess returns of sin stocks identified with our textual analysis approach over a portfolio of comparables. We report equal-, value-, and sin-weighted returns with respect to the [FF3](#), [CH4](#), and [FF6](#) factor models, respectively. Panel A reports the results of our baseline approach requiring a min. frequency ≥ 20 & unique ≥ 2 . Panels B and C report results with a min. frequency of 100 and a min. frequency of 200, respectively. Furthermore, we report the average number of sin stocks in the long leg (\overline{LL}) and the average number of comparable stocks in the short leg (\overline{SL}) over our sample period from July 1997 to December 2021.

Panel A: SIN TA 20									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Equal-weighted			Value-weighted			Sin-weighted		
α	0 (.03)	.02 (.17)	.05 (.43)	.13 (.73)	.18 (1.06)	.05 (.31)	.29** (2.28)	.29** (2.4)	.34*** (2.74)
MKT	.11*** (4.61)	.1*** (4.03)	.08*** (3.16)	.07 (1.45)	.03 (.59)	.07 (1.35)	.07*** (3.14)	.07*** (2.64)	.05* (1.74)
SMB	.12*** (4.83)	.13*** (4.9)	.13*** (3.78)	.15* (1.69)	.16* (1.72)	.25** (2.04)	.07** (2.14)	.07** (2.15)	.06* (1.65)
HML	.11*** (4.31)	.1*** (3.45)	.14*** (3.13)	.27*** (4.88)	.23*** (3.42)	.12* (1.75)	.09*** (2.81)	.09** (2.48)	.14*** (3.09)
MOM		-.02 (-1.09)	-.02 (-.95)		-.09 (-1.11)	-1 (-1.3)		-.01 (-.23)	0 (-.07)
RMW			0 (-.06)			.24** (2.3)			-.03 (-.7)
CMA			-.1 (-1.26)			.04 (.33)			-.12 (-1.29)
\overline{LL}	84	84	84	84	84	84	84	84	84
\overline{SL}	227	227	227	227	227	227	227	227	227
Obs	294	294	294	294	294	294	294	294	294

Panel B: SIN TA 100									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Equal-weighted			Value-weighted			Sin-weighted		
α	.22 (1.63)	.23* (1.78)	.26* (1.92)	.16 (.80)	.23 (1.19)	.09 (.46)	.36** (2.48)	.36*** (2.64)	.40*** (2.88)
\overline{LL}	58	58	58	58	58	58	58	58	58
\overline{SL}	227	227	227	227	227	227	227	227	227

Panel C: SIN TA 200									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Equal-weighted			Value-weighted			Sin-weighted		
α	.30** (2.04)	.31** (2.16)	.35** (2.31)	.36* (1.66)	.43** (1.99)	.35* (1.65)	.40*** (2.67)	.40*** (2.82)	.43*** (2.90)
\overline{LL}	43	43	43	43	43	43	43	43	43
\overline{SL}	227	227	227	227	227	227	227	227	227

Table 5: Industry classification sin stock excess returns over comparables

The table reports monthly excess returns of sin stocks identified with an industry classification approach over a portfolio of comparables. We report equal-, value-, and sin-weighted returns with respect to the **FF3**, **CH4**, and **FF6** factor models, respectively. Furthermore, we report the average number of sin stocks in the long leg (\overline{LL}) and the average number of comparable stocks in the short leg (\overline{SL}) over our sample period from July 1997 to December 2021.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Equal-weighted			Value-weighted			Sin-weighted		
α	.45*** (3.03)	.48*** (3.3)	.46*** (2.78)	.18 (.98)	.23 (1.31)	.09 (.52)	.55*** (3.57)	.58*** (3.86)	.56*** (3.71)
MKT	.18*** (5.52)	.16*** (4.02)	.16*** (3.83)	.03 (.74)	0 (-.03)	.05 (.84)	.17*** (5.08)	.16*** (3.84)	.16*** (3.66)
SMB	.02 (.3)	.02 (.39)	.08 (.96)	.13 (1.36)	.14 (1.38)	.22* (1.73)	.05 (.86)	.05 (.92)	.09 (1.24)
HML	.18*** (3.53)	.15** (2.53)	.15* (1.85)	.26*** (4.44)	.22*** (2.98)	.09 (1.22)	.16*** (3.54)	.15*** (2.77)	.15* (1.91)
MOM		-.05 (-.73)	-.06 (-.76)		-.09 (-1.05)	-.1 (-1.25)		-.04 (-.65)	-.04 (-.67)
RMW			.12 (1.29)			.24** (2.11)			.1 (1.01)
CMA			-.13 (-.92)			.09 (.75)			-.11 (-.83)
\overline{LL}	45	45	45	45	45	45	43	43	43
\overline{SL}	227	227	227	227	227	227	227	227	227
Obs	294	294	294	294	294	294	294	294	294

Appendix for Measuring Business Social Irresponsibility: The Case of Sin Stocks

Abstract

This document provides further information. Table [A.1](#) shows summary statistics of all words used in our sin dictionary. Table [A.2](#) shows all companies identified by the TA or IC method. Table [A.3](#) shows portfolio returns in excess of the risk-free rate of sin stocks identified by our TA method. Table [A.4](#) shows portfolio returns in excess of the risk-free rate of sin stocks identified using the IC method. Tables [A.5](#) and [A.6](#) show portfolio returns in excess of comparables of TA and IC sin stocks excluding the covid period (1997-2019), respectively.

Table A.1: Sin dictionary summary statistics in 10-K filings

This table reports the words used in our sin dictionary by sin industry. Furthermore, we report the number of overall occurrences (N), the mean (Mean), the standard deviation (SD), the minimum (Min), the median (p50), and the maximum (Max) amount of occurrences in all (not just sin stocks) 10-K filings, respectively. The lower case "s" after a word indicates that we searched for the singular and plural of the specific word. The upper case "S" after a word indicates that we searched for the singular and plural of multiple different spellings of the specific word. The (-) between two words indicates that we searched for cases with a hyphen and cases with a space between the two words. The different shades of gray within the industries indicate the 4 clusters that were used to determine the unique score. Words without gray shading were used to avoid false positive identifications.

	Words	N	Mean	SD	Min	p50	Max
Tobacco	TOBACCOs	4133	26	121	1	1	1885
	VAPING	50	8	13	1	4	70
	VAPORIZER'S ¹	192	6	14	1	2	93
	SNUFF	108	25	50	1	2	222
	CIGARs	480	18	44	1	2	355
	CIGARETTEs	1299	34	79	1	2	556
	CIGARILLOS	38	2	2	1	1	9
	HOOKAHs	34	1	1	1	1	6
	CANNABIS	270	21	55	1	3	416
	HEMP	173	22	45	1	4	291
	MARIJUANA	341	4	6	1	2	40
	SMOKING CESSATION'S ²	551	5	10	1	2	83
Alcohol	ALCOHOL	4646	4	12	1	1	220
	ALCOHOLIC BEVERAGEs	3371	5	6	1	3	82
	BEERs	2974	11	38	1	2	548
	WINEs	3406	11	40	1	1	505
	LIQUOR'S ³	3315	5	13	1	2	209
	SPIRITs	3605	8	33	1	1	506
	GINs	233	3	4	1	2	35
	TEQUILAs	160	8	12	1	2	71
	VODKAs	163	13	27	1	5	161
	WHISKY'S ⁴	227	16	23	1	5	118
	BRANDY	153	5	10	1	2	63
	SCOTCH	197	4	5	1	2	32
	RUMs	158	10	18	1	2	113
	BOURBONs	310	5	10	1	1	79
	CIDERs	194	10	14	1	3	61
	BREWERs	950	9	16	1	2	141
	BREWERYs	917	18	39	1	2	244
	BREWING	902	16	40	1	1	297
	DISTILLERs	213	4	5	1	1	29
	DISTILLERYs	231	10	21	1	2	168
	DISTILLING	198	4	9	1	2	76
	WINEMAKERs	81	2	2	1	2	11
WINEMAKING	105	6	5	1	6	26	
WINERYs	717	10	23	1	1	144	
VINEYARDs	826	15	42	1	2	268	

¹VAPORIZER'S= VAPORIZER + VAPORIZERS + VAPORISER + VAPORISERS

²SMOKING CESSATION'S= SMOKING CESSATION + NICOTINE ADDICTION + TOBACCO ADDICTION; to identify and exclude companies that solely try to help people to stop smoking, we subtract the SMOKING CESSATION'S word counts first from the TOBACCOs word counts and any remaining counts from the CIGARETTEs word counts.

³LIQUOR'S= LIQUEUR + LIQUEURS + LIQUOR + LIQUORS

⁴WHISKY'S= WHISKEY + WHISKEYS + WHISKY + WHISKYS

Table A.1: (continued)

	Words	N	Mean	SD	Min	p50	Max
	VINTNERs	129	3	8	1	1	51
	COFFEE ⁵	3708	14	54	1	1	524
	INDUSTRY ALCOHOL'S ⁶	10782	11	45	1	2	849
Gambling	GAMBLING	1618	5	7	1	2	72
	GAMBLEs	3533	3	10	1	1	174
	GAMBLER	33	2	3	1	1	15
	BETs	957	7	15	1	2	145
	BETTING	662	6	11	1	2	88
	WAGERs	812	9	18	1	4	171
	WAGERING	942	23	48	1	4	436
	WAGERED	547	4	4	1	3	25
	GAMING INDUSTRY'S ⁷	1582	14	16	1	8	77
	BINGO	632	16	50	1	2	326
	HORSE(-)RACINGs	594	11	19	1	2	151
	HORSE RACEs	266	6	6	1	3	32
	LOTTERYs	2286	16	61	1	2	637
	SPORTS BETTING	281	9	27	1	2	281
	CASINOs	3550	51	110	1	3	1156
	HORSE RACETRACKs	115	5	8	1	2	57
	SLOT MACHINES	1093	16	18	1	10	126
	GAMING MACHINES	784	14	37	1	2	333
Adult Entertainment	PORNOGRAPHY'S ⁸	892	1	2	1	1	22
	SEXUALLY(-)EXPLICIT	294	2	1	1	1	10
	SEXUALLY(-)ORIENTED	117	5	8	1	1	41
	ADULT(-)ORIENTED	133	3	3	1	1	16
	ADULT BUSINESS	20	2	1	1	1	5
	ADULT COMPANIES	6	1	0	1	1	1
	ADULT ENTERTAINMENT	134	14	14	1	10	54
	ADULT INDUSTRY	22	2	1	1	2	6
	NUDITY	75	1	0	1	1	4
	ADULT ENTERTAINERS	6	6	2	5	6	8
	ADULT PUBLICATIONs	21	2	1	1	2	4
	ADULT MAGAZINEs	39	2	1	1	1	4
	ADULT WEBSITE'S ⁹	29	3	3	1	2	12
	ADULT MEDIA	15	24	16	1	35	42
	ADULT TELEVISION'S ¹⁰	83	5	6	1	3	26
	ADULT MATERIAL	49	1	0	1	1	2
	ADULT PAY(-)TV'S ¹¹	34	3	2	1	3	10
	ADULT INTERNET	33	4	3	1	2	14
	ADULT NIGHTCLUBs	33	12	4	3	12	17

⁵To avoid false positives for COFFEE companies, we first subtract the COFFEE word counts from the BREWING word counts and the remaining counts from the BREWER word counts.

⁶INDUSTRY ALCOHOL'S= SPECIALTY ALCOHOLs + ETHANOL + POLYETHYLENE + ETHYLE + POLYMERS; to identify and exclude companies that solely use alcohol for production purposes, we subtract the INDUSTRY ALCOHOL'S word counts first from the DISTILLERs word counts and any remaining counts from the ALCOHOL word counts.

⁷GAMING INDUSTRY'S= GAMING INDUSTRY + GAMING LICENSE + GAMING AUTHORITIES

⁸PORNOGRAPHY'S= PORN + PORNOGRAPHY

⁹ADULT WEBSITE'S= ADULT WEBSITEs + ADULT WEB SITES + ADULT INTERNET WEBSITEs

¹⁰ADULT TELEVISION'S= ADULT TELEVISION + ADULT(-)TV + ADULT PROGRAMMING

¹¹ADULT PAY(-)TV'S= ADULT PAY(-)TV + ADULT PAY(-)PER(-)VIEW + ADULT(-)PPV + ADULT(-)VOD

Table A.1: (continued)

Words	N	Mean	SD	Min	p50	Max
ADULT CLUBs	22	3	2	1	3	7
CABARETs	167	18	36	1	2	162
GENTLEMEN CLUBs	33	4	2	1	5	8
STRIP CLUBs	4	1	0	1	1	1
ADULT STUDIOs	18	2	1	1	2	3
ADULT VIDEOS	26	3	2	1	2	7
ADULT SERVICEs	58	3	6	1	2	34
ADULT FILM'S ¹²	60	6	6	1	5	39
ADULT MOTION PICTUREs	19	2	1	1	3	3
TOPLESS	41	6	4	1	4	22
ADULT CONTENT	173	4	6	1	1	38
EXPLICIT SEXUAL CONTENT	7	1	0	1	1	1
SEXUAL CONTENT	42	1	0	1	1	1
ADULT(-)RELATED CONTENT	19	3	1	1	3	4

¹²ADULT FILM'S= ADULT FILM's + ADULT MOVIE's + EROTIC FILM's

Table A.2: List of sin stocks

The table reports all identified sin firms, which are sin firms in both TA and IC at some point in time (Panel A), only sin firms in TA (Panel B), and only sin firms in IC (Panel C), respectively. Furthermore, we report the permno, the sin industry, the first and last occurrence, the number of sin years in IC and TA, the mean frequency of sin words ($\overline{\text{Freq.}}$), and the mean market capitalization ($\overline{\text{ME}}$) throughout the history of the sin firm. Within each panel, the firms are sorted in descending order of $\overline{\text{Freq.}}$ within each sin industry. The order of the industries (Industry) implies that a firm has more industry-related words of the first mentioned sin industry.

Panel A: Sin firms in both TA + IC approach								
Permno	Company name	Industry	$\overline{\text{Freq.}}$	$\overline{\text{ME}}$	First	Last	IC	TA
86946	REYNOLDS AMERICAN INC	Tobacco	1868	22591	2000	2016	17	17
17279	LORILLARD INC	Tobacco	844	15208	2008	2014	7	7
87816	ROCK CREEK PHARMACEUTICALS INC	Tobacco	830	201	2000	2014	12	15
14523	22ND CENTURY GROUP INC	Tobacco	667	238	2014	2021	8	8
75233	VECTOR GROUP LTD	Tobacco	623	1270	1996	2021	26	26
16083	TURNING POINT BRANDS INC	Tobacco	580	549	2016	2021	6	6
79689	STANDARD DIVERSIFIED INC	Tobacco	578	135	2018	2019	2	2
81667	MAFCO CONSOLIDATED GROUP INC	Tobacco	511	590	1996	1996	1	1
13901	ALTRIA GROUP INC	Tobacco/Alcohol	456	98133	1996	2021	26	26
16555	UNIVERSAL CORPORATION	Tobacco	444	1191	1996	2021	13	25
92602	PHILIP MORRIS INTERNATIONAL INC	Tobacco	422	127387	2008	2021	14	14
83851	CARIBBEAN CIGAR CO	Tobacco	414	25	1997	1998	2	2
76597	NABISCO GROUP HOLDING CORP	Tobacco	411	8632	1996	1999	3	4
83819	CONSOLIDATED CIGAR HOLDINGS INC	Tobacco	407	220	1996	1997	2	2
26710	LOEWS CORP	Tobacco	385	13253	1996	2007	12	12
84374	SWISHER INTERNATIONAL GROUP INC	Tobacco	369	83	1996	1998	3	3
84580	GENERAL CIGAR HOLDINGS INC	Tobacco	345	285	1997	1999	3	3
15077	U S T INC	Tobacco/Alcohol	317	6511	1996	2007	12	12
82649	SCHWEITZER MAUDUIT INTL INC	Tobacco	268	803	1996	2021	22	25
12044	CULBRO CORP	Tobacco	53	257	1996	1996	1	1
79615	OROAMERICA INC	Tobacco	46	41	1997	1999	1	3
20654	DUCKHORN PORTFOLIO INC	Alcohol	844	2513	2021	2021	1	1
69796	CONSTELLATION BRANDS INC	Alcohol	685	10629	1997	2021	25	25
82634	BOSTON BEER INC	Alcohol	615	1414	1996	2021	26	26
82176	CRAFT BREW ALLIANCE INC	Alcohol	588	118	1996	2019	24	24
80955	WILLAMETTE VALLEY VINYDS INC	Alcohol	540	22	1996	2021	26	26
86301	GOLDEN STATE VINTNERS INC	Alcohol	502	24	1999	2003	5	5
82627	PETES BREWING CO	Alcohol	481	65	1996	1997	2	2
22569	CHALONE WINE GROUP LTD	Alcohol	478	96	1996	2003	7	7
91192	CASTLE BRANDS INC	Alcohol	476	103	2007	2019	13	13
16887	EASTSIDE DISTILLING INC	Alcohol/Tobacco	470	29	2017	2020	4	4
59248	MOLSON COORS BEVERAGE CO	Alcohol	428	7676	1996	2021	26	26
86843	RAVENSWOOD WINERY INC	Alcohol	419	57	1999	2000	2	2
22417	WINC INC	Alcohol	407	69	2021	2021	1	1
83525	LION BREWERY INC	Alcohol	380	18	1996	1997	2	2
77984	CRUZAN INTERNATIONAL INC	Alcohol	348	65	1996	2005	10	10
82808	NOR WESTER BREWING INC	Alcohol	315	9	1996	1996	1	1
84539	INDEPENDENCE BREWING COMPANY	Alcohol	305	1	1997	1998	2	2
10225	BEAM INC	Alcohol/Tobacco	299	8374	1996	2013	18	18
82710	PYRAMID BREWERIES INC	Alcohol	296	22	1996	2007	12	12
89890	ASCONI CORP	Alcohol	294	56	2003	2004	2	2
13970	TRUETT HURST INC	Alcohol	278	11	2013	2018	6	6
82517	R H PHILLIPS INC	Alcohol	260	18	1996	1999	4	4
83624	BIG BUCK BREWERY AND STEAKHS INC	Alcohol	226	13	1996	2001	6	6

Table A.2: (continued)

Permno	Company name	Industry	Freq.	ME	First	Last	IC	TA
29946	BROWN FORMAN CORP	Alcohol	225	7402	1996	2021	26	26
12226	M G P INGREDIENTS INC	Alcohol	204	387	1996	2021	25	20
38608	GENESEE CORP	Alcohol	175	42	1996	2002	4	6
79715	M B C HOLDING CO	Alcohol	148	9	1996	2000	5	5
79289	ROBERT MONDAVI CORP THE	Alcohol	144	300	1996	2004	9	9
83240	FREDERICK BREWING CO	Alcohol	109	7	1996	1998	3	3
59184	ANHEUSER BUSCH COS INC	Alcohol	109	35090	1996	2007	12	12
85456	BERINGER WINE ESTATES HLDGNS INC	Alcohol	72	736	1998	2000	3	2
20415	LEUCADIA NATIONAL CORP	Alcohol/Gambling	63	4747	1998	2012	1	13
84185	ANDEAN DEVELOPMENT CORP	Alcohol	18	9	1997	1999	2	1
86446	J C C HOLDING CO	Gambling	1156	17	1998	1999	2	2
19333	RUSH STREET INTERACTIVE INC	Gambling	768	738	2020	2021	2	2
76746	POWERHOUSE TECHNOLOGIES INC	Gambling	762	107	1996	1998	3	3
79171	HOLLYWOOD CASINO CORP	Gambling	718	128	1996	2001	6	6
79338	SCIENTIFIC GAMES CORP	Gambling	642	1612	1996	2021	26	26
80563	PENN NATIONAL GAMING INC	Gambling	628	2501	1996	2021	24	26
81676	TRUMP HOTELS & CASINO RESRTS INC	Gambling	597	98	1996	2003	8	8
56434	SHOWBOAT INC	Gambling	591	376	1996	1997	2	2
86447	CAESARS ENTERTAINMENT INC	Gambling	581	3701	1999	2004	6	6
79791	CENTURY CASINOS INC	Gambling	539	119	1996	2021	26	26
18425	BALLY'S CORP	Gambling	525	1493	2019	2021	3	3
65533	MANDALAY RESORT GROUP	Gambling	496	2470	1996	2005	10	10
77028	GRAND CASINOS INC	Gambling	462	568	1996	1997	2	2
80153	ANCHOR GAMING	Gambling	461	746	1996	2001	6	6
78147	M T R GAMING GROUP	Gambling	440	155	1996	2013	18	18
42140	PINNACLE ENTERTAINMENT INC	Gambling	437	779	1996	2015	20	20
79026	CHURCHILL DOWNS INC	Gambling	432	1774	1996	2021	16	25
14882	CAESARS ENTERTAINMENT INC DE	Gambling	425	5867	2014	2021	8	8
86578	GOLDEN ENTERTAINMENT INC	Gambling/Alcohol	420	281	1999	2021	22	22
90505	LAS VEGAS SANDS CORP	Gambling	420	34734	2004	2021	18	18
83479	AMERICAN WAGERING INC	Gambling	414	65	1997	2000	4	4
77897	ISLE OF CAPRI CASINOS INC	Gambling/Alcohol	398	372	1996	2016	20	20
50657	GREATER BAY CASINO CORP DEL	Gambling	396	8	1996	1997	2	2
79758	BOYD GAMING CORP	Gambling	386	1971	1996	2021	26	26
79795	AMERISTAR CASINOS INC	Gambling/Alcohol	384	682	1996	2012	17	17
79790	EMPIRE RESORTS INC	Gambling	380	178	1996	2018	22	22
18778	GOLDEN NUGGET ONLINE GAMING INC	Gambling	337	596	2020	2021	2	2
79606	LADY LUCK GAMING CORP	Gambling	333	34	1996	1998	3	3
76090	HARRAHS ENTERTAINMENT INC	Gambling	325	6560	1996	2007	12	11
79490	FULL HOUSE RESORTS INC	Gambling	323	64	1996	2021	21	22
89332	DOVER DOWNS GAMING & ENTMT INC	Gambling	318	72	2002	2018	17	17
79192	STATION CASINOS INC	Gambling	303	1885	1997	2006	10	10
89533	WYNN RESORTS LTD	Gambling	293	10030	2002	2021	20	20
38149	BALLY TECHNOLOGIES INC	Gambling	290	1077	1996	2014	17	19
79308	SODAK GAMING INC	Gambling	290	228	1996	1998	2	3
16019	RED ROCK RESORTS INC	Gambling	288	2020	2016	2021	6	6
60709	NEVADA GOLD & CASINOS INC	Gambling	284	56	2002	2018	17	17
79153	BLACK HAWK GAMING & DEV CO INC	Gambling	280	25	1996	2000	5	5
80199	HARVEYS CASINO RESORTS	Gambling	276	210	1996	1998	3	3
88925	G B HOLDINGS INC	Gambling	276	30	2001	2003	3	3
75900	AZTAR CORP	Gambling	276	610	1996	2005	10	10
82181	VENTURE CATALYST INC	Gambling	272	16	1996	2001	4	5

Table A.2: (continued)

Permno	Company name	Industry	$\overline{\text{Freq.}}$	$\overline{\text{ME}}$	First	Last	IC	TA
15129	INSPIRED ENTERTAINMENT INC	Gambling	266	211	2017	2021	3	5
18835	DRAFTKINGS INC	Gambling	261	14827	2020	2021	2	2
83795	SILICON GAMING INC	Gambling	261	109	1996	1998	1	3
78867	ARGOSY GAMING CO	Gambling	257	541	1996	2004	9	9
17252	PLAYAGS INC	Gambling	256	437	2018	2021	4	4
23309	HILTON HOTELS CORP	Gambling	254	6966	1996	1997	2	2
90911	TRUMP ENTERTAINMENT RESORTS INC	Gambling	244	313	2005	2008	4	4
87005	YOUBET COM	Gambling	234	76	1999	2009	11	10
13267	CAESARS ENTERTAINMENT CORP	Gambling	232	3874	2012	2019	8	8
79507	MONARCH CASINO & RESORT INC	Gambling	229	362	1996	2021	26	26
83458	RIVIERA HOLDINGS CORP	Gambling/Alcohol	228	104	1996	2008	13	13
10857	PLAYERS INTERNATIONAL INC	Gambling	225	198	1996	1999	4	4
81145	LITTLEFIELD CORP	Gambling	225	18	1996	2002	6	5
11891	M G M RESORTS INTERNATIONAL	Gambling/Alcohol	223	9715	1996	2021	26	25
83529	MULTIMEDIA GAMES HOLDING CO INC	Gambling	221	272	1996	2014	12	17
81182	CANTERBURY PARK HOLDING CORP	Gambling	218	44	1996	2021	7	26
45277	INTERNATIONAL GAME TECHNOLOGY	Gambling	217	5927	1996	2014	19	19
80225	BOARDWALK CASINO INC	Gambling	215	37	1996	1997	2	2
16001	PINNACLE ENTERTAINMENT INC NEW	Gambling	214	1348	2016	2017	2	2
78194	PRESIDENT CASINOS INC	Gambling	213	19	1997	1998	2	2
78023	CASINO MAGIC CORP	Gambling	213	65	1996	1997	2	1
11561	GLOBAL CASINOS INC	Gambling	204	5	1996	1999	4	4
85386	NEVSTAR GAMING CORP	Gambling	199	9	1997	1998	2	2
79578	CASINO RESOURCE CORP	Gambling	196	13	1996	1998	3	2
78021	BOOMTOWN INC	Gambling	182	38	1996	1996	1	1
82776	I T T CORP NEV	Gambling	176	5047	1996	1996	1	1
84673	W H G RESORTS & CASINOS INC	Gambling	172	66	1997	1997	1	1
79297	PRIMADONNA RESORTS INC	Gambling	170	496	1996	1997	2	2
49286	FLORIDA GAMING CORP	Gambling	165	20	1996	1997	2	2
81478	COLORADO CASINO RESORTS INC	Gambling/Alcohol	152	54	1996	1998	3	3
90176	ENTERTAINMENT GAMING ASIA INC	Gambling	147	65	2004	2016	6	13
12395	RIO HOTEL & CASINO INC	Gambling	143	383	1996	1997	2	2
91687	SANDS REGENT	Gambling	143	29	1996	2006	11	11
60441	MIRAGE RESORTS INC	Gambling	139	3402	1996	1999	4	4
90327	W P T ENTERPRISES INC	Gambling	138	78	2005	2007	2	3
80282	STRATOSPHERE CORP	Gambling	128	49	1996	1996	1	1
19893	FUBOTV INC	Gambling	127	2140	2020	2021	1	2
84026	DOVER DOWNS ENTERTAINMENT INC	Gambling	127	135	1997	2000	4	4
21036	ESPORTS TECHNOLOGIES INC	Gambling	127	447	2021	2021	1	1
85037	C C A COMPANIES INC	Gambling	127	31	1998	1999	2	2
80801	BALLY'S GRAND INC	Gambling	123	383	1996	1996	1	1
46181	JACKPOT ENTERPRISES INC	Gambling	115	103	1996	2000	5	5
68567	SANTA FE GAMING CORP	Gambling	107	8	1996	1998	3	3
75781	EUROPA CRUISES CORP	Gambling	107	17	1996	1997	2	2
81158	TRANS WORLD GAMING CORP	Gambling	91	3	1996	1996	1	1
11174	CROWN GROUP INC	Gambling	87	36	1996	1999	4	3
79502	LANDRYS RESTAURANTS INC	Gambling/Alcohol	80	476	2004	2009	5	6
79980	HOLLY HOLDINGS INC	Gambling	47	8	1996	1997	2	1
91050	M G T CAPITAL INVESTMENTS INC	Gambling	40	12	2013	2015	3	3
16975	ALLIED ESPORTS ENTERTAINMENT INC	Gambling	28	61	2019	2020	2	1
13169	ZYNGA INC	Gambling	19	3586	2011	2021	7	2

Panel B: Sin firms only in TA approach

Permno	Company name	Industry	Freq.	ME	First	Last	IC	TA
18556	GREENLANE HOLDINGS INC	Tobacco	472	59	2019	2021	0	3
29867	PYXUS INTL INC	Tobacco	457	375	1996	2020	0	24
76950	VAPOR CORP	Tobacco	419	13	2014	2015	0	2
91018	CORE MARK HOLDING CO INC	Tobacco	295	886	2005	2020	0	16
85553	HOLTS CIGAR HOLDINGS INC	Tobacco	279	31	1998	2000	0	3
85016	800 JR CIGAR INC	Tobacco	260	242	1997	1999	0	3
16715	STANDARD COMMERCIAL CORP	Tobacco	248	160	1996	2004	0	9
85310	PREMIUM CIGARS INTERNATIONAL LTD	Tobacco	247	5	1997	1998	0	2
79500	SYNERGY BRANDS INC	Tobacco	175	11	1997	2007	0	11
81670	M & F WORLDWIDE CORP	Tobacco	145	328	1996	2010	0	15
80483	GRAND HAVANA ENTERPRISES INC	Tobacco	116	12	1996	1998	0	3
82171	AMCON DISTRIBUTING CO	Tobacco	92	32	1996	2021	0	25
18642	WHOLE EARTH BRANDS INC	Tobacco	82	416	2020	2021	0	2
85200	PC411 INC	Tobacco	66	1	1998	1998	0	1
91507	SUSSER HOLDINGS CORP	Tobacco/Alcohol	63	509	2007	2013	0	7
15724	PERFORMANCE FOOD GROUP CO	Tobacco	63	4853	2019	2021	0	3
18844	A R K O CORP	Tobacco	62	1104	2020	2021	0	2
14076	MURPHY USA INC	Tobacco	59	3113	2013	2021	0	9
86993	PANTRY INC	Tobacco	59	384	1999	2014	0	11
92553	SHINER INTERNATIONAL INC	Tobacco	58	26	2008	2011	0	4
57154	GLATFELTER P H CO	Tobacco	57	786	1997	1997	0	1
13831	C S T BRANDS INC	Tobacco	57	3192	2013	2016	0	4
12918	AMERICAN FILTRONA CORP	Tobacco	45	159	1996	1996	0	1
85982	GRIFFIN LAND AND NURSERIES INC	Tobacco	43	107	1998	2009	0	12
10790	SHOREWOOD PACKAGING CORP INC	Tobacco	42	429	1996	1999	0	2
77015	7 ELEVEN INC	Tobacco	40	1226	1997	2004	0	8
21742	CASEYS GENERAL STORES INC	Tobacco	39	3088	2006	2021	0	16
54843	MONARCH SERVICES INC	Tobacco	37	3	2002	2004	0	3
91215	DELEK U S HOLDINGS INC NEW	Tobacco	35	1569	2006	2021	0	12
84132	STAGE STORES INC	Tobacco	25	620	1998	1999	0	2
10908	UNI MARTS INC	Tobacco	25	13	1998	2003	0	4
14682	RAYONIER ADVANCED MATERIALS INC	Tobacco	20	669	2016	2016	0	1
85213	SCHEID VINEYARDS INC	Alcohol	458	11	1997	2005	0	9
86249	CENTRAL EUROPEAN DISTRIBUTN CORP	Alcohol	342	702	1998	2012	0	15
80628	GEERLINGS AND WADE INC	Alcohol	295	15	1996	2002	0	7
80736	ROCK BOTTOM RESTAURANTS INC	Alcohol	228	59	1996	1998	0	3
84062	BJS RESTAURANTS INC	Alcohol	211	563	1996	2021	0	26
89102	GRANITE CITY FOOD & BREWERY LTD	Alcohol	182	27	2001	2012	0	12
54244	FLANIGANS ENTERPRISES INC	Alcohol	162	21	1996	2020	0	25
16590	IFRESH INC	Alcohol	103	28	2020	2020	0	1
78913	BEV TYME INC NEW	Alcohol	81	0	1996	1996	0	1
92449	REEDS INC	Alcohol	66	47	2007	2021	0	15
16175	LONG ISLAND ICED TEA CORP	Alcohol	64	32	2016	2016	0	1
13874	FAIRWAY GROUP HOLDINGS CORP	Alcohol	62	211	2014	2015	0	2
85167	CAPITAL BEVERAGE CORP	Alcohol	57	8	1998	2000	0	3
11786	S V B FINANCIAL GROUP	Alcohol	57	7609	2004	2021	0	18
13548	BLOOMIN BRANDS INC	Alcohol	57	2101	2012	2021	0	10
13408	IGNITE RESTAURANT GROUP INC	Alcohol	55	197	2012	2016	0	5
88982	SMITH & WOLLENSKY REST GP INC	Alcohol	53	44	2005	2006	0	2
77157	O I GLASS INC	Alcohol	53	3496	2010	2020	0	10
15383	ONE GROUP HOSPITALITY INC	Alcohol	52	65	2015	2016	0	2
13869	DIVERSIFIED RESTAURANT HLDGS INC	Alcohol	50	72	2013	2018	0	6

Table A.2: (continued)

Permno	Company name	Industry	$\overline{\text{Freq.}}$	$\overline{\text{ME}}$	First	Last	IC	TA
76702	O S I RESTAURANT PARTNERS INC	Alcohol	47	2415	1996	2006	0	11
87162	1 800 FLOWERS COM INC	Alcohol	46	138	2005	2013	0	9
85222	CHAMPPS ENTERTAINMENT INC DEL	Alcohol	46	93	2002	2007	0	6
85651	F F P MARKETING CO	Alcohol	45	8	2001	2001	0	1
15167	SHAKE SHACK INC	Alcohol	43	1537	2015	2020	0	6
89904	BUFFALO WILD WINGS INC	Alcohol	43	1297	2003	2016	0	14
18570	BOB EVANS FARMS INC	Alcohol	42	917	2008	2012	0	5
90865	KONA GRILL INC	Alcohol	37	88	2005	2018	0	14
13520	CHUYS HOLDINGS INC	Alcohol	36	466	2012	2021	0	10
15426	FOGO DE CHAO INC	Alcohol	35	386	2015	2017	0	3
13524	DEL FRISCOS RESTAURANT GROUP INC	Alcohol	35	401	2012	2018	0	7
85214	TOTAL ENTERTAINMENT REST CORP	Alcohol	35	74	2000	2004	0	5
13455	CHANTICLEER HOLDINGS INC	Alcohol	34	24	2013	2015	0	3
88031	HANSEN NATURAL CORP	Alcohol	34	57	2001	2003	0	3
15648	J ALEXANDERS HOLDINGS INC	Alcohol	34	140	2015	2020	0	6
87532	READING INTERNATIONAL INC	Alcohol	33	238	2016	2021	0	6
14329	ARAMARK	Alcohol	32	8696	2014	2021	0	8
90359	HERITAGE OAKS BANCORP	Alcohol	32	79	2004	2004	0	1
17322	IPIC ENTERTAINMENT INC	Alcohol	31	22	2018	2018	0	1
77088	AVADO BRANDS INC	Alcohol	31	283	1996	2000	0	5
93397	LIMONEIRA CO	Alcohol	31	347	2016	2019	0	4
86534	P F CHANGS CHINA BISTRO INC	Alcohol	30	776	2006	2011	0	6
85378	IL FORNAIO AMERICA CORP	Alcohol	30	61	1997	2000	0	4
81655	DARDEN RESTAURANTS INC	Alcohol	30	10174	2012	2021	0	10
78099	VALUE HOLDINGS INC	Alcohol	30	4	1996	1996	0	1
77405	LONE STAR STEAKHOUSE & SALOON IN	Alcohol	30	502	1996	2005	0	10
89567	COSI INC	Alcohol	29	91	2003	2015	0	13
17369	BURGERFI INTERNATIONAL INC	Alcohol	29	240	2020	2020	0	1
84203	BBQ HOLDINGS INC	Alcohol	29	82	2010	2021	0	12
13959	NOODLES & CO	Alcohol	28	700	2013	2015	0	3
79994	GRILL CONCEPTS INC	Alcohol	28	17	1996	2007	0	10
18898	KURA SUSHI USA INC	Alcohol	28	239	2019	2021	0	3
57007	TEJON RANCH CO	Alcohol	28	460	1996	2021	0	11
13375	FIESTA RESTAURANT GROUP INC	Alcohol	28	951	2014	2017	0	4
90871	RUTHS HOSPITALITY GROUP INC	Alcohol	28	414	2005	2021	0	17
76385	MAIN STREET RESTAURANT GROUP INC	Alcohol	27	47	2000	2005	0	6
82482	JERRYS FAMOUS DELI INC	Alcohol	27	19	1997	2000	0	4
16633	I C C HOLDINGS INC	Alcohol	26	45	2017	2020	0	4
89548	DRIVE SHACK INC	Alcohol	26	179	2019	2021	0	3
76318	BOSTON RESTAURANT ASSOCIATES INC	Alcohol	26	7	1999	1999	0	1
90427	TEXAS ROADHOUSE INC	Alcohol	26	2424	2005	2021	0	17
89829	ANCHOR GLASS CONTAINER CORP NEW	Alcohol	26	279	2003	2004	0	2
14131	CLUBCORP HOLDINGS INC	Alcohol	26	1144	2013	2014	0	2
91117	MORTONS RESTAURANT GROUP INC NEW	Alcohol	26	131	2006	2010	0	5
80583	WESTERN COUNTRY CLUBS INC	Alcohol	25	7	1996	1996	0	1
79732	REUNION INDUSTRIES INC	Alcohol	25	13	1996	1999	0	4
77902	CHEESECAKE FACTORY INC	Alcohol	25	1817	2005	2013	0	6
86302	HERITAGE COMMERCE CORP	Alcohol	25	118	2002	2003	0	2
85849	STEAKHOUSE PARTNERS INC	Alcohol	24	14	1998	2000	0	3
12348	BRAVO BRIO RESTAURANT GROUP INC	Alcohol	24	215	2010	2017	0	8
93347	CHINA JO JO DRUGSTORES INC	Alcohol	24	35	2011	2011	0	1
90286	MCCORMICK & SCHMICKS SEAFOOD RES	Alcohol	24	195	2004	2010	0	7

Table A.2: (continued)

Permno	Company name	Industry	Freq.	ME	First	Last	IC	TA
89247	ARAMARK CORP	Alcohol	24	2710	2002	2006	0	5
86858	BUCA INC	Alcohol	24	127	2001	2007	0	7
93069	FUTURE FINTECH GROUP INC	Alcohol	23	56	2010	2017	0	3
23297	BRINKER INTERNATIONAL INC	Alcohol	23	1954	2020	2021	0	2
77448	RARE HOSPITALITY INTL INC	Alcohol	23	1011	2005	2006	0	2
77801	AUSTINS INTERNATIONAL INC	Alcohol	22	7	1996	1996	0	1
89499	LANCER CORP TX	Alcohol	22	158	2004	2004	0	1
89453	RED ROBIN GOURMET BURGERS INC	Alcohol	22	533	2002	2021	0	9
76360	GOOD TIMES RESTAURANTS INC	Alcohol	22	61	2018	2021	0	2
83362	MEXICAN RESTAURANTS INC	Alcohol	22	21	2002	2009	0	8
57568	BALL CORP	Alcohol	22	4522	2007	2007	0	1
49736	GROUND ROUND RESTAURANTS INC	Alcohol	22	27	1996	1996	0	1
76223	O CHARLEYS INC	Alcohol	21	193	2006	2011	0	5
76718	BERTUCCIS HOLDING CORP	Alcohol	21	52	1996	1997	0	2
86165	ALTO INGREDIENTS INC	Alcohol	20	350	2021	2021	0	1
80359	QUALITY DINING INC	Alcohol	20	29	1998	2001	0	4
87619	MAGNA ENTERTAINMENT CORP	Gambling	742	212	2000	2007	0	8
77803	GTECH HOLDINGS CORP	Gambling	580	1893	1997	2006	0	10
84742	INTERACTIVE SYSTEMS WORLDWIDE IN	Gambling	378	23	1997	2006	0	10
64135	W M S INDUSTRIES INC	Gambling	365	1002	1996	2013	0	18
91079	FORTUNET INC	Gambling	351	75	2006	2008	0	3
14252	GAMING & LEISURE PROPERTIES INC	Gambling	339	4505	2013	2013	0	1
14304	CAESARS ACQUISITION CO	Gambling	325	1463	2013	2016	0	4
85550	GAMETECH INTERNATIONAL INC	Gambling	324	48	1998	2010	0	10
17989	STUART ENTERTAINMENT INC	Gambling	273	12	1997	1997	0	1
79850	PROGRESSIVE GAMING INTL CORP	Gambling	231	127	1996	2007	0	12
79078	CASINO DATA SYSTEMS	Gambling	224	83	1996	2000	0	5
67619	INTERNATIONAL THOROUGHbred BRDRS	Gambling	202	49	1996	1997	0	2
12627	AMERICAN VANTAGE COS	Gambling	195	18	1996	1998	0	3
79675	ACRES GAMING INC	Gambling	172	56	1996	2003	0	7
88459	PACIFICNET INC	Gambling	170	61	2006	2007	0	2
78200	S H F L ENTERTAINMENT INC	Gambling	160	437	1996	2012	0	17
80357	GAMING PARTNERS INTL CORP	Gambling	159	64	1997	2018	0	20
80447	GAME FINANCIAL CORP	Gambling	159	34	1996	1996	0	1
83910	TRANSACT TECHNOLOGIES INC	Gambling	135	78	2001	2021	0	21
79172	INNOVATIVE GAMING CP OF AMERICA	Gambling	127	16	1996	2001	0	6
84600	COLONIAL DOWNS HOLDINGS INC	Gambling	122	8	1997	1999	0	3
90958	POKERTEK INC	Gambling	120	35	2005	2013	0	9
90885	EVERI HOLDINGS INC	Gambling	119	673	2005	2021	0	17
80562	P D S GAMING CORP	Gambling	118	10	1997	2003	0	7
89978	CASH SYSTEMS INC	Gambling	115	126	2004	2007	0	4
16874	ACCEL ENTERTAINMENT INC	Gambling	105	1083	2020	2021	0	2
77238	NUWAY MEDICAL INC	Gambling	99	6	1996	2002	0	7
18573	SCIPLAY CORP	Gambling	97	311	2019	2021	0	3
87244	LEISURE TIME CASINOS & RESRTS IN	Gambling	89	13	2000	2000	0	1
85586	ARK RESTAURANTS CORP	Gambling/Alcohol	83	72	2006	2021	0	6
77351	LAS VEGAS ENTERTAINMENT NTWK INC	Gambling	79	9	1996	1998	0	3
20447	PLAYTIKA HOLDING CORP	Gambling	75	7083	2021	2021	0	1
45509	INTERNATIONAL LOTTERY & T S INC	Gambling	73	14	1996	1997	0	2
91096	MORGANS HOTEL GROUP CO	Gambling	68	305	2007	2009	0	3
85307	ON STAGE ENTERTAINMENT INC	Gambling	58	20	1997	1998	0	2
84005	GEMSTAR TV GUIDE INTL INC	Gambling	56	1804	2002	2007	0	6

Table A.2: (continued)

Permno	Company name	Industry	$\overline{\text{Freq.}}$	$\overline{\text{ME}}$	First	Last	IC	TA
20277	PLAYSTUDIOS INC	Gambling	48	434	2021	2021	0	1
83403	PLANET HOLLYWOOD INTL INC	Gambling	43	1286	1997	1997	0	1
84226	INTERNATIONAL SPEEDWAY CORP	Gambling	42	704	2008	2010	0	3
81486	INTERACTIVE FLIGHT TECHS INC	Gambling	36	44	1996	1999	0	3
92555	CAPE BANCORP INC	Gambling	36	110	2010	2014	0	4
86982	S S P SOLUTIONS INC	Gambling	35	15	2002	2002	0	1
85579	SPORTSLINE COM INC	Gambling	35	90	2000	2002	0	3
88231	I A GLOBAL INC	Gambling	34	32	2004	2005	0	2
34067	MAXXAM INC	Gambling	32	180	1998	2008	0	10
12279	FIRST ENTERTAINMENT INC	Gambling	30	4	1997	1997	0	1
76548	DIGITAL BIOMETRICS INC	Gambling	25	53	1996	2000	0	4
50550	PERINI CORP	Gambling	23	1147	2007	2008	0	2
81242	GLOBAL PAYMENT TECHNOLOGIES INC	Gambling	21	39	1998	2000	0	3
38914	CERIDIAN CORP	Gambling	21	2792	1996	1996	0	1
75593	T S F COMMUNICATIONS CORP	Gambling	21	93	1996	1996	0	1
82518	RCI HOSPITALITY HOLDINGS INC	Adult Ent./Alcohol	205	98	1996	2021	0	26
86744	PRIVATE MEDIA GROUP INC	Adult Ent.	146	138	1999	2010	0	12
79974	SPICE ENTERTAINMENT COS INC	Adult Ent.	92	31	1996	1997	0	2
78871	INTERACTIVE BRAND DVLPMNT INC	Adult Ent.	89	8	2004	2004	0	1
90218	V C G HOLDING CORP	Adult Ent./Alcohol	84	61	2004	2010	0	7
85838	NEW FRONTIER MEDIA INC	Adult Ent.	78	93	1998	2012	0	15
12793	FRIENDFINDER NETWORKS INC	Adult Ent.	64	22	2011	2012	0	2
76218	PLAYBOY ENTERPRISES INC	Adult Ent./Tobacco	53	319	1998	2007	0	6
19644	PLBY GROUP INC	Adult Ent.	38	1102	2021	2021	0	1
Panel C: Sin firms only in IC approach								
36281	SEABOARD CORP	Alcohol	3	452	1997	2001	5	0
85269	VALERO ENERGY CORP NEW	Alcohol	2	20261	2008	2016	9	0
11262	ELMERS RESTAURANTS INC	Gambling	18	11	2002	2004	3	0
75748	CREATIVE LEARNING PRODUCTS INC	Gambling	15	14	1996	1997	2	0
79678	ACTIVISION BLIZZARD INC	Gambling	4	11226	2008	2008	1	0
79867	SANTA FE FINANCIAL CORP	Gambling	0	13	1998	2003	6	0

Table A.3: Textual analysis sin stock excess returns over risk-free rate

The table reports monthly excess returns of sin stocks identified with our textual analysis approach over the risk-free rate. We report equal-, value-, and sin-weighted returns with respect to the [FF3](#), [CH4](#), and [FF6](#) factor models, respectively. Panel A reports the results of our baseline approach requiring a min. frequency ≥ 20 & unique ≥ 2 . Panels B and C report results with a min. frequency of 100 and a min. frequency of 200, respectively. Furthermore, we report the average number of sin stocks in the long leg (\overline{LL}) over our sample period from July 1997 to December 2021.

Panel A: SIN TA 20									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Equal-weighted			Value-weighted			Sin-weighted		
α	.20 (1.01)	.35* (1.74)	.22 (1.15)	.21 (.86)	.31 (1.28)	-.10 (-.46)	.49** (2.58)	.63*** (3.33)	.51*** (3.04)
MKT	.93*** (14.01)	.83*** (11.77)	.86*** (12.47)	.8*** (13.03)	.74*** (11.45)	.88*** (15.75)	.89*** (15.66)	.8*** (12.9)	.83*** (13.64)
SMB	.66*** (6.93)	.69*** (9.65)	.83*** (10.46)	-.03 (-.3)	-.01 (-.1)	.23* (1.81)	.6*** (6.34)	.63*** (8.61)	.76*** (10.07)
HML	.39*** (4.32)	.28*** (3.47)	.2** (2.17)	.54*** (5.58)	.47*** (4.16)	.09 (.84)	.37*** (4.36)	.27*** (3.73)	.21** (2.43)
MOM		-.25*** (-2.97)	-.26*** (-3.51)		-.17 (-1.41)	-.2** (-2)		-.23*** (-2.93)	-.24*** (-3.39)
RMW			.34*** (2.83)			.67*** (4.58)			.31*** (2.77)
CMA			-.13 (-.98)			.34** (2.05)			-.15 (-1.2)
\overline{LL}	84	84	84	84	84	84	84	84	84
Obs	294	294	294	294	294	294	294	294	294

Panel B: SIN TA 100									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Equal-weighted			Value-weighted			Sin-weighted		
α	.42** (2.00)	.56*** (2.66)	.44** (2.18)	.25 (.87)	.36 (1.33)	-.07 (-.25)	.55*** (2.72)	.70*** (3.42)	.58*** (3.16)
\overline{LL}	58	58	58	58	58	58	58	58	58
Obs	294	294	294	294	294	294	294	294	294

Panel C: SIN TA 200									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Equal-weighted			Value-weighted			Sin-weighted		
α	.50** (2.49)	.64*** (3.23)	.52*** (2.89)	.44 (1.57)	.56** (1.98)	.19 (.76)	.59*** (2.96)	.73*** (3.71)	.60*** (3.5)
\overline{LL}	43	43	43	43	43	43	43	43	43
Obs	294	294	294	294	294	294	294	294	294

Table A.4: Industry classification sin stock excess returns over risk-free rate

The table reports monthly excess returns of sin stocks identified with an industry classification approach over the risk-free rate. We report equal-, value-, and sin-weighted returns with respect to the [FF3](#), [CH4](#), and [FF6](#) factor models, respectively. Furthermore, we report the average number of sin stocks in the long leg (\overline{LL}) over our sample period from July 1997 to December 2021.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Equal-weighted			Value-weighted			Sin-weighted		
α	.65*** (3.05)	.82*** (3.71)	.63*** (3.14)	.42 (1.65)	.52** (2.07)	.09 (.39)	.75*** (3.5)	.91*** (4.17)	.73*** (4.07)
MKT	1.00*** (14.28)	.89*** (11.97)	.94*** (13.82)	.77*** (12.34)	.7*** (10.78)	.85*** (14.44)	1.00*** (14.16)	.89*** (11.59)	.94*** (12.9)
SMB	.55*** (4.47)	.58*** (5.72)	.77*** (6.17)	-.05 (-.51)	-.03 (-.31)	.2 (1.48)	.58*** (4.88)	.61*** (6.23)	.79*** (6.66)
HML	.46*** (4.37)	.33*** (3.52)	.22* (1.7)	.53*** (5.43)	.46*** (3.92)	.06 (.58)	.45*** (4.3)	.33*** (3.6)	.22* (1.71)
MOM		-.28** (-2.16)	-.29** (-2.44)		-.16 (-1.37)	-.2* (-1.94)		-.26** (-2.28)	-.28*** (-2.61)
RMW			.47*** (2.81)			.66*** (4.29)			.44*** (2.6)
CMA			-.16 (-.94)			.39** (2.28)			-.14 (-.85)
\overline{LL}	45	45	45	45	45	45	43	43	43
Obs	294	294	294	294	294	294	294	294	294

Table A.5: Textual analysis sin stock excess returns over comparables excluding covid period

The table reports monthly excess returns of sin stocks identified with our textual analysis approach over a portfolio of comparables. We report equal-, value-, and sin-weighted returns with respect to the [FF3](#), [CH4](#), and [FF6](#) factor models, respectively. Panel A reports the results of our baseline approach requiring a min. frequency ≥ 20 & unique ≥ 2 . Panels B and C report results with a min. frequency of 100 and a min. frequency of 200, respectively. Furthermore, we report the average number of sin stocks in the long leg (\overline{LL}) over our sample period from July 1997 to December 2019.

Panel A: SIN TA 20									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Equal-weighted			Value-weighted			Sin-weighted		
α	.01 (.08)	.02 (.16)	.03 (.31)	.17 (.92)	.23 (1.23)	.08 (.43)	.29** (2.17)	.29** (2.26)	.32** (2.47)
MKT	.08*** (3.89)	.08*** (3.26)	.07** (2.43)	.05 (.99)	.01 (.26)	.08 (1.36)	.06** (2.48)	.06** (2.07)	.05 (1.41)
SMB	.12*** (4.61)	.12*** (4.61)	.11*** (3.12)	.13 (1.31)	.15 (1.36)	.25* (1.75)	.08** (2.47)	.08** (2.41)	.06 (1.59)
HML	.1*** (3.68)	.09*** (3.21)	.11** (2.21)	.28*** (4.02)	.24*** (3.18)	.09 (1.18)	.09** (2.35)	.09** (2.21)	.12** (2.2)
MOM		-.01 (-.64)	-.01 (-.52)		-.09 (-1.02)	-.11 (-1.27)		0 (-.02)	0 (.1)
RMW			-.03 (-.72)			.3** (2.45)			-.05 (-.9)
CMA			-.01 (-.13)			.08 (.63)			-.04 (-.48)
\overline{LL}	86	86	86	86	86	86	86	86	86
\overline{SL}	234	234	234	234	234	234	234	234	234
Obs	270	270	270	270	270	270	270	270	270

Panel B: SIN TA 100									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Equal-weighted			Value-weighted			Sin-weighted		
α	.22 (1.52)	.22 (1.62)	.23 (1.62)	.22 (.99)	.29 (1.37)	.12 (.57)	.33** (2.21)	.34** (2.34)	.35** (2.4)
\overline{LL}	60	60	60	60	60	60	60	60	60
\overline{SL}	234	234	234	234	234	234	234	234	234
Obs	270	270	270	270	270	270	270	270	270

Panel C: SIN TA 200									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Equal-weighted			Value-weighted			Sin-weighted		
α	.31* (1.95)	.31** (2.03)	.32** (2.02)	.42* (1.89)	.49** (2.20)	.38* (1.75)	.38** (2.42)	.38** (2.52)	.38** (2.44)
\overline{LL}	44	44	44	44	44	44	44	44	44
\overline{SL}	234	234	234	234	234	234	234	234	234
Obs	270	270	270	270	270	270	270	270	270

Table A.6: Industry classification sin stock excess returns over comparables excluding covid period

The table reports monthly excess returns of sin stocks identified with an industry classification approach over a portfolio of comparables. We report equal-, value-, and sin-weighted returns with respect to the [FF3](#), [CH4](#), and [FF6](#) factor models, respectively. Furthermore, we report the average number of sin stocks in the long leg (\overline{LL}) over our sample period from July 1997 to December 2019.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Equal-weighted			Value-weighted			Sin-weighted		
α	.44*** (2.79)	.47*** -3	.42** (2.4)	.23 (1.2)	.28 (1.51)	.12 (.66)	.52*** (3.22)	.54*** (3.48)	.49*** (3.17)
MKT	.15*** (4.62)	.13*** (3.26)	.16*** (3.34)	.03 (.55)	-.01 (-.13)	.07 (1.09)	.14*** (4.4)	.13*** (3.08)	.15*** (3.32)
SMB	.02 (.27)	.02 (.35)	.07 (.7)	.12 (1.11)	.13 (1.15)	.23 (1.57)	.05 (.85)	.05 (.88)	.08 (.98)
HML	.17*** (2.83)	.15** (2.34)	.1 (1.15)	.28*** (3.85)	.24*** (2.9)	.07 (.81)	.17*** (3.11)	.15*** (2.7)	.1 (1.23)
MOM		-.05 (-.59)	-.05 (-.66)		-.09 (-.97)	-.11 (-1.22)		-.03 (-.56)	-.04 (-.63)
RMW			.12 (1.09)			.3** (2.36)			.09 (.9)
CMA			.01 (.06)			.13 (.91)			.03 (.24)
\overline{LL}	46	46	46	46	46	46	44	44	44
\overline{SL}	234	234	234	234	234	234	234	234	234
Obs	270	270	270	270	270	270	270	270	270

References

- Carhart, M. M., 1997. On persistence in mutual fund performance. *Journal of Finance* 52, 57–82.
- Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3–56.
- Fama, E. F., French, K. R., 2018. Choosing factors. *Journal of Financial Economics* 128, 234–252.