Estimating Firms' Responses to Securities Regulation Using a Bunching Approach

Finance Working Paper N° 867/2023
January 2023

Dhammika Dharmapala
University of Chicago, Australian National University, CESifo and Oxford University

ECGI Working Paper Series in Finance

Estimating Firms' Responses to Securities Regulation Using a Bunching Approach

Working Paper N° 867/2023
January 2023

Dhammika Dharmapala

I thank the Editor (JJ Prescott), two anonymous referees, Bernie Black, Ofer Eldar, Feng Gao, Michael Guttentag, Todd Henderson, Birgit Hüsecken, Prasad Krishnamurthy, John Landon-Lane, Christian Leuz, Kate Litvak, Anup Malani, Ben Marx, Roberta Romano, Holger Spamann, Juan Carlos Suarez-Serrato, David Weber, Laura Wellman, Eleanor Wilking, conference participants at the American Law and Economics Association annual meetings, the Conference on Empirical Legal Studies, and the International Institute of Public Finance annual Congress, and workshop participants at the University of California at Berkeley, the University of Chicago, Rutgers University, and the University of Virginia for helpful conversations and comments. I am especially grateful to Daniel Marcin (formerly of the Coase-Sandor Institute for Law and Economics at the University of Chicago Law School) for writing the Python code used to collect the data on public float, to Shreya Ram for research assistance, and to Miguel Almunia, Michael Carlos Best and Ben Marx for kindly sharing their Stata code for implementing the bunching analysis. I also acknowledge the financial support of the Lee and Brena Freeman Faculty Research Fund at the University of Chicago Law School. Any remaining errors or omissions are, of course, my own.

Abstract

Many important provisions of US securities law – most notably, crucial elements of the Sarbanes- Oxley (SOX) legislation enacted in 2002 – apply only to firms that have a public float of at least $75 million. Public float (i.e., the market value of shares held by non-insiders) is not comprehensively reported in standard databases, so I “scrape” public float data from firms’ 10-K filings for an extensive sample of reporting entities over fiscal years 1993-2015. I use a bunching approach that compares the number of observations immediately below the $75 million threshold to a smooth counterfactual density. Prior to SOX (i.e., over 1993-2002), there is no detectable bunching. Following SOX (i.e., over 2003-2015), there is statistically significant evidence of bunching. However, the magnitude of bunching is relatively modest. Moreover, bunching is concentrated in the early post-SOX years (2003-2009), and is virtually absent in later years (2010- 2015). The magnitude of bunching is not a sufficient statistic for the compliance costs of securities regulation because the costs of managing public float are unobservable. Nonetheless, the results of our bunching analysis cast some doubt on widespread claims that the regulatory burdens of these securities law provisions are large.

Keywords: Securities regulation; Sarbanes-Oxley; Public float; Compliance costs; Bunching analysis

JEL Classifications: G38; K22

Dhammika Dharmapala
Paul H. and Theo Leffmann Professor of Law
University of Chicago, The Law School
1111 E. 60th St.
Chicago IL 60637, USA
phone: +1 773 702 9494
e-mail: dharmap@uchicago.edu

Estimating Firms' Responses to Securities Regulation Using a Bunching Approach

Dhammika Dharmapala
University of Chicago Law School
dharmap@uchicago.edu

First draft: August 2016
Revised version: December 2022

Abstract

Many important provisions of US securities law – most notably, crucial elements of the Sarbanes-Oxley (SOX) legislation enacted in 2002 – apply only to firms that have a public float of at least $75 million. Public float (i.e., the market value of shares held by non-insiders) is not comprehensively reported in standard databases, so I "scrape" public float data from firms' 10-K filings for an extensive sample of reporting entities over fiscal years 1993-2015. I use a bunching approach that compares the number of observations immediately below the $75 million threshold to a smooth counterfactual density. Prior to SOX (i.e., over 1993-2002), there is no detectable bunching. Following SOX (i.e., over 2003-2015), there is statistically significant evidence of bunching. However, the magnitude of bunching is relatively modest. Moreover, bunching is concentrated in the early post-SOX years (2003-2009), and is virtually absent in later years (2010-2015). The magnitude of bunching is not a sufficient statistic for the compliance costs of securities regulation because the costs of managing public float are unobservable. Nonetheless, the results of our bunching analysis cast some doubt on widespread claims that the regulatory burdens of these securities law provisions are large.

JEL Classification: K22; G38

Keywords: Securities regulation; Sarbanes-Oxley; Public float; Compliance costs; Bunching analysis

Acknowledgments: I thank the Editor (JJ Prescott), two anonymous referees, Bernie Black, Ofer Eldar, Feng Gao, Michael Guttentag, Todd Henderson, Birgit Hüsecken, Prasad Krishnamurthy, John Landon-Lane, Christian Leuz, Kate Litvak, Anup Malani, Ben Marx, Roberta Romano, Holger Spamann, Juan Carlos Suarez-Serrato, David Weber, Laura Wellman, Eleanor Wilking, conference participants at the American Law and Economics Association annual meetings, the Conference on Empirical Legal Studies, and the International Institute of Public Finance annual Congress, and workshop participants at the University of California at Berkeley, the University of Chicago, Rutgers University, and the University of Virginia for helpful conversations and comments. I am especially grateful to Daniel Marcin (formerly of the Coase-Sandor Institute for Law and Economics at the University of Chicago Law School) for writing the Python code used to collect the data on public float, to Shreya Ram for research assistance, and to Miguel Almunia, Michael Carlos Best and Ben Marx for kindly sharing their Stata code for implementing the bunching analysis. I also acknowledge the financial support of the Lee and Brena Freeman Faculty Research Fund at the University of Chicago Law School. Any remaining errors or omissions are, of course, my own.

1) Introduction

An extensive literature across law, accounting, economics and finance has analyzed the costs and benefits of securities regulation. This paper brings both a novel dataset and a new empirical approach to bear on this important question. The empirical strategy exploits the fact that many provisions of US securities law apply only to firms at or above a threshold of $75 million of "public float" (i.e., the market value of shares held by non-insiders).[1] The most notable of these provisions is Section 404(b) of the Sarbanes-Oxley (hereafter, SOX) legislation. Enacted in 2002 in the wake of accounting scandals in the early 2000s, SOX introduced an array of new measures, including requirements of enhanced internal controls with respect to firms' financial disclosures. Section 404(b) requires that external auditors attest to the quality of these internal controls. Coates and Srinivasan (2014) provide a comprehensive assessment of the scholarly literature on SOX; they conclude that while substantial progress has been made in understanding the effects of SOX, research on its net social welfare consequences remains inconclusive.

Although it is a crucial concept for determining regulatory obligations, public float is not reported in standard financial databases, with some partial exceptions (discussed in Section 3 below). The previous literature has hand-collected public float data from firms' annual 10-K filings with the Securities Exchange Commission (SEC) for various subsamples of firms. This paper constructs a much larger dataset on public float by using Python code to "scrape" this information from firms' 10-K filings.[2] Using this method, I collect public float information for an extensive sample of reporting entities for fiscal years 1993-2015. This period spans the introduction of the $75 million threshold, the SOX legislation and several subsequent changes in its implementation.

This paper applies to this new dataset an empirical approach - not previously used in the study of securities regulation[3] - that draws on a growing literature in economics analyzing "bunching" around bright-line tax and regulatory thresholds.[4] In a setting characterized by a size-based threshold, a bunching analysis examines the divergence between the number of firms around the threshold and the counterfactual density (i.e., the number of firms that would be expected to

[1] Insiders (or "affiliates") typically include officers, directors and large blockholders.
[2] These filings are available through the SEC website at: http://www.sec.gov/edgar/searchedgar/companysearch.html
[3] Note, however, that some recent working papers that have appeared subsequent to earlier versions of this paper use variants of the bunching approach (e.g., Ewens, Xiao and Xu, 2021).
[4] See e.g., Saez, 2010; Chetty et al., 2011; Kleven and Waseem, 2013; Almunia and Lopez-Rodriguez, 2018; Marx, 2018. For a general overview, see Slemrod (2013), and for a theoretical analysis of such thresholds see Dharmapala, Slemrod and Wilson (2011).

be found around the threshold, absent the regulation). The latter is calculated by fitting a flexible polynomial function to the observed density of firms, excluding an interval close to the threshold. This approach assumes that the behavior of firms far from the threshold is not affected by its presence, and that the density of firms around the threshold would be smooth absent the threshold.

If securities regulation and enforcement make firms' financial disclosures more credible to investors, then firms may bunch just above $75 million. If the compliance costs of the regulation exceed the credibility (and other) benefits, then firms may bunch below the threshold.[5] Thus, bunching provides a direct source of evidence on firms' perceptions of the net value of regulations that is difficult to obtain using other types of data or empirical approaches. The estimated magnitude of bunching allows inferences to be made about the extent to which firms are willing to change their public float in response to securities regulation.

In 2002, the SEC allowed firms with public float below $75 million a later deadline to file their disclosures. More consequentially, in 2003, the SEC initially exempted these firms – known as "nonaccelerated filers" or NAFs - from SOX compliance. The SEC's timeline for phased-in SOX compliance for NAFs was repeatedly delayed until finally (in 2010) the Dodd-Frank legislation permanently exempted them. Thus, NAFs have been subject to more limited securities law obligations than "accelerated filers" or AFs (firms with public float of at least $75 million) throughout the period since the introduction of the threshold.[6]

As discussed in the accounting literature, firms have available a variety of mechanisms through which public float might potentially be "managed," such as insider purchases of stock, disclosures that reduce stock price, or increases in leverage (Gao, Wu and Zimmerman, 2009; Nondorf, Singer and You, 2012; Gao, 2016; Weber and Yang, 2020). Of course, these strategies entail various types of costs. A simple conceptual framework is discussed in Section 3 below in which firms' insiders trade off these costs against the net costs of securities regulation in determining what public float to report. Under certain assumptions, the magnitude of observed bunching below the threshold can be used to infer the amount by which firms reduce their reported public float. However, the magnitude of bunching is not a sufficient statistic for the net compliance cost of securities regulation. In particular, inferring this cost requires knowing the costs of

[5] There may be other explanations for bunching below the threshold, for instance the possibility that insiders may lose private benefits of control due to the regulations. These are discussed later in the paper (see in particular Section 5).
[6] The details of the regulations and their timeline are quite complex, and are explained in Section 2 below.

managing public float, which are not readily observable. Intuitively, this means that an absence of bunching may be attributable either to low compliance costs or to high costs of managing public float.

The empirical analysis in this paper aggregates observations of public float at the firm-year level into "bins" with a width of $1 million, and focuses on those bins ranging from $50 million to $150 million of public float. The counterfactual number of firms in bins around the $75 million threshold is estimated by fitting a flexible fifth-order polynomial function of public float to the observed density of firms, excluding an interval – $66-$83 million of public float – relatively close to the threshold.[7] Prior to the introduction of the threshold and the SOX legislation (1993-2002), there is no detectable divergence between the actual and counterfactual density around $75 million. After SOX (2003-2015), there is an excess number of firms below the $75 million threshold, relative to a smooth counterfactual density, that is statistically significant using bootstrapped standard errors. However, the magnitude of bunching is arguably relatively modest. The excess mass consists of 151 excess observations, relative to a counterfactual number of 1121 observations in the range $66-$75 million.[8] Importantly, this bunching behavior is almost entirely concentrated in the earlier years of the post-SOX era (2003-2009). Bunching virtually disappears during the latter part of the sample period (2010-2015).

An alternative estimation approach uses the observed pre-SOX (1993-2002) density away from the threshold to estimate the counterfactual density (rather than estimating the counterfactual density using data from the SOX period itself). This takes account of the possibility that SOX (or other factors that were different over 2003-2015) may have changed the density of firms in regions far away from the threshold. This alternative approach leads to very similar conclusions (while more satisfactorily reconciling the observed amount of bunching below the threshold with the observed missing mass above the threshold).

As previously noted, I cannot rule out the possibility that a modest magnitude of bunching is due to large costs of managing public float; limited bunching may, in principle, be consistent with large compliance costs. Even so, however, the results cast some doubt on widespread claims

[7] This excluded interval is determined using an iterative procedure described in Section 4 below.
[8] The estimated bunching parameter of 1.2 indicates that bunching firms reduce their public float by up to $1.2 million. As discussed in Section 5.1 below, the magnitude of bunching – while modest - is broadly consistent with estimates in the prior literature of other types of responses to SOX, such as the number of going-private transactions (e.g., Kamar, Karaca-Mandic and Talley, 2009).

that the regulatory burdens of these securities law provisions are (at least currently) large. This is especially the case because the accounting literature that focuses directly on firms' management of public float suggests that such management is not prohibitively costly.[9]

While the accounting literature finds evidence of firms manipulating public float in an apparent attempt to remain below the $75 million threshold, no previous study has used a formal bunching analysis to address this issue, and none has collected extensive data on public float.[10] The most closely related prior paper is Iliev (2010), which also uses the $75 million threshold as a central element of its research design. However, Iliev (2010) uses a quite different approach: a regression discontinuity (RD) design that compares the value of firms just above the $75 million threshold for the application of SOX Section 404 in 2004 with the value of firms just below this threshold. Using hand-collected public float data for a sample of 301 firms, Iliev (2010) notes that there is some apparent bunching below the threshold. For Iliev's (2010) RD design, any manipulation of public float is a source of potential bias. To address this problem, Iliev (2010) instruments for 2004 values of public float using 2002 values (which could not have been manipulated in response to the SOX threshold). In contrast, the empirical approach in this paper is premised on the ability of firms to manipulate public float.

[9] Gao, Wu and Zimmermann (2009) construct a treatment group of 806 firms that reported being NAFs in their 10-K filings with the SEC during the 2003-2005 period with a control group of 485 firms that reported being AFs and had market capitalization below $150 million over the same period. They find evidence that NAFs (when compared to the control group and to their own behavior in the pre-SOX period) engaged in various actions to remain below the $75 million threshold. These included reducing investment, increasing payout, and reducing the number of shares held by outsiders. They also find that NAFs disclose negative news and report lower accounting earnings in the second fiscal quarter (when public float is calculated). Nondorf, Singer and You (2012) analyze a sample of 257 firms that have public float around the $75 million threshold. They find evidence of changes in ownership by insiders, and of the use of earnings management in financial reporting during the second fiscal quarter, that are consistent with attempts to remain below the threshold. Gao (2016) explores in more detail how firms use their discretion in reporting public float, using a sample of 716 firms that had market capitalization between $75 million and $150 million over 2003-2006. Financial and ownership data for these firms is used to construct a benchmark public float, which is then compared to the reported public float (hand-collected from the SEC filings). Gao (2016) finds that firms that are predicted to face higher SOX 404 compliance costs count more shares as being "affiliated" in order to report a lower public float.

[10] There is an extensive accounting literature (beginning with Burgstahler and Dichev (1997)) that studies the density of reported earnings around relevant thresholds (such as zero or analysts' forecasts). However, this literature does not formally estimate a counterfactual density or derive an estimate of the magnitude of manipulation. Chhaochharia, Ott and Vig (2011) analyze the $75 million threshold, but their focus is on its impact on mergers and acquisitions that would result in the threshold being crossed. There is a large empirical literature on the impact of SOX across several academic disciplines (e.g., Litvak, 2007; Leuz, 2007; Zhang, 2007; Chhaochharia and Grinstein, 2007; Kamar, Karaca-Mandic and Talley, 2009); however, none of these prior contributions use a bunching approach.

This paper proceeds as follows. Section 2 provides some background on the $75 million threshold. Section 3 describes the data. Section 4 develops the conceptual framework and empirical strategy. Section 5 reports and discusses the results. Section 6 concludes.

2) The Role of the $75 Million Threshold in US Securities Law

The $75 million threshold and the concomitant division of issuers into AFs (with public float of at least $75 million in fiscal year 2002 or thereafter) and NAFs (with public float below $75 million) were introduced by the SEC in September 2002. The initial context was in relation to new rules relating to deadlines for filing 10-K and related forms with the SEC. AFs, as the term suggests, were subject to an earlier filing deadline than were NAFs.[11] Public float is defined as the "aggregate worldwide market value of the voting and non-voting common equity held by its non-affiliates . . . as of the last business day of the issuer's most recently completed second fiscal quarter . . ."[12] Affiliates generally include managers, directors, and large blockholders (typically those owning 10% or more of the firm's shares, though sometimes a 5% threshold is used).[13] Thus, public float reflects the value of the firm that is held by outside shareholders.[14]

While originally introduced in the context of filing deadlines, the $75 million threshold soon became an important factor in the implementation of SOX. Although SOX included a large number of provisions, the central elements for the purposes of this paper are contained in Section 404. In particular, Section 404(a) requires management to provide an annual assessment of the issuer's internal controls over financial reporting.[15] Section 404(b) requires that the registered public accounting firm that audits the company's financial statements must provide an attestation

[11] Firms with over $700 million of public float – known as "large accelerated filers" (LAFs) – were subject to even earlier deadlines. Because of the relatively small number of LAFs and because the SOX 404 treatment of AFs and LAFs is very similar, we do not emphasize the distinction between LAFs and AFs in this paper.
[12] 17 CFR 240.12b-2.
[13] An affiliate is defined as "a person that directly, or indirectly through one or more intermediaries, controls, or is controlled by, or is under common control with" the issuer (17 CFR 240.12b-2).
[14] Even if it meets the public float threshold, a firm is not defined as an AF until it has been required to file disclosures with the SEC for a year. Firms that have not satisfied this requirement and are close to the $75 million threshold will thus not face immediate regulatory consequences from crossing the threshold. This is unlikely to be of much importance in firms' choices, however, as such a firm would anticipate becoming subject to the relevant provisions in the near future.
[15] The SEC rules state that: "The internal control report must include: a statement of management's responsibility for establishing and maintaining adequate internal control over financial reporting for the company; management's assessment of the effectiveness of the company's internal control over financial reporting as of the end of the company's most recent fiscal year; a statement identifying the framework used by management to evaluate the effectiveness of the company's internal control over financial reporting" - see http://www.sec.gov/rules/final/33-8238.htm#ia

of the firm's internal controls over financial reporting. The latter is generally thought to be the most costly SOX provision, especially for small firms, entailing additional audit fees as well as other types of costs (e.g., Alexander et al., 2013). It is thus on Section 404(b) that most of the empirical literature on SOX has focused.

The SOX legislation did not include an exemption for smaller firms, and was initially expected to apply to all issuers on the US market. However, on June 5, 2003, the SEC announced a phased-in implementation schedule under which smaller firms would be permitted a longer period in which to move towards compliance with Section 404. Specifically, AFs were required to comply with SOX Section 404 starting in 2004. NAFs were granted a temporary delay in the date of expected compliance, initially until 2005.[16] In 2005, the SEC extended the exemption from Section 404 for NAFs by an additional year. This extension was repeated in 2006.[17] In 2007, all firms became subject to SOX Section 404(a).[18] However, NAFs continued to receive temporary exemption from 404(b).[19] The temporary exemption from Section 404(b) for NAFs was further extended in 2008 and 2009. In 2010, Congress enacted the Dodd-Frank Act, which (among other things) made the exemption of NAFs from SOX 404(b) permanent.

The history of SOX 404 implementation is thus rather complex. For the purposes of this paper, the crucial point is that the $75 million threshold on which I focus has continued to be important throughout the 2003-2015 period. In particular, NAFs have been exempt (either on an ostensibly temporary basis or permanently) from SOX 404(b) throughout the period that SOX has been in operation. NAFs have also been subject to later filing deadlines over this period. It is important to note that the timing of the various changes in SEC rules does not permit us to distinguish between the impact of Section 404 and that of the filing deadlines.

In addition, "smaller reporting companies" (SRCs) were defined similarly to NAFs in terms of having public float below $75 million during our sample period.[20] SRCs have been eligible for scaled disclosure requirements since 2008. In 2018, the definition of SRCs was amended to include

[16] SEC release 33-8238, available at: http://www.sec.gov/rules/final/33-8238.htm
[17] See e.g., Table 1 in Gao, Wu and Zimmerman (2009) and Figure 1 in Ge, Koester and McVay (2017) for detailed timelines.
[18] Thus, until 2007 firms that became AFs thereby became subject to both Sections 404(a) and 404(b). Consequently, it is difficult during the period until 2007 to distinguish between firms that reduced their public float in order to avoid Section 404(b) and those that did so to avoid Section 404(a) (or were seeking to avoid both provisions).
[19] Also in 2007, the Public Company Accounting Oversight Board (PCAOB) introduced Accounting Standard 5 (AS5), with the apparent intention of lowering the compliance costs of Section 404(b) (Gao, 2016).
[20] If public float could not be calculated, a firm could also qualify as a smaller reporting company if its revenue was below $50 million

all firms with public float below $250 million, along with entities that had annual revenue of less than $100 million in the previous year and either no public float or a public float of less than $700 million.[21] These changes became effective on September 10, 2018 and thus do not affect our sample. However, over the period from 2008-2015, it is difficult to distinguish between the effects of SOX 404(b), accelerated filing deadlines, and scaled disclosure. Any observed bunching is assumed to be in response to some combination of these regulatory provisions.

As public float is calculated as of the last business day of the second fiscal quarter (typically, the last business day in June), 2003 was the first year in which firms could have adjusted their public float in response to SOX or to the accelerated filing deadlines (Iliev, 2010). I thus divide the sample into a period prior to the introduction of the threshold (1993-2002, also referred to as the pre-SOX period) during which there was no incentive for bunching, and a period after the introduction of the threshold (2003-2015, also referred to as the SOX period), during which firms in the region of the threshold may or may not have used the various strategies discussed in Section 4 below to manage their public float.

There are certain categories of firms that do not face any incentives to remain below the threshold. Most importantly, once a firm becomes an AF as a result of crossing the $75 million threshold, it remains one in the future unless its public float falls below $50 million.[22] This implies that firms with public float below $75 million that had become AFs in the past have no incentive to remain below $75 million to avoid regulatory burdens (which they would face in any event). As described below in Section 3, the analysis seeks to take account of this asymmetry in a number of ways.

Certain other categories of firms may also be indifferent to the threshold. AFs that are foreign private issuers (FPIs, a category of foreign companies defined by the SEC on the basis of the degree of US share ownership and business contacts) were temporarily exempt from SOX 404(b) until 2007. Some firms not subject to SOX 404(b) voluntarily complied with its provisions (e.g., Ge, Koester, and McVay, 2017). The Jumpstart Our Business Startups (JOBS) Act, enacted on April 5, 2012, relaxed disclosure and compliance obligations for a new category of firms, known as "emerging growth companies" (EGCs), defined on the basis of revenue (and certain

[21] SEC release 33-10513, available at: https://www.sec.gov/rules/final/2018/33-10513.pdf

[22] This rule has applied since 2005; a stricter rule, requiring public float to fall below $25 million and certain other conditions to be met, applied before 2005.

other characteristics) rather than public float, for a period of 5 years from their initial issuance of securities (e.g., Dharmapala and Khanna, 2016). Among other things, EGCs were permitted an exemption from SOX 404(b) for that 5-year period, even if they are AFs on the basis of their public float.

Until 2007, firms wishing to undertake seasoned equity offerings (SEOs) were permitted to use a simplified form (Form S-3) only if they had public float exceeding $75 million (e.g., Gao, 2016). This created an incentive for firms undertaking SEOs to manage public float to exceed the $75 million threshold, and so may lead to the magnitude of bunching observed below $75 million being smaller than might otherwise be the case. However, SEOs are relatively uncommon. In addition, the SEC extended eligibility for the simplified form to firms with public float below $75 million in 2007, subject to certain conditions. Thus, this issue does not affect the majority of the post-SOX years, and our estimates for the 2010-2015 period are unaffected.

3) Data

3.1) Data on Public Float

As is evident from the account in Section 2, public float is a crucial concept in determining firms' securities law obligations. Nonetheless, research using public float has been limited because it is not reported in standard archival financial databases such as Compustat and CRSP, with some partial exceptions. The Audit Analytics database reports firms' accelerated filer status, and has been used quite extensively in the accounting literature (e.g., Randhawa, 2009; Kinney and Shepardson, 2011; Jia, Xie and Ziebart, 2014; Ge, Koester and McVay, 2017). The Datastream database provided by Thomson's Financial reports firms' "free float" from 2002 (e.g., Gao, 2010). However, this is only available for a limited sample of firms, has very limited pre-SOX coverage, and uses a somewhat different definition than the SEC's notion of public float. Importantly, the SEC has recently made available "Structured" Financial Statements datasets.[23] These provide machine-readable data on public float, but only from 2009. This data is used in Section 3.2 below to seek to validate this paper's dataset over the overlapping sample period (2009-2015).

The previous literature has hand-collected the value of public float from firms' annual 10-K filings with the SEC for various subsamples of firms (e.g., Gao, Wu and Zimmerman, 2009: Iliev, 2010; Dharmapala and Khanna, 2016; Weber and Yang, 2020). This paper constructs a much

[23] See https://www.sec.gov/dera/data/financial-statement-data-sets.html

larger dataset on public float by using Python code to "scrape" information on public float from firms' 10-K filings.[24] The filings are accessed through the SEC's Electronic Data Gathering and Retrieval (EDGAR) system.[25] The data-gathering process is facilitated by the relatively uniform nature of public float reporting. On the first page of the 10-K form, a reporting entity states its public float (determined as of the last day of the second fiscal quarter), as well as checking one of a number of checkboxes specifying its filing status. A typical example of the language used when reporting public float is the following:[26]

> "The aggregate market value of common stock held by non-affiliates of the registrant based on the closing price of the registrant's common stock as reported on the NASDAQ Global Market on June 28, 2013, was $234,272,491. Shares of voting and non-voting stock held by executive officers, directors and holders of more than 5% of the outstanding stock have been excluded from this calculation because such persons or institutions may be deemed affiliates. This determination of affiliate status is not a conclusive determination for other purposes."

The Python code automatically extracts the number that follows the phrase: "The aggregate market value . . . was" – in this instance, $234,272,491.

Using this method, I collect public float information for a large sample of reporting entities for fiscal years 1993-2015. This period spans the introduction of the threshold, the SOX legislation, and the various changes in its implementation that were outlined in Section 2. The full dataset contains 160,988 observations on public float at the firm-year level on 23,719 distinct reporting entities over 1993-2015. Reporting entities are identified by the Central Index Key (CIK) number that is assigned by the SEC. While this dataset is very large in comparison to those used in prior studies, it is important to bear in mind that it is relatively small in relation to the size that is ideal for the implementation of bunching analysis. This sample size limitation affects aspects of the analysis, as noted below.

The $75 million threshold described in Section 2 exceeds the public float reported for the majority of observations in this dataset. Even excluding observations with public float below $1 million (which typically involve small entities that become subject to SEC reporting requirements by issuing debt securities), the threshold falls just below the median reported public float (about $80 million). Firms close to the threshold are thus fairly representative of the population of

---

[24] I am grateful to Daniel Marcin (formerly of the Coase-Sandor Institute for Law and Economics at the University of Chicago Law School) for writing this code.
[25] See http://www.sec.gov/edgar/searchedgar/companysearch.html
[26] This example is from the 2013 10-K form for Brightcove, Inc.

reporting entities. However, the threshold is far below the mean level of public float (about $60 billion) due to a skewed distribution in which a small number of entities have very large public float. This skewed distribution does not affect the results, however, as the analysis focuses only on observations with public float in the interval of $50 million to $150 million (see below).

In addition, four checkboxes are listed on the first page of the 10-K form to indicate the issuer's filing status. A firm must report being an accelerated filer, a large accelerated filer, a nonaccelerated filer, or a smaller reporting company (SRC). The automated "scraping" process also collects this data on firms' filing status from the checkboxes. However, the coverage of the checkbox data is substantially more limited than that of the numerical public float data.[27] There is nonetheless some evidence for the validity of the checkbox data that I are able to collect. For instance, for virtually all (specifically, 99.75% of) observations in which a firm reports being an AF in the checkbox data, I are able to verify that its current public float, and/or its public float in a prior year, exceeds $75 million.

The number of reporting entities is large in relation to the number of publicly-traded companies on the US stock market, as it includes, for instance, very small entities that become subject to SEC reporting requirements by issuing debt securities.[28] However, as our interest is in the behavior of firms that are relatively close to the $75 million threshold, I focus only on issuers with public float in the range of $50-$150 million (which are more likely to represent the conventional publicly-traded companies that have been the focus of discussion in the literature). There are 21,848 firm-year observations in this range.[29] Of these observations, 10,247 observations are from the pre-SOX period (1993-2002) and 11,601 from the SOX period (2003-2015).

As noted in Section 2, once a firm passes the $75 million threshold, it remains an AF unless its public float subsequently falls below $50 million. Thus, an AF will retain this status absent a quite large future reduction in public float. Firms that have previously crossed the threshold and become AFs will thus in most circumstances be indifferent to the threshold in their future decisions about reporting public float. For instance, consider a firm that is already an AF (because it passed

[27] For instance, there are 88,793 observations on filing status from the checkbox data, compared to nearly 161,000 observations on public float.

[28] However, the dataset does not include registered investment companies regulated by the Investment Company Act of 1940, for which SOX rules are not relevant, or other entities that do not report public float to the SEC.

[29] In most cases, the scraping procedure obtains one value of public float for each firm-year. However, firms in some situations file amended 10-K forms, and the automated data collection process collects data from both the original and amended forms. The dataset has been cleaned to ensure that there is only one observation per firm-year, using the latest filing by a firm for a given filing year.

the threshold in the past) and currently has public float of $70 million. If it can freely choose next year's public float anywhere in the range $73 million to $77 million, it should be equally likely to locate just below or just above the threshold.

Thus, taken by itself, the existence of firms that are indifferent to the threshold because they are already AFs would, in expectation, affect the observed excess mass and missing mass symmetrically. However, the discussion below focuses particularly on the former. Moreover, AFs with public float below $75 million are likely to locate disproportionately close to the threshold (relative to NAFs), as public float tends to exhibit persistence over time. For instance, suppose that the hypothetical firm in our example reported $76 million of public float in the previous year. It would be expected to be substantially more likely to report $70 million this year than to report $60 million this year. AFs with public float below $75 million are thus likely to affect the measured excess mass, while having little impact on the estimated counterfactual density.

There appears to be no straightforward solution to this issue. The standard bunching approach presumes that all firms are potentially sensitive to the threshold. However, any AFs that are found in the bunching region cannot be viewed as being bunchers or as part of the pool of potential bunchers. Thus, the baseline analysis below excludes firm-years involving AFs reporting public float below $75 million. In particular, I use firms' responses to the checkboxes in the 10-K filing that identify their filing status (specifically, whether they report being an AF) to identify firm-year observations in year t in which the firm had public float below $75 million in year ($t$ – 1) but nonetheless reported being an AF in year (t – 1).[30] Using this approach, I exclude from the analysis 1508 firm-years for which, in the prior year, the firm reported public float below $75 million but either also reported being an AF or had ever reported a public float of $75 million or more in 2002 or any subsequent year. This leaves 10,093 firm-year observations in the post-SOX (2003-2015) period.

While this exclusion is used for the baseline results, it is not a complete solution and may itself generate potential bias in certain circumstances. Thus, I also verify that the results and patterns of bunching are similar when AFs with public float below $75 million are included in the analysis (see Figure 9 and the discussion in Section 5.5 below).

[30] As the coverage of the checkbox data is limited relative to the numerical public float data, we also use data on past values of public float to infer whether firms became AFs in the past. In particular, we compute the maximum past value of public float reported by a firm since fiscal year 2002, and use this to infer which firms are likely to already be AFs (even when their checkbox data is missing).

Descriptive statistics for public float are provided in Table 1 for issuers with public float in the range of $50-$150 million in the pre-SOX and SOX periods. It is noteworthy that the mean and dispersion of public float across the pre-SOX and SOX periods are quite similar.

3.2) Validating the Data on Public Float

An important question relates to the extent and representatives of the coverage of the scraped data. While the aim is to encompass the universe of public float reports, achieving this is not straightforward. Moreover, in the absence of a comprehensive data source on public float – which is precisely what necessitates the web scraping approach - it is not possible to fully answer this question. However, a validation exercise that provides some reassurance on this issue is made possible by the availability of the SEC's structured dataset (introduced in Section 3.1 above) that covers years from 2009. For this validation exercise, I collect data on public float for the period 2009-2020 from the SEC's Structured Financial Statements database, and focus particularly on the period over which the two datasets overlap (2009-2015).

We begin with the observations in the SEC structured dataset (identified by the CIK number of the reporting entity and the fiscal year). I then determine how many of those observations can be matched (by CIK number and fiscal year) to observations found in the scraped data. For the years 2009-2014, 93% of the firm-year observations in the SEC structured data are also found in the scraped data. Of these, 95% are exact matches (i.e., reported public float is identical across the two datasets).[31] Where differences exist, they tend to be relatively small: the median difference for observations that are not exact matches is about $37,000. The scraped data for 2009-2014 thus appears to be substantially, though, not entirely, complete.[32] For 2015, the scraped data has a lower extent of coverage. However, the results (described in Section 5 below) are virtually identical when I exclude the 2015 data from the analysis. Thus, the baseline analysis includes the 2015 data, despite its more limited coverage, in order to maximize sample size. However, it should be emphasized that none of the paper's claims hinge on the inclusion of the 2015 data.

---

[31] In addition, there is a similarly high rate of exact matches (93%) when comparing the scraped data to the data on public float in fiscal year 2012 hand-collected for a subset of firms by Dharmapala and Khanna (2016).

[32] For the relatively small number of observations where the data sources differ, it would be possible to use the SEC structured data's report of public float in the analysis (instead of using the scraped value of public float). We do not do this, in order to maintain comparability of the scraped data across different years; however, doing so leads to virtually identical results.

Table 2 underscores the similarity of the two datasets. It reports descriptive statistics for the dataset used in this paper and the SEC structured dataset for the years in which the coverage of the two datasets overlap (2009-2015). The comparison is restricted to observations with public float that falls within the range of $50 million to $150 million. While discrepancies between the two datasets are uncommon, this range is defined using public float as reported in this paper's dataset (and hence corresponds to the range of observations used in the baseline analysis below). As shown in Table 2, the mean and standard deviation of the SEC structured data are very similar to those of the scraped data. Note also that the SEC structured data provides an opportunity to validate the bunching results using the scraped data for the later years of the sample – see the discussion in Section 5.5 below in relation to Figure 10.

A potential concern is that the coverage of the scraped data might be particularly sparse at around the time that the threshold was introduced and SOX was enacted (for instance, because the language used in the disclosures may have been less standardized at that time). This may potentially create challenges for the empirical approach used below. Table 3 reports the number of public float observations in the scraped dataset that fall within the range of $50 million to $150 million for each year in a five-period (2001-2005) around the enactment of SOX. The coverage of the scraped data is fairly stable (at around 1000 observations per year) throughout this period. Indeed, there is a slight increase – rather than a decrease - in coverage in 2003 (the first year in which SOX applied) relative to the prior year. Thus, it does not appear that there were substantial changes in the coverage of the scraped data around the enactment of SOX that may bias the results.

4) Conceptual Framework and Empirical Strategy

4.1) Conceptual Framework

We apply to the setting described in Section 2 an empirical technique that draws on a growing literature in economics analyzing bunching around tax and regulatory thresholds. Bunching analysis was originally applied to the study of responses to taxation (e.g., Saez, 2010; Chetty et al., 2011). However, it has increasingly been applied to the analysis of the consequences of regulation, exploiting bright-line thresholds – often based on firm size – that are commonly used to determine the applicability of various types of regulations.[33] In the terminology of Slemrod (2013), these types of thresholds represent "notches" at which there is a discrete jump in regulatory obligations.

It is important to emphasize that what these studies aim to do is not to highlight distortions arising from the particular choice of a bright-line threshold, but rather to use the existence of this threshold as a source of quasi-experimental variation to draw wider inferences about the costs and benefits of the regulation. Note, however, that these wider inferences may be limited by the specific threshold that is chosen; for instance, it is not in general possible to rule out the possibility that the bunching behavior observed around a $75 million threshold may have been different if the threshold had been set at $500 million. Nonetheless, the main conclusion in Section 5 below is that the magnitude of bunching is fairly modest, and (if compliance costs are largely fixed) it is likely that a higher threshold would tend to reinforce this result. Note also that the bunching literature does not take a stand on whether bright-line thresholds are optimal as a normative matter. In theory, it is possible that in the presence of administrative costs, an optimally-designed tax or regulatory system may include bright-line thresholds, notwithstanding the distortionary behavior that they induce immediately around the threshold (for a formal model, see Dharmapala, Slemrod and Wilson (2011)).

Consider a firm with exogenous fundamental value V that is controlled by an insider who owns a fraction (1 - α) of its stock (where α is the fraction of the firm owned by outsiders). Public float (denoted PF) can then be expressed as: $PF = \alpha V$. Assume that the insider has a certain amount of discretion in the choice of α (for instance, by buying or selling more or less of the firm's stock).[34] This choice potentially affects firm value through effects on the firm's governance, the incentives of the insider to monitor the firm's operations, or the firm's access to external capital. Absent the regulatory threshold introduced below, there is assumed to be an optimal choice of α, denoted α^*. Deviating from this choice entails costs to the insider (for instance, of increased risk-bearing due to reduced portfolio diversification when inside ownership is high) and to all

[33] For instance, Almunia and Lopez-Rodriguez (2018) analyze the responses of Spanish firms to a threshold at €6 million of revenue above which they are subjected to greater scrutiny by the tax authorities. Marx (2018) studies the responses of US charities to an income threshold above which they face more onerous reporting requirements in order to establish their tax-exempt status.

[34] In reality, there may be other mechanisms to achieve this aim, such as disclosures that reduce stock price or reclassification of the insider status of blockholders (Gao, Wu and Zimmerman, 2009; Nondorf, Singer and You, 2012; Gao, 2016).

shareholders (for instance, the value of the firm will be lower with more limited access to external capital).

Now, suppose that a regulatory threshold $\overline{PF}$ is introduced, where firms with $PF \geq \overline{PF}$ are subject to an additional set of regulatory obligations with a (fixed) net compliance cost *C*. For instance, SOX 404(b) entails higher audit fees, the potential diversion of managerial effort into compliance activity, liability risk, and other possible costs. On the other hand, internal control attestation may make firms' financial disclosures more credible to investors, and firms not subject to the regulation may not be able to replicate the same level of credibility through voluntary mechanisms (for instance, because sanctions are less severe). *C* represents the excess of the costs over the benefits.

Let $\bar{V} = \overline{PF}/\alpha^*$ be the fundamental value of a firm that just has a sufficiently large (undistorted) public float to satisfy the regulatory threshold. Firms with $V < \bar{V}$ are unaffected by the introduction of the threshold, and the insiders of these firms will choose α^* and the same *PF* as in the absence of the threshold. For a firm with *V* at or slightly above $\bar{V}$, for a sufficiently large *C* and a sufficiently small cost of managing public float, the insider will choose to reduce outside investors' stake to a value that is just low enough to reduce *PF* below $\overline{PF}$. This behavior will generate bunching, as some firms with $V \geq \bar{V}$ report public float $PF < \overline{PF}$. If the costs of reducing public float below $\overline{PF}$ are increasing in $(V - \overline{PF})$, there will be some fundamental value $V_m > \bar{V}$ at which the insider will be indifferent between bunching and choosing the undistorted ownership structure α^*.[35]

In this simple framework, all firms with fundamental value from $\bar{V}$ to V_m will bunch – i.e., will reduce their outside ownership to the (minimal) extent required to remain below the threshold. These firms will report public float below $\overline{PF}$. There will be an excess mass of firms just below the threshold $\overline{PF}$, as those firms that would counterfactually locate in this region are joined by the bunching firms. Conversely, there will be a missing region above the threshold, with no firms reporting public float in the range $\overline{PF}$ to PF_m. From PF_m onwards, firms will report undistorted values of their public float. Thus, firms with fundamental value $V > V_m$ will not change their ownership structure due to the regulatory threshold, although they will of course bear the net compliance cost *C*.

[35] For a formal characterization of this bunching behavior, see the theoretical model developed in an earlier version of this paper (Dharmapala, 2016).

This pattern of outcomes is illustrated in Figure 1. Here, $\Delta PF = PF_m - \overline{PF}$ represents the amount (in millions of dollars) by which the marginal buncher (the firm whose insider is indifferent between bunching and not bunching) reduces its public float in response to the threshold. Let $f_0(PF)$ be the counterfactual density of firms defined over the possible values of public float. Assuming that $B = H$, the magnitude of the excess mass (i.e., the area of B in Figure 1) can in general be defined as:

$$B = \int_{\overline{PF}}^{PF_m} f_0(PF)dPF \tag{1}$$

Assuming that $B \approx H$ and that $f_0(PF)$ is approximately flat in the region around the \$75 million threshold (as shown in Figure 1), it is possible to use the following approximation for the magnitude of the excess mass:

$$B \approx H \approx \Delta PF * f_0(75) \tag{2}$$

where $f_0(75)$ is the counterfactual density evaluated at the threshold. Rearranging Equation (2) results in a simple approximation for ΔPF:

$$\Delta PF \approx \frac{B}{f_0(75)} \equiv \hat{b} \tag{3}$$

Here, $\hat{b}$ is the "bunching ratio" that is commonly estimated in bunching studies: i.e., the ratio of the excess mass (B) to the height of the counterfactual density at the threshold (i.e., $f_0(75)$).

There are a number of important caveats to be noted. In common with most applications of the bunching approach, there is not a completely missing region above the threshold. Rather, there is only a partially missing region, with many firms being insensitive to the threshold for various reasons. I do not impose the restriction that $B = H$ in the basic analysis below because it is possible that firms just above the threshold may go private and therefore not appear in the data (however, one of our specifications - using pre-SOX data to compute the counterfactual density - generates approximately equal values of B and H; see Section 5.3 below). In addition, the estimated counterfactual density is declining rather than flat. Thus, it should be borne in mind that the analysis described here relies on a number of simplifications; nonetheless, the approximation given above by $\hat{b}$ remains valuable in interpreting the magnitude of the observed bunching behavior.

Unfortunately, however, $\hat{b}$ is not a sufficient statistic for the net compliance cost C. In particular, it can be shown that C depends not only on ΔPF (and hence on $\hat{b}$), but also on other (unobservable) terms, such as the various costs of managing public float.[36] The net compliance cost *C* thus cannot be directly inferred from the observed magnitude of bunching. Intuitively, this implies, for instance, that an absence of bunching may be attributable either to low compliance costs or to high costs of managing public float (even when compliance costs are large).[37]

4.2) Empirical Strategy

The empirical analysis involves aggregating the firm-year-level data described in Section 3.1 to a bin-level dataset that measures the number of firms in each bin of public float (i.e., firm density). As entities with zero or very low values of public float and larger firms with public float that substantially exceeds the threshold are unlikely to be of much relevance, I restrict attention to values of public float in the range of $50 million to $150 million. I start at $50 million because the density appears substantially different below this level. The range is extended to $150 million (rather than the more natural $50-$100 million) in order to provide a sufficient number of bin-level observations to meaningfully estimate the counterfactual density using a fifth-order polynomial; in any event, the density over $100-$150 million does not appear substantially different from that for the preceding bins.

[36] For a formal derivation, see Dharmapala (2016). Note that "net compliance costs" should be understood fairly broadly in this context. For example, if managing public float to remain below the threshold sends a negative signal to investors about the credibility of the firm's disclosures, then firms may refrain from bunching even when compliance costs are quite large. Note also that the discussion in this section assumes that firms can choose their public float deterministically. In practice, while firms can adjust insider ownership and other relevant variables, the ultimate impact on public float will be mediated by stock price (and general market) volatility. Thus, public float management may be stochastic in nature. Under certain assumptions (including firms being risk-neutral with respect to the costs and benefits of crossing the threshold, and stock price volatility being symmetric on the upside and downside), the stochastic character of public float will not necessarily affect firms' choices of the public float for which to aim. However, the potential impact on the observed amount of bunching is not straightforward. For example, upside volatility may result in some bunchers who seek to remain just below the threshold having realized public float above the threshold; on the other hand, downside volatility may result in some nonbunchers who aim for a public float above the threshold having realized public float in the bunching region. In general, the expected return from any action taken in an attempt to keep public float below the threshold will be lower in a stochastic scenario than in one where public float is deterministic. This could, however, be conceptualized as part of the cost of managing public float. Thus, it does not fundamentally alter the point that bunching results from trading off the net costs of regulatory compliance against the costs of managing public float (and that in the absence of direct information on the latter, the observed amount of bunching is not a sufficient statistic for the former). A further complication is that (prior to the Dodd-Frank legislation) firms faced regulatory uncertainty about whether SOX exemptions would be extended in the future. This, too, would reduce the expected return from actions taken to limit the growth of public float. However, it is of less relevance for the later years of the sample period.

[37] A recent working paper that has appeared subsequent to earlier versions of this paper adopts a "fuzzy bunching" approach and seeks to develop a more concrete characterization of compliance costs (Ewens, Xiao and Xu, 2021). The latter exercise, however, requires strong assumptions; Ewens, Xiao and Xu (2021) assume that public float is managed entirely by increasing debt, and impose a specific functional form to characterize the costs to firms of modifying their capital structure.

We divide the \$50-\$150 million range into bins with a uniform width of \$1 million for the formal analysis. For instance, "bin 75" includes observations of public float that range from \$74 million up to (but not including) \$75 million. This definition ensures that no bin crosses the regulatory threshold. This bin width is narrower than would be ideal, creating a substantial degree of noise in the observed density. The histograms shown below thus use wider bin widths of \$5 million for greater visual clarity (histograms using \$1 million bins show very similar, although somewhat noisier, patterns). However, the use of \$1 million bins allows for a sufficient number of bin-level observations to implement the analysis, while the noise may be expected to be taken account of in the standard errors.

Observations of public float are pooled across years within the pre-SOX period to calculate a variable N_i^{PreSOX} (the number of firm-year-level observations of public float that are found within bin *i* over the pre-SOX period (1993-2002)). Similarly, observations of public float are pooled across years within the SOX period to calculate a variable N_i^{SOX} (the number of firm-year-level observations of public float that are found within bin *i* over the SOX period (2003-2015)).

Figure 2 is a histogram representing the number of firm-year observations of public float during the pre-SOX period using bins of width \$5 million. In general, the density declines as public float rises – i.e., there are fewer firm-years with larger values of public float, even in our restricted range. However, \$75 million of public float (indicated by the vertical red line) does not appear to be associated with any particular departure from the general pattern. In contrast, Figure 3 - a histogram representing the number of firm-year observations of public float during the SOX period, also using \$5 million bins - suggests some visual evidence of bunching below the \$75 million threshold.

Our formal methodology is based on the estimation of a counterfactual density and a bunching parameter $\hat{b}$ (using bins of \$1 million width to provide a sufficient number of bin-level observations), and on obtaining standard errors for $\hat{b}$ through bootstrapping. The regression specification used to estimate the counterfactual density can be represented as follows:

$$N_i^{SOX} = \alpha + \sum_{j=1}^{5}[\beta_j (Mid_i)^j] + \gamma_i I_i + \epsilon_i \tag{4}$$

Here, *i* indexes bins (where, for example bin 55 is the bin that includes public float observations that fall in the range from \$54 million to just under \$55 million). N_i^{SOX} is the number of firm-year-level observations of public float that are found within bin i over the SOX period (for instance, N_{55}^{SOX} represents the number of observations of public float that fall in the range from \$54 million to just under \$55 million over the SOX period). Mid_i is the midpoint of the range of public floats included within bin i – for instance, for bin 55 (i.e., i = 55), Mid_{55} = 54.5 (where public float is measured in millions of US \$). The summation over the values of j represents a flexible fifth-order polynomial that is used to estimate the counterfactual density.

The estimation of the counterfactual density excludes an interval around the \$75 million threshold. This excluded interval is denoted by [L, U], where L and U are the bins that represent the lower and upper bounds of the excluded interval. Its exclusion is accomplished in Equation (4) by adding a series of indicator variables I_i = 1 if $i \in$ [L, U] and 0 otherwise, that remove the influence of the bins in the excluded interval from the estimation of the counterfactual density. In choosing the excluded interval [L, U], I begin with the visual inspection of the bin-level data to determine reasonable lower and upper bounds, and then follow an iterative process of including indicators for the bins in and around that interval until I exhaust the bins that exhibit a significantly different observed density relative to the counterfactual density. Based on this iterative process, I use an excluded interval ranging from \$66-\$83 million.[38]

The counterfactual density generated by Equation (4) is represented in Figure 4, along with the number of public float observations in each bin of width \$1 million. Although the density exhibits some noise within these narrower bins, the counterfactual density captures the actual density quite closely outside the excluded interval. Within the excluded interval, there is (as in Figure 3) some evidence of bunching below, and missing mass above, the threshold.

Let the predicted values from Equation (4) – i.e., the estimated counterfactual density – be denoted by $\widehat{N}_i^{SOX}$; then, the excess mass B can be estimated as follows:

$$\widehat{B} = \sum_{i=L}^{75} (N_i^{SOX} - \widehat{N}_i^{SOX}) \tag{5}$$

The magnitude of the missing mass H can be estimated as follows:

[38] Kleven and Waseem (2013) develop an iterative procedure that determines the upper limit of the excluded region by a process of repeatedly increasing this limit until the area of H converges to that of B (see also Almunia and Lopez-Rodriguez, 2018). This imposes the assumption that firms that find themselves just above the threshold do not exit the dataset in substantial numbers. We do not impose the assumption that B = H in our baseline analysis because, for instance, it is possible that firms that would counterfactually be just above the threshold may go private and therefore not appear in the data. However, one of our specifications - using pre-SOX data to compute the counterfactual density - generates approximately equal values of B and H (as discussed in Section 5.3 below).

$$\widehat{H} = \sum_{i=76}^{U} (\widehat{N}_i^{SOX} - N_i^{SOX}) \tag{6}$$

The height of the counterfactual density $f_0(75)$ can be estimated by averaging the values of $\widehat{N}_i^{SOX}$ over the 9 bins (representing public float of \$66-\$75 million) within the excluded region and under the threshold:

$$\hat{f}_0(75) = \frac{\sum_{i=L}^{75} \widehat{N}_i^{SOX}}{9} \tag{7}$$

This approach smooths the estimated height of the counterfactual density over the bunching region, and thus takes account of the fact that the counterfactual density in this region is not flat but rather is declining (and so is not necessarily well approximated by simply evaluating f_0 at the threshold). Finally, the bunching ratio defined in Equation (3) can be estimated as:

$$\hat{b} = \frac{\widehat{B}}{\hat{f}_0(75)} \tag{8}$$

5) Results and Discussion

5.1) Basic Results

Table 4 reports the estimates of $\widehat{B}$, $\widehat{H}$ and $\hat{b}$. I compute bootstrapped standard errors using 200 replications (the standard approach in this literature – e.g., Almunia and Lopez-Rodriguez (2018)). The first row of Table 4 reports these parameter estimates for the pre-SOX period (1993-2002); this involves using N_i^{PreSOX} as the dependent variable in Equation (4), with the subsequent equations defined analogously. As would be expected from Figure 2, the estimated $\hat{b}$ is relatively small and statistically insignificant; indeed, it has a negative sign, suggesting that there are slightly fewer firm-year observations immediately below \$75 million than would be expected. This casts doubt on the possibility that there is some economic reason underlying bunching under the \$75 million threshold during the SOX period.

It is possible that bunching under the \$75 million threshold may be manifested in the pre-SOX period at a value of public float that corresponds not to the nominal \$75 million level, but to its inflation-adjusted equivalent. Thus, I use the inflation data provided by the Bureau of Labor

Statistics[39] to adjust the threshold for the rate of inflation from the midpoint of the pre-SOX period to the midpoint of the SOX period. When implementing a similar analysis using the inflation-adjusted threshold of $57 million, the estimated $\hat{b}$ remains small and statistically insignificant.

The second row of Table 4 reports $\hat{B}$, $\hat{H}$ and $\hat{b}$ for the SOX period (2003-2015). There are approximately 151 excess firm-years below the threshold; the counterfactual number of firm-year observations in the interval from $66-$75 million is 1121, and the observed number is 1272. Hence, about 12% of observations in this range represent "bunchers" (although it is not possible to identify *which* of these observations are bunchers and which are not). The estimated $\hat{b}$ is approximately 1.2, and is statistically significant at the 5% confidence level (using bootstrapped standard errors, as described above).

While it is statistically significant (as well as being visibly apparent in Figure 3), the magnitude of bunching is arguably relatively modest. The estimated bunching parameter of 1.2 implies that bunching firms reduce their public float by up to $1.2 million in order to remain below the $75 million threshold. In a stylized setting in which all firms are responsive to the threshold, this means that firms whose (undistorted) public float would fall in the range [$75 million, $76.2 million] would choose to bunch just below $75 million. While this estimate should be interpreted with caution because it pertains to a highly stylized scenario, the range over which firms are willing to bunch appears to be relatively narrow range.

The number of excess public float observations below the threshold (151) amounts to about 12 per year over 2003-2015 (or about 20 per year if I assume they are concentrated in 2003-2009, as suggested by Figures 5 and 6 below). This seems modest, but it is important to place this estimate in context, particularly in relation to other types of responses by firms to SOX. For instance, Kamar, Karaca-Mandic and Talley (2009) study the impact of SOX on the propensity of small public firms to engage in going-private transactions through acquisition by a private acquirer. The descriptive statistics that they report in their Table 1 (p. 117) imply that there were 24 "excess" private acquisitions of small US public firms over their post-SOX period (August 1, 2002 to December 31, 2004) of about two years (i.e., about 12 per year).[40] Although they focus on

[39] See http://www.bls.gov/data/inflation_calculator.htm.

[40] Before SOX (from January 1, 2000 to July 30, 2002, their Table 1 reports 974 acquisitions of US public firms; of these, 22% of targets were "small" and 40% of "small" targets were acquired by private acquirers – i.e. there were 86 private acquisitions of small targets. After SOX (from August 1, 2002 to December 31, 2004), they find 679 acquisitions of US public firms; of these, 30% of targets were "small" and 54% of "small" targets were acquired by private acquirers – i.e. there were 110 private acquisitions of small targets.

a very different margin of firm responsiveness, the magnitude of the going-private response that Kamar, Karaca-Mandic and Talley (2009) find seems highly comparable to the magnitude of the bunching response that I find here.

While there are approximately 151 excess firm-years below the threshold, the missing mass above the threshold consists only of about 56 firm-years, thus implying that $B > H$. This is difficult to explain. In the SOX literature, much attention has been paid to the exit of firms subject to SOX from the market (e.g., Bartlett, 2009; Kamar, Karaca-Mandic and Talley, 2009) or from reporting entity status (e.g., Leuz, Triantis and Wang, 2008). However, exit from the sample (e.g., through a "going private" transaction) by firms just above the threshold would make the missing mass above the threshold appear larger (rather than smaller) in relation to the excess mass below the threshold. Moreover, substantial entry onto the public markets in the SOX period by firms just above the threshold appears implausible. I seek to address this issue and to reconcile the estimated values of B and H issue in Section 5.3 below.

5.2) Comparing Subperiods within the SOX Period

The results in Table 4, Row 2 use the entire SOX period (2003-2015). There is some evidence that SOX compliance costs were particularly large in the early years of implementation (e.g., Grundfest and Bocher, 2007). Thus, firms' possible aversion to crossing the $75 million threshold may have arisen due to high initial costs of SOX 404 compliance, and may have disappeared over time as accounting and legal practitioners became more familiar with the regime and compliance costs fell. To address this possibility, I compare earlier and later subperiods within the SOX period. In particular, I divide the SOX period into two subperiods – 2003-2009 and 2010-2015 (when the exemption from SOX 404(b) was made permanent). I plot the histograms and estimate the magnitude of bunching for each of these subperiods. Unfortunately, results from estimating the magnitude of bunching for shorter subperiods within the SOX period (such as for each year) are less meaningful because of the limited number of observations.

Figure 5 shows a histogram using firm-year observations from 2003-2009. The visual evidence of bunching is somewhat more pronounced than in Figure 3 (for 2003-2015). The estimates in Row 3 of Table 4 show that almost all of the SOX period bunching is concentrated in the 2003-2009 subperiod. Of the 151 excess observations just below the $75 million threshold over 2003-2015, 146 occur in the 2003-2009 subperiod. The estimated bunching parameter of about 1.6

is larger than that in Row 2 for the entire SOX period (implying that firms are willing to reduce their public float by a larger amount, up to $1.6 million).

Figure 6 shows a histogram using firm-year observations from 2010-2015. There is very little, if any, visual evidence of bunching in Figure 6. Consistent with this, the estimates in Row 4 of Table 4 show no detectable evidence of bunching in the 2010-2015 subperiod. The estimated excess number of observations is only 5. This is statistically insignificant, but the estimate is sufficiently precise that an excess mass of more than 44 observations can be ruled out at the 95% level. The estimated bunching parameter of about 0.15 is quite close to zero.

As discussed previously, the magnitude of bunching is not a sufficient statistic for the net compliance cost, and so an absence of bunching may be consistent with high compliance costs if the costs of managing public float are large. Nonetheless, Figures 5 and 6 (and the corresponding estimates in Rows 3 and 4 of Table 4) suggest an apparent decline in firms' willingness to bunch over the course of the 2003-2015 period. Assuming that the costs of managing public float did not change over this period, this is consistent with the SOX legislation creating substantial but transitory costs of compliance for firms. It is impossible to rule out the possibility that compliance costs continue to be substantial (albeit perhaps lower than in 2003-2009). However, the evidence here that firms in recent years do not seem to view compliance costs as being large in relation to the costs of managing public float is quite relevant for current debates about the reform of securities regulation (especially in the direction lowering compliance burdens on smaller firms and raising the thresholds for the application of particular provisions of the securities laws). In addition, it should be remembered that the accounting literature on firms' management of public float (e.g., Gao, Wu and Zimmerman, 2009; Nondorf, Singer and You, 2012; Gao, 2016) – as well as the bunching result above for 2003-2009 - suggest that the management of public float is by no means prohibitively costly.

Alternative explanation for the bunching patterns I observe over time might be constructed along the following lines. Suppose that there exists a subset of firms that find SOX compliance to be particularly costly and/or can manage public float at low cost. When SOX is introduced, these firms are especially likely to bunch. Over time, however, the costs of maintaining a low public float may grow, leading them to cross the threshold and thereby cease to bunch. This explanation requires that new firms with high compliance costs and/or low costs of managing public float do not enter the bunching region in later years (either by growing from lower initial levels of public float or by entering the public-traded market). If that is the case, then this explanation is difficult to distinguish from explanations involving learning over time by firms and by accounting and legal professionals that reduces compliance costs.

Alternatively, suppose that compliance costs were initially modest but that uncertainty over whether future regulation of AFs would be more onerous created a substantial option value to remaining in NAF status (especially given the one-way ratchet that makes escaping AF status very difficult). As regulatory uncertainty diminished over time, this option value would become smaller, inducing less bunching. All of these explanations are consistent with a pattern of relatively more bunching immediately after SOX.[41]

5.3) A "Pre-Reform Counterfactual" Approach

Equation (4) implies that the counterfactual density is estimated using data for the SOX period. This conforms to standard bunching analysis, in which a maintained assumption is that the threshold does not affect the behavior of firms that are located far from it. However, in contrast to many bunching studies (in which the regulatory threshold does not change over time), I observe pre-SOX data. Thus, it is also possible to use the density of firms in the pre-SOX period to control for the density of firms in the SOX period.

In particular, it is possible to imagine two alternative counterfactuals – one in which the threshold does not exist because *no* firms (at least in the $0-150 million range) are subject to the regulations, and one where the threshold does not exist because *all* firms are subject to the regulations. Arguably, the former is more relevant for policy purposes, given that the exemption for smaller firms has been made permanent and that policy debates have focused on increasing the threshold for exemption. The pre-SOX counterfactual density provides us with a glimpse into the situation where no firms are subject to the regulations. In such a scenario, there may be more observations of public float above the excluded interval than in the actual SOX period density (for instance, if some firms above the excluded interval go private). If, as in Equation (4), I use the SOX period to derive the counterfactual density, then the counterfactual density may arguably be biased downward (relative to the "true" counterfactual); in turn, this may lead to H being underestimated.

[41] The pattern of bunching in Figure 5 appear to be even stronger when restricting attention to a shorter span of time following SOX (such as 2003-2005; however, the pattern is noisier due to the smaller number of observations.

The “prereform counterfactual” approach can be implemented as follows. Define:

$$N_i^{PreC} = \begin{cases} N_i^{SOX} & if i \in [L, U] \\ kN_i^{PreSOX} & otherwise \end{cases} \tag{9}$$

Here, k is an arbitrary constant that scales the pre-SOX density to correct for the fact that our dataset has a different number of observations in the SOX period than in the pre-SOX period, and to account for possible differences in the SOX and pre-SOX densities in bins outside the excluded interval.[42] I then estimate the counterfactual density as follows:

$$N_i^{PreC} = \alpha + \sum_{j=1}^{5} [\beta_j (Mid_i)^j] + \gamma_i I_i + \epsilon_i \tag{10}$$

where the variables are as defined earlier.

We choose the constant k in Equation (9) based on an iterative procedure that starts with a value of k = 0.985 (to scale down the pre-SOX density by the ratio of SOX to pre-SOX observations, as reported in the descriptive statistics in Table 1). I then generate values of B and H using this candidate k. As $B > H$ for this candidate k, I then iteratively reduce the value of k in order to further lower the counterfactual density and increase the estimated H until convergence is reached (i.e., when I approximate the condition that $B = H$). It turns out that $B = H$ is satisfied when $k = 0.938$.

Note that in using the pre-SOX density, I use the nominal dollar values of public float, and do not adjust for inflation. As noted in Section 5.1 above, adjusting for inflation makes little difference to the conclusions I draw from the pre-SOX density. Moreover, the density of firms in bins below \$50 million looks quite different (and involves substantially larger numbers of observations) in both the pre-SOX and SOX periods. As the inflation-adjusted value of \$75 million is around \$57 million in the pre-SOX period, bins below \$50 million would have to be used in computing the inflation-adjusted counterfactual density, potentially introducing extraneous factors that are less relevant to firms that are close to the threshold.

The bunching parameters that I estimate using this approach are reported in Row 5 of Table 4. The estimated excess mass below the threshold is about 107 observations (as is the estimated missing mass above the threshold, by construction). The estimated $\hat{b}$ is approximately 0.8, and is statistically significant at the 5% confidence level (using bootstrapped standard errors as described

[42] It is also possible to use an additive constant instead; this leads to very similar results.

above). Thus, the specification in Equation (10) enables us to more satisfactorily reconcile the observed amount of bunching below the threshold with the observed missing region above the threshold. However, the results are otherwise quite similar to the baseline results in Row 2 of Table 4. Indeed, the magnitude of observed bunching below the threshold is even more modest than in the baseline estimates.

5.4) Serial Dependence and Dynamic Bunching Estimation

The baseline analysis treats each underlying firm-year observation as being independent, even though the underlying data has a panel structure with (in many cases) repeated observations on the public float of the same reporting entity. This is the standard ("static") approach that is widely used in the bunching literature. However, Marx (2018) argues that this "static" approach can lead to potential bias in the estimation of the counterfactual density using panel data due to serial dependence among repeated observations on the same unit: "If the running variable [here, public float] exhibits positive serial dependence, then reducing its value today in order to bunch will lower its value tomorrow. This is true regardless of where that value lies in tomorrow's distribution, implying the potential for distortion far from the notch . . ." (Marx, 2018, p. 3).

This serial dependence may arise because reducing public float in order to remain below the threshold may persistently lower future values of the firm's public float, even at values far from the threshold. For instance, consider a firm with public float of $60 million, that would counterfactually grow by 10% to $66 million in the subsequent year. Suppose this firm reduces its growth rate to zero in order to avoid approaching the $75 million threshold in the future. This firm remains below the excluded interval; by increasing the observed density below the excluded interval, its behavior affects the estimated counterfactual density. Moreover, suppose that several years later, this firm crosses the threshold. Even so, its public float may remain permanently lower than in the absence of its response to the threshold. If its public float is $80 million when it would counterfactually have been $86 million, the observed density above the excluded interval will be lower (again, affecting the estimated counterfactual density). Thus, serial dependence can potentially bias the estimation of the counterfactual density and hence the estimated bunching parameter.

Marx (2018) proposes an alternative "dynamic" bunching estimation approach that addresses the problem of serial dependence. This focuses on computing the growth rate of the running variable for bins around the threshold. Formal implementation of this approach ideally requires a large number of observations in narrow bins around the threshold. However, a less formal test in the spirit of Marx (2018) involves computing the median growth rate of public float – i.e., the rate of growth from the base year *t* to the subsequent year (*t* + 1) – for firm-years in each bin of public float in the base year *t*. A counterfactual distribution of median growth rates across bins can be estimated using a fifth-order polynomial function of public float (excluding the interval ranging from \$66-\$83 million, as before).

$$G_i = \alpha + \sum_{j=1}^{5} [\beta_j (Mid_i)^j] + \gamma_i I_i + \epsilon_i \tag{11}$$

Here, G_i is the median growth rate of public float - from the base year *t* to the subsequent year (*t* + 1) – for observations in bin *i* of public float in the base year *t*.[43] Other variables are as defined earlier. Essentially, this approach replaces the count of observations in each bin (used in Equation (4)) with the median growth rate of the observations in that bin.

The approach in Equation (11) allows us to test whether growth rates are lower just below the threshold, and also provides a simple visual test of whether serial dependence affects the counterfactual density further away from the threshold. Figure 7 plots the median growth rates across bins and the counterfactual density estimated using Equation (11) for the pre-SOX period (1993-2002). Figure 8 plots the median growth rates across bins and the counterfactual density estimated using Equation (11) for the SOX period (2003-2015).[44] In Figure 8, the growth rate of public float in the SOX period is negative, and lower than the counterfactual growth rate, just below the \$75 million threshold. However, the growth rate is not much further below the counterfactual in the SOX period than in the pre-SOX period (Figure 7). Overall, this pattern seems generally consistent with the earlier findings of a modest amount of bunching after SOX.

To test for the effects of serial dependence, I can examine the growth rates of public float before and after SOX for bins below the excluded interval (covering public float from \$50 million to \$65 million). Comparing Figures 7 and 8, it appears that public float growth rates are higher – rather than lower – for these bins during the SOX period relative to the pre-SOX period. Thus, it

[43] Median – rather than mean – growth rates of public float are used because mean growth rates exhibit considerable volatility; however, the general patterns are similar for mean growth rates. Note also that the bins used here are of width \$5 million (rather than \$1 million), in order to smooth the variability of growth rates. Results using bins of width \$1 million yield similar – although noisier – patterns.
[44] Note that these figures are analogous to Figure 5 in Marx (2018).

appears unlikely that the estimated counterfactual density in the baseline analysis is significantly affected by serial dependence in public float.

5.5) Other Issues and Robustness Checks

As described in Section 3.1, the SOX period data used in the baseline analysis excludes 1508 observations for which, in the prior year, the firm reported public float below $75 million but either also reported being an AF or had ever reported a public float of $75 million or more in 2002 or any subsequent year. This was on the grounds that such firms are insensitive to the threshold and are not part of the pool of potential bunchers. Figure 9 is a histogram representing the number of firm-year observations of public float during the SOX period, with the 1508 excluded observations on these "insensitive" firms included. The pattern of visually apparent but modest bunching in Figure 9 does not seem fundamentally different from that in Figure 3 (where the observations on "insensitive" firms are excluded).

While the primary concern discussed in Section 3.1 above is that aspects of the definition of AFs may make some firms insensitive to the threshold, there are also some scenarios in which the estimated magnitude of bunching may be a biased upwards. If so, even the modest estimates of bunching found here may overstate firms' responsiveness to the threshold. For instance, the prior literature has in some instances found a negative market reaction to the imposition of SOX 404(b); Iliev (2010, p. 1190) finds about a 4% abnormal return for small firms that received a temporary exemption from compliance, relative to other firms. In this scenario, a firm that crosses the threshold will potentially experience a decline in market value and hence in public float. A decline in price upon crossing the threshold may result in public float remaining below the threshold. It is thus possible that some of the observations that appear to be "bunching" below the threshold actually represent firms that sought to locate *above* the threshold.

It should also be noted that observed bunching below the threshold may potentially be due not to the compliance costs of securities regulation, but to the loss of insiders' private benefits of control from crossing the threshold. In this scenario, observed bunching may overstate the impact of compliance costs. The past literature does not suggest that private benefits of control are a significant factor in decisions to avoid SOX 404(b). For example, Gao, Wu and Zimmermann (2009) find no evidence that the NAFs that engage in strategies to remain below the threshold have weaker governance, as measured by standard indices. In addition, *a priori* considerations suggest that private benefits of control are likely to be determined more strongly by state corporate law than by federal securities regulation. In principle, however, it is possible that the information provided in an auditor attestation may provide grounds for an action alleging a breach of fiduciary duty under corporate law, and thereby constrain insiders' extraction of private benefits.

Finally, the existence of the SEC's structured data from 2009 provides an opportunity for validating the paper's findings in a way that goes beyond the discussion in Section 3.2 above (which primarily assessed the extent of the coverage of public float reports by the scraped dataset). In particular, it is possible to analyze bunching by firms under the $75 million threshold in the SEC data for the period over which I have the structured public float data (2009-2020). Of course, this does not allow us to compare the pre-SOX and SOX periods. Nonetheless, it provides additional evidence concerning the apparent absence of bunching in later years of our sample period (as shown in Figure 6 for 2010-2015). Figure 10 is a histogram representing the number of firm-year observations of public float in the SEC's structured data over 2009-2020, using bins of width $5 million. There is little visual evidence of bunching below the $75 million threshold, although there is some modest apparent missing mass just above the threshold. Overall, the evidence from the SEC's structured data is highly consistent with the conclusion reached using this paper's dataset, namely that there is very limited, if any, evidence of bunching by firms in response to securities regulation in recent years.

6) Conclusion

This paper brings both a new dataset and a new empirical approach to bear on longstanding but unresolved questions regarding the consequences of securities regulation. I collect public float data for a large sample of reporting entities for fiscal years 1993-2015 by "scraping" this information from 10-K filings. I apply a bunching approach to this new dataset, comparing the number of observations immediately below the $75 million public float threshold to a smooth counterfactual density. The $75 million threshold is of particular importance because many important provisions of US securities law, most notably certain widely-discussed elements of the SOX legislation, apply only above this threshold.

Prior to SOX (i.e., over 1993-2002), there is no detectable bunching. Following the SOX legislation (i.e., over 2003-2015), I document statistically significant bunching under the $75 million regulatory threshold. However, the estimated magnitude of bunching is relatively modest.

Importantly, this bunching is concentrated in the earlier part of the period (2003-2009), and is virtually absent later in our sample period (2010-2015). Although an absence of bunching is consistent with high compliance costs if the costs of managing public float are large, the evidence in this paper suggests that firms in recent years do not seem to view compliance costs as being large in relation to the costs of managing public float. The thrust of recent discussions about the reform of securities regulation has been on lowering compliance costs for smaller firms and raising the thresholds for the application of particular provisions of the securities laws. The evidence in this paper is thus particularly relevant for these debates.

References

Alexander, C. R., S.W. Bauguess, G. Bernile, Y.H.A. Lee, and J. Marietta-Westberg (2013) "The Economic Effects of SOX Section 404 Compliance: A Corporate Insider Perspective" *Journal of Accounting and Economics*, 56, 267-290.

Almunia, M., and López-Rodríguez, D. (2018) "Under the Radar: The Effects of Monitoring Firms on Tax Compliance" *American Economic Journal: Economic Policy*, 10, 1-38.

Bartlett III, R. P. (2009) "Going Private but Staying Public: Reexamining the Effect of Sarbanes-Oxley on Firms' Going-Private Decisions" *University of Chicago Law Review*, 76, 7-44.

Burgstahler, D. and I. Dichev (1997) "Earnings management to avoid earnings decreases and losses" *Journal of Accounting and Economics*, 24(1), 99-126.

Chhaochharia, V. and Y. Grinstein (2007) "Corporate Governance and Firm Value: The Impact of the 2002 Governance Rules" *Journal of Finance*, 62, 1789-1825.

Chhaochharia, V., Otto, C. A., and Vig, V. (2011) "The Unintended Effects of the Sarbanes–Oxley Act" *Journal of Institutional and Theoretical Economics*, 167(1), 149-164.

Coates IV, J. C. and S. Srinivasan (2014) "SOX After Ten Years: A Multidisciplinary Review" *Accounting Horizons*, 28, 627-671.

Chetty, R., Friedman, J. N., Olsen, T., & Pistaferri, L. (2011) "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records" *Quarterly Journal of Economics*, 126(2), 749.

Dharmapala, D. (2016) "Estimating the Compliance Costs of Securities Regulation: A Bunching Analysis of Sarbanes-Oxley Section 404(b)" CESifo Working Paper No. 6180.

Dharmapala, D. and V. S. Khanna (2016) "The Costs and Benefits of Mandatory Securities

Regulation: Evidence from Market Reactions to the JOBS Act of 2012" *Journal of Law, Finance, & Accounting*, 1, 139-186.

Dharmapala, D., Slemrod, J., and Wilson, J. D. (2011) "Tax policy and the missing middle: Optimal tax remittance with firm-level administrative costs" *Journal of Public Economics*, 95(9), 1036-1047.

Ewens, M., K. Xiao, and T. Xu (2021) "Regulatory Costs of Being Public: Evidence from Bunching Estimation" NBER Working Paper 29143.

Gao, F. (2010) "Essays on the Incentives to Manage Public Float" Unpublished PhD dissertation, University of Rochester.

Gao, F. (2016) "To Comply or Not to Comply: Understanding the Discretion in Reporting Public Float and SEC Regulations" *Contemporary Accounting Research*, 33(3), 1075-1100.

Gao, F., Wu, J. S., & Zimmerman, J. (2009) "Unintended Consequences of Granting Small Firms Exemptions from Securities Regulation: Evidence from the Sarbanes-Oxley Act" *Journal of Accounting Research*, 47(2), 459-506.

Ge, W., Koester, A. and McVay, S. (2017) "Benefits and costs of Sarbanes-Oxley Section 404(b) exemption: Evidence from small firms' internal control disclosures" *Journal of Accounting and Economics*, 63(2-3), 358-384.

Grundfest, J. A. and S. E. Bocher (2007) "Fixing 404" *Michigan Law Review*, 105, 1643-1676.

Iliev, P. (2010) "The Effect of SOX Section 404: Costs, Earnings Quality, and Stock Prices" *Journal of Finance*, 65, 1163-1196.

Jia, H., Xie, H. and Ziebart, D.A. (2014) "An Analysis of the Costs and Benefits of Auditor Attestation of Internal Control over Financial Reporting" *Accounting Horizons*, *14*(3), pp.325-341.

Kamar, E., P. Karaca-Mandic, and E. Talley (2009) "Going-Private Decisions and the Sarbanes-Oxley Act of 2002: A Cross-Country Analysis" *Journal of Law, Economics, & Organization*, 25, 107-133.

Kinney, Jr., W. R. and M. L. Shepardson (2011) "Do Control Effectiveness Disclosures Require SOX 404(b) Internal Control Audits? A Natural Experiment with Small U.S. Public Companies" *Journal of Accounting Research*, 49, 413-448.

Kleven, H. J., and Waseem, M. (2013) "Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan" *Quarterly Journal of Economics*, 128(2), 669-723.

Leuz, C. (2007) "Was the Sarbanes–Oxley Act of 2002 really this costly? A discussion of evidence from event returns and going-private decisions" *Journal of Accounting and Economics*, 44(1), 146-165.

Leuz, C., Triantis, A., & Wang, T. Y. (2008) "Why do firms go dark? Causes and economic consequences of voluntary SEC deregistrations" *Journal of Accounting and Economics*, 45(2), 181-208.

Litvak, K. (2007) "The Effect of the Sarbanes-Oxley Act on Non-US Companies Cross-Listed in the US" *Journal of Corporate Finance*, 13, 195-228.

Marx, B. M. (2018) "Dynamic Bunching Estimation with Panel Data" MPRA Working Paper No. 88647.

Nondorf, M. E., Singer, Z., and You, H. (2012) "A study of firms surrounding the threshold of Sarbanes–Oxley Section 404 compliance" *Advances in Accounting*, 28(1), 96-110.

Randhawa, S. P. (2009) "The Sarbanes-Oxley Act and Small Public Companies" Working paper.

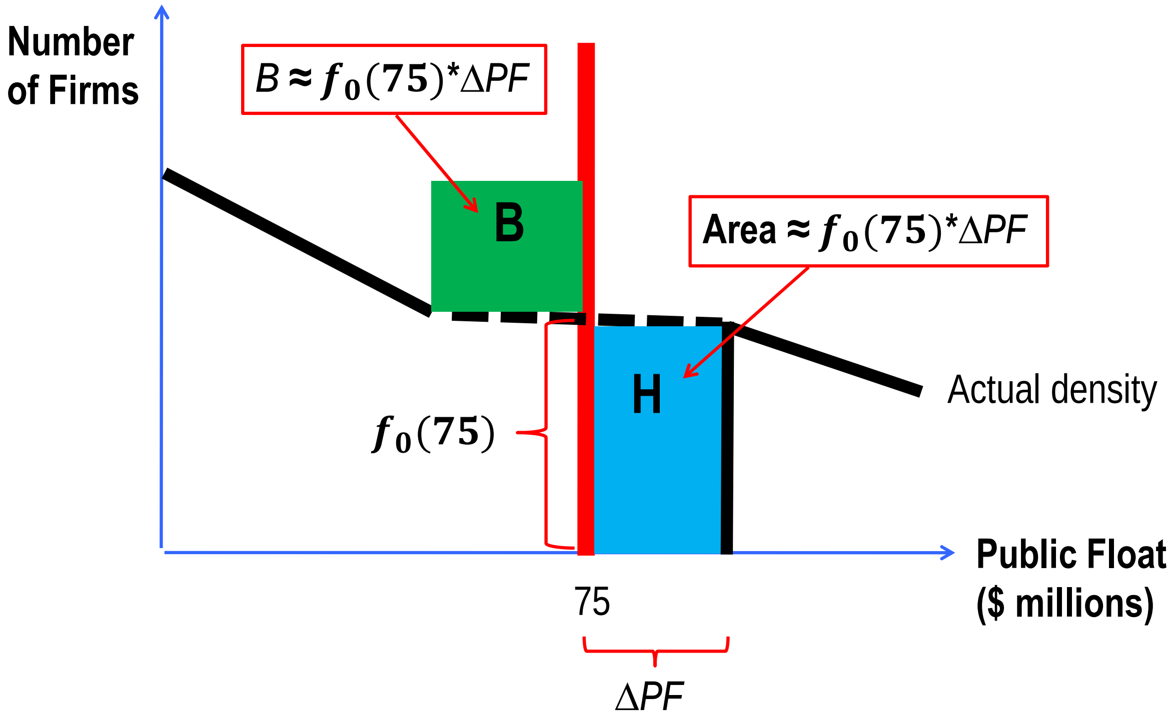
Saez, E. (2010) "Do taxpayers bunch at kink points?" *American Economic Journal: Economic Policy*, 2, 180-212.

Slemrod, J. (2013) "Buenas notches: lines and notches in tax system design" *eJournal of Tax Research*, 11 (3), 259-283.

Weber, D. and Y. S. Yang (2020) "The debt-equity choice when regulatory thresholds are based on equity values: Evidence from SOX 404" *The Accounting Review*, 95(2), 339-364.

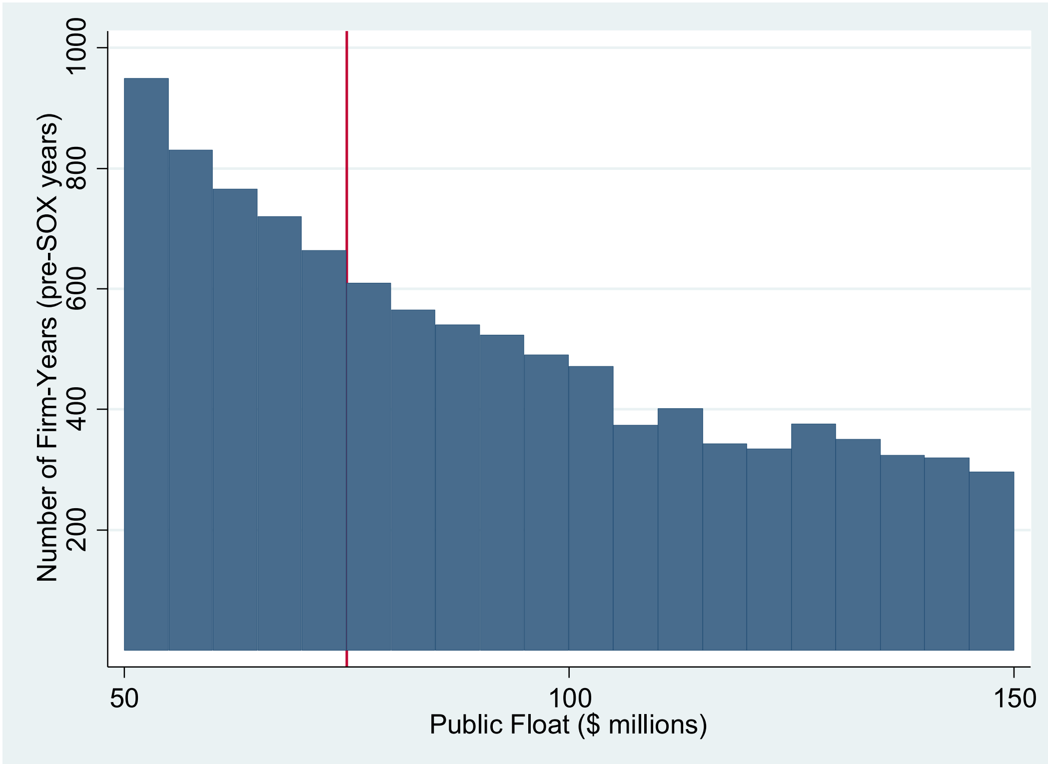
Zhang, I. X. (2007) "Economic consequences of the Sarbanes–Oxley Act of 2002" *Journal of Accounting and Economics*, 44(1), 74-115.

Figure 1: Estimating the Magnitude of Bunching



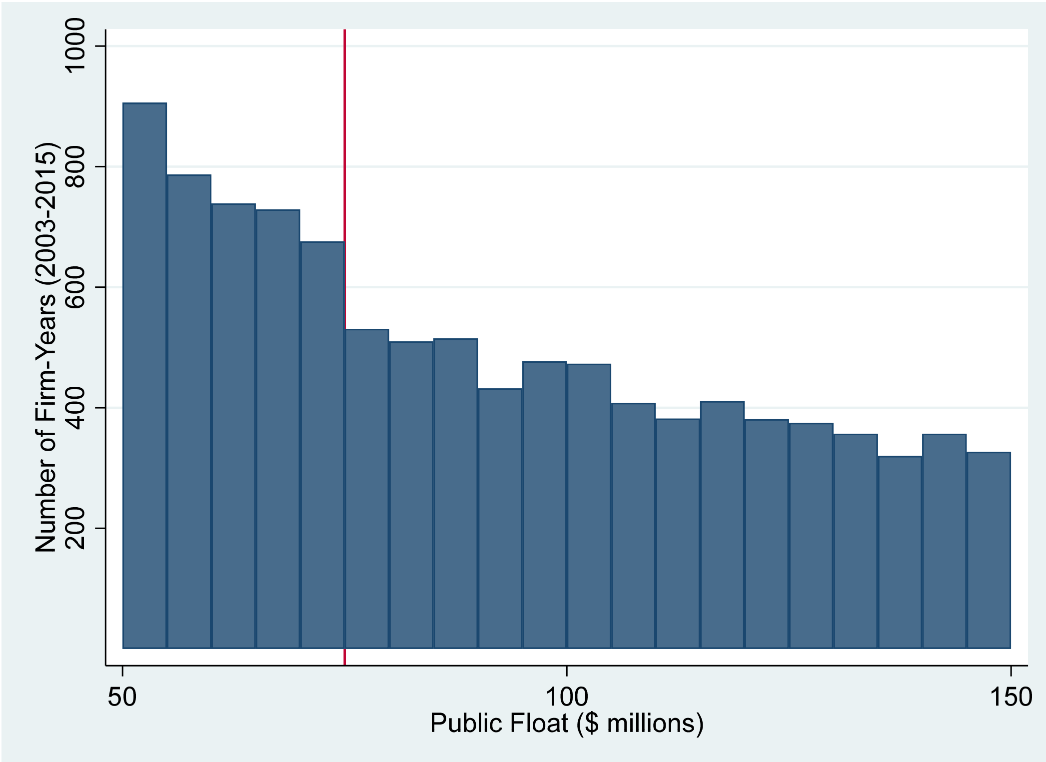
Notes: See the text for an explanation. B is the area of the bunching region, and H is the area of the missing region.

Figure 2: The Density of Public Float Observations in the Pre-SOX period (1993-2002)



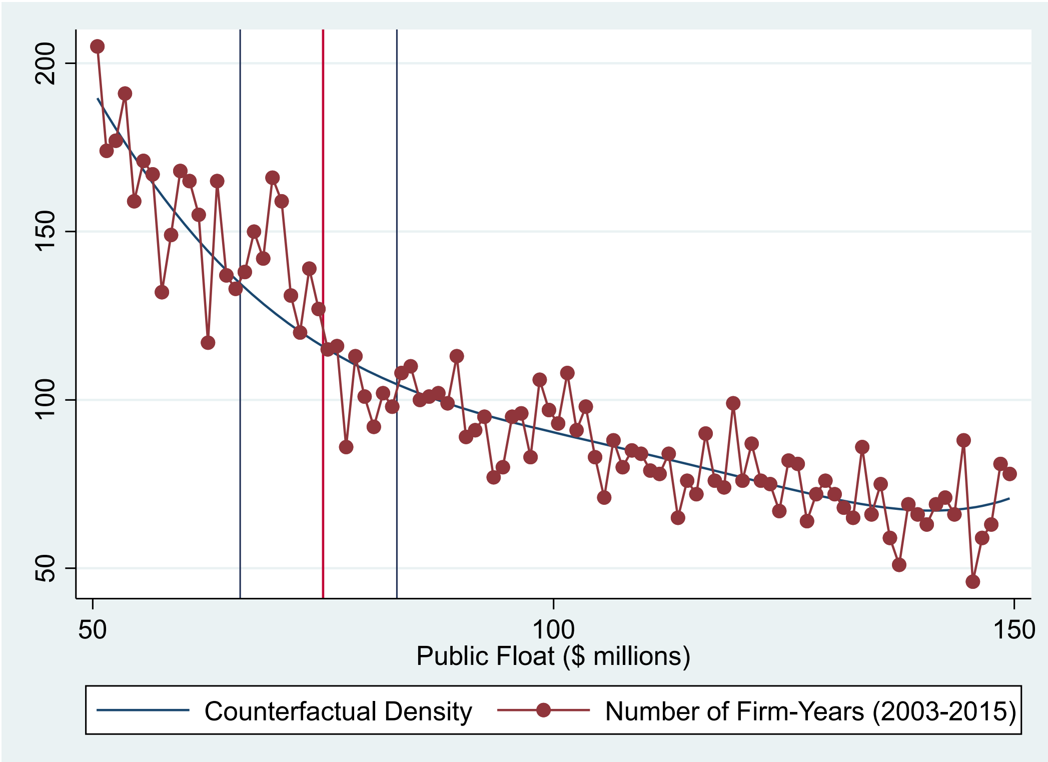
Note: This graph represents the number of firm-year level observations of public float during the pre-SOX period (1993-2002) within each of the bins representing public float observations in the range $50 million to $150 million. The vertical red line represents public float of $75 million. The bin width is $5 million, and the underlying number of public float observations is 10,247.

Figure 3: The Density of Public Float Observations in the SOX period (2003-2015)



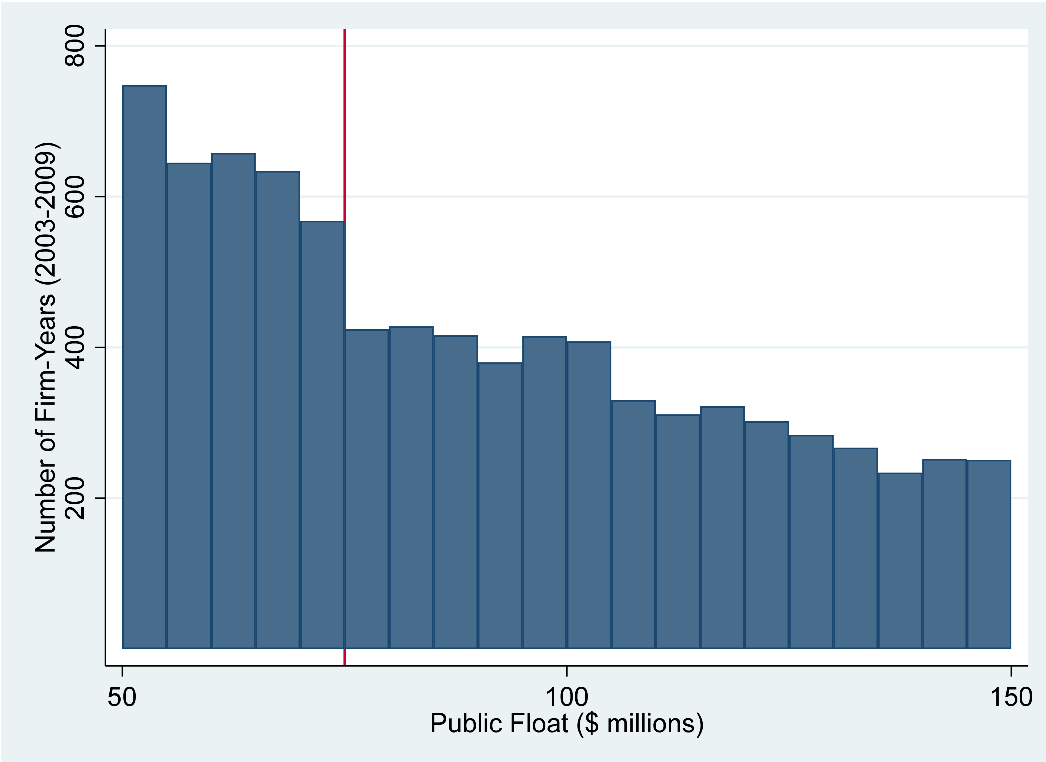
Note: This graph represents the number of firm-year level observations of public float during the SOX period (2003-2015) within each of the bins representing public float observations in the range $50 million to $150 million. The vertical red line represent public float of $75 million. The bin width is $5 million. The analysis excludes firm-years for which, in the prior year, the firm reported public float below $75 million but also reported being an AF (or had a prior public float report of $75 million or more). Taking account of this exclusion, the underlying number of public float observations is 10,093.

Figure 4: Actual and Counterfactual Densities for the SOX Period (2003-2015), using $1 Million Bins



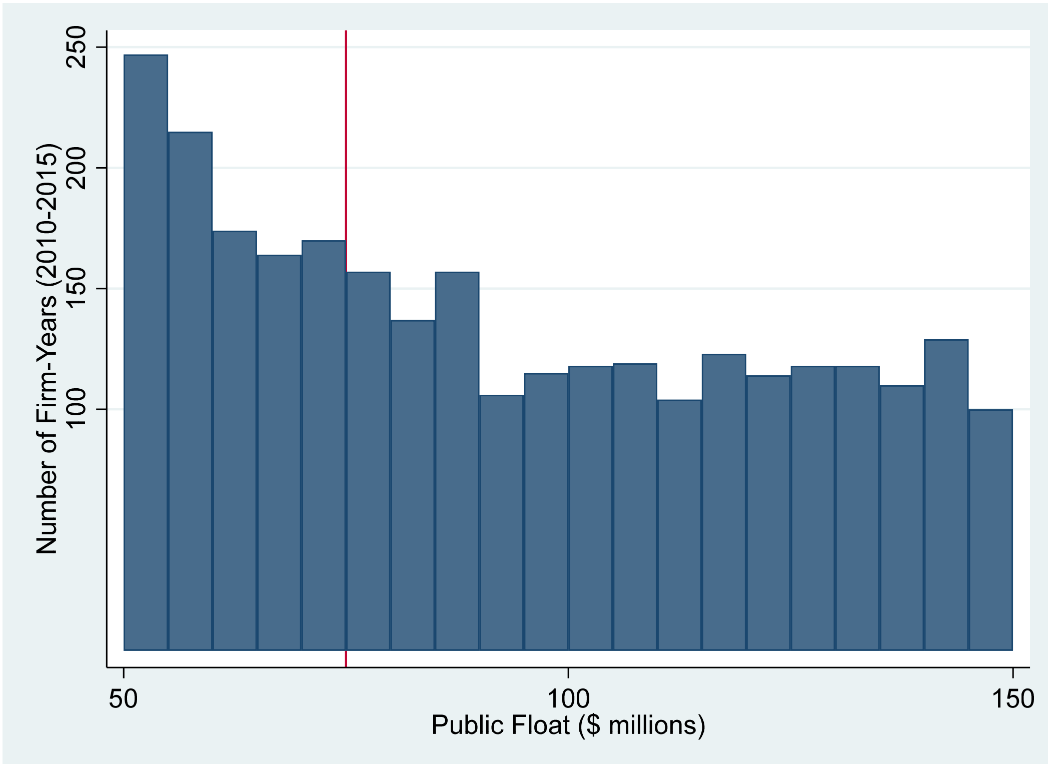
Note: This graph represents the number of firm-year level observations of public float during the SOX period (2003-2015) within each of the bins representing public float observations in the range $50 million to $150 million (using bins of $1 million width). The vertical red line represent public float of $75 million, and the vertical blue lines mark the limits of the excluded region (public float of $66 million to $83 million). The counterfactual density is computed using the fifth-order polynomial represented in Equation (4), using bins outside the excluded interval. The analysis excludes firm-years for which, in the prior year, the firm reported public float below $75 million but also reported being an AF (or had a prior public float report of $75 million or more). Taking account of this exclusion, the underlying number of public float observations is 10,093.

Figure 5: The Density of Public Float Observations in the Early SOX period (2003-2009)



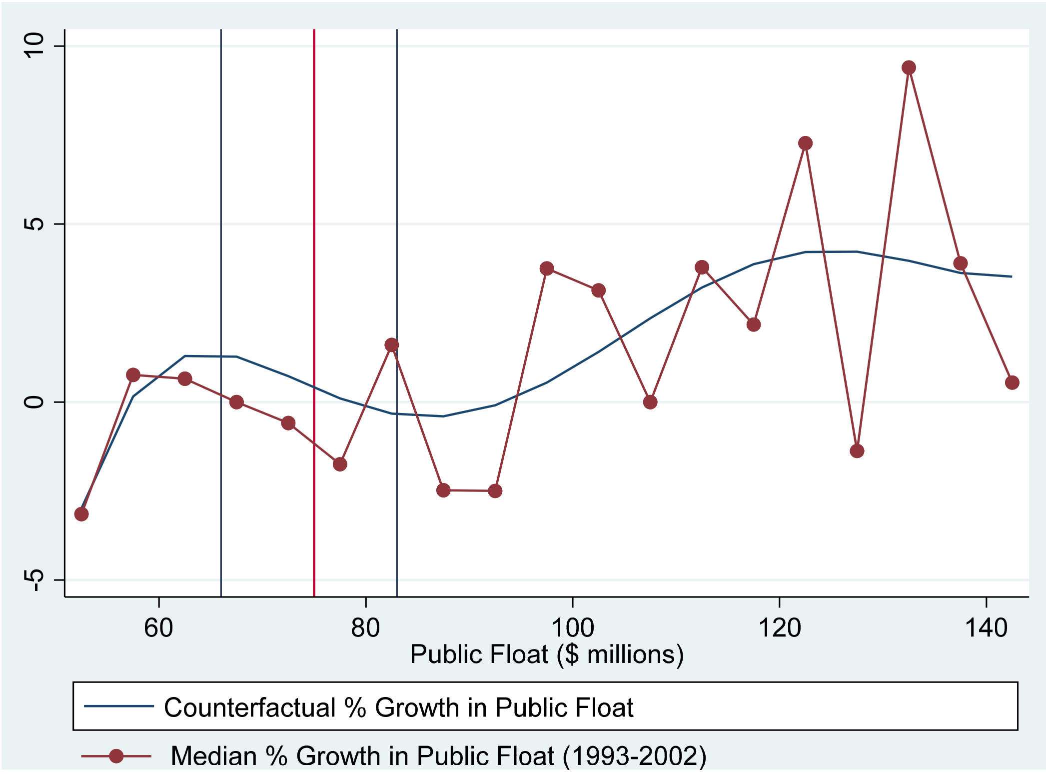
Note: This graph represents the number of firm-year level observations of public float during 2003-2009 within each of the bins representing public float observations in the range $50 million to $150 million. The vertical red line represent public float of $75 million. The bin width is $5 million. The analysis excludes firm-years for which, in the prior year, the firm reported public float below $75 million but also reported being an AF (or had a prior public float report of $75 million or more). Taking account of this exclusion, the underlying number of public float observations is 7,298.

Figure 6: The Density of Public Float Observations in the Later SOX period (2010-2015)



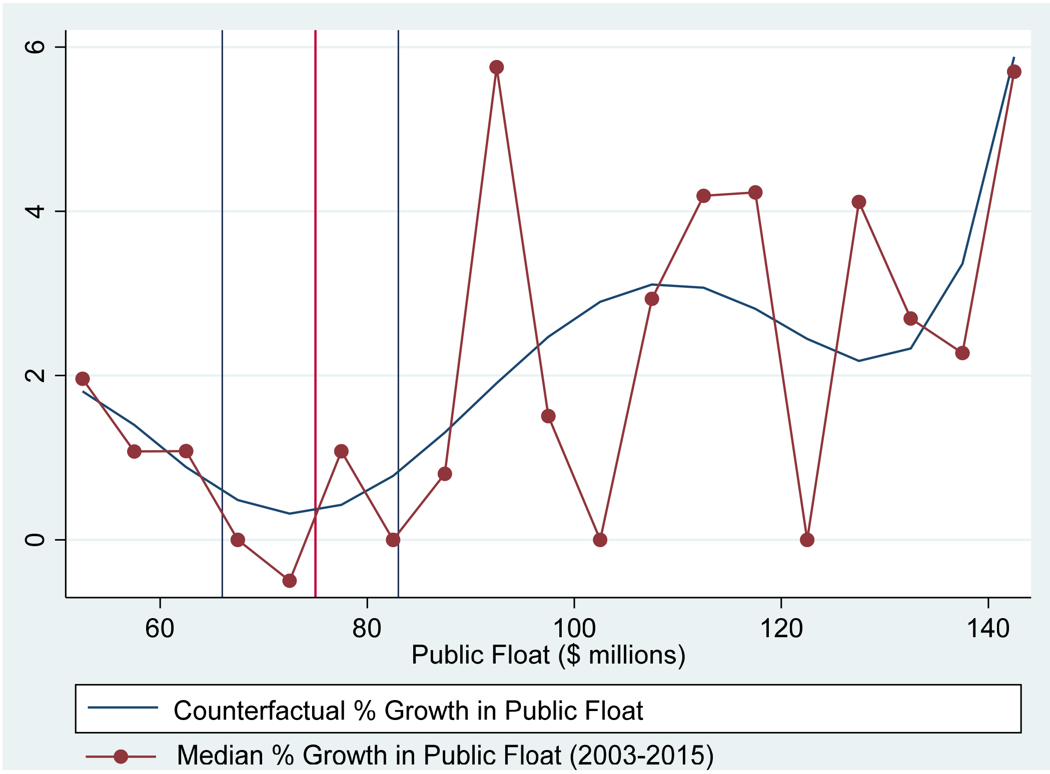
Note: This graph represents the number of firm-year level observations of public float during 2010-2015 within each of the bins representing public float observations in the range $50 million to $150 million. The vertical red line represent public float of $75 million. The bin width is $5 million. The analysis excludes firm-years for which, in the prior year, the firm reported public float below $75 million but also reported being an AF (or had a prior public float report of $75 million or more). Taking account of this exclusion, the underlying number of public float observations is 2,795.

Figure 7: Median Growth Rate of Public Float in the Pre-SOX Period (1993-2002)



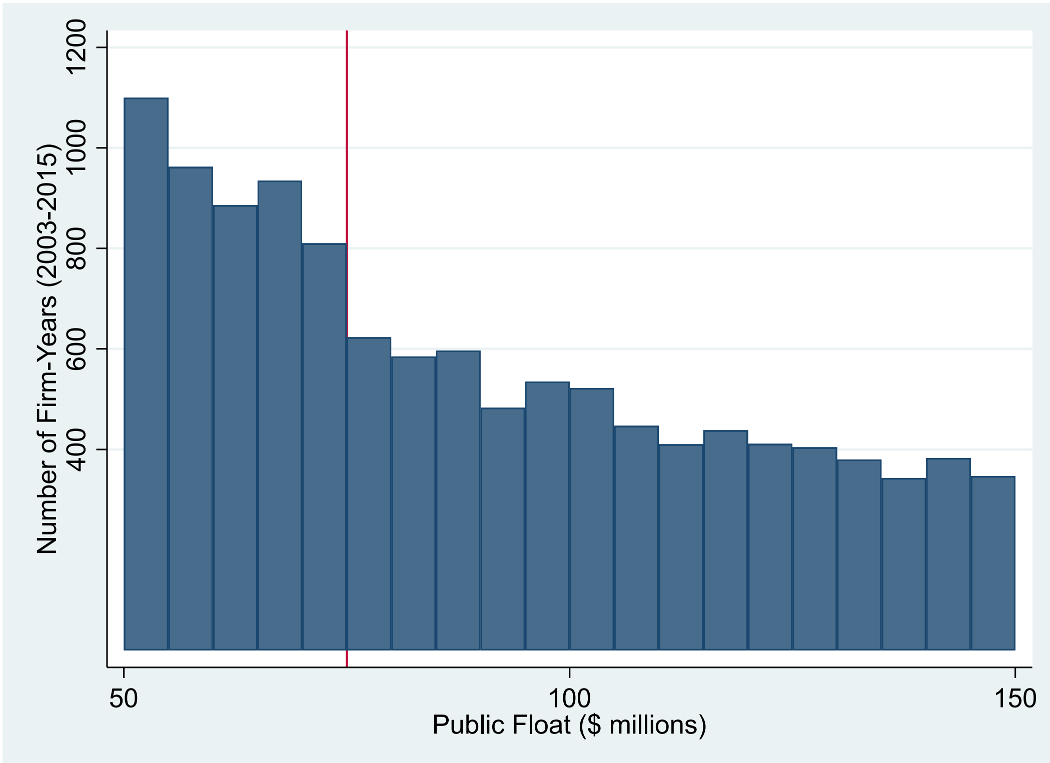
Note: This graph plots the median growth rate of public float from the current year to the next year, as a function of public float in the current year, for the pre-SOX period (1993-2002). The growth rate is shown for each of the bins representing public float observations in the range $50 million to $150 million (using bins of $5 million width). The vertical red line represent public float of $75 million, and the vertical blue lines mark the limits of the excluded region (public float of $66 million to $83 million). The counterfactual density is computed using the fifth-order polynomial represented in Equation (11), using bins outside the excluded interval.

Figure 8: Median Growth Rate of Public Float in the SOX Period (2003-2015)



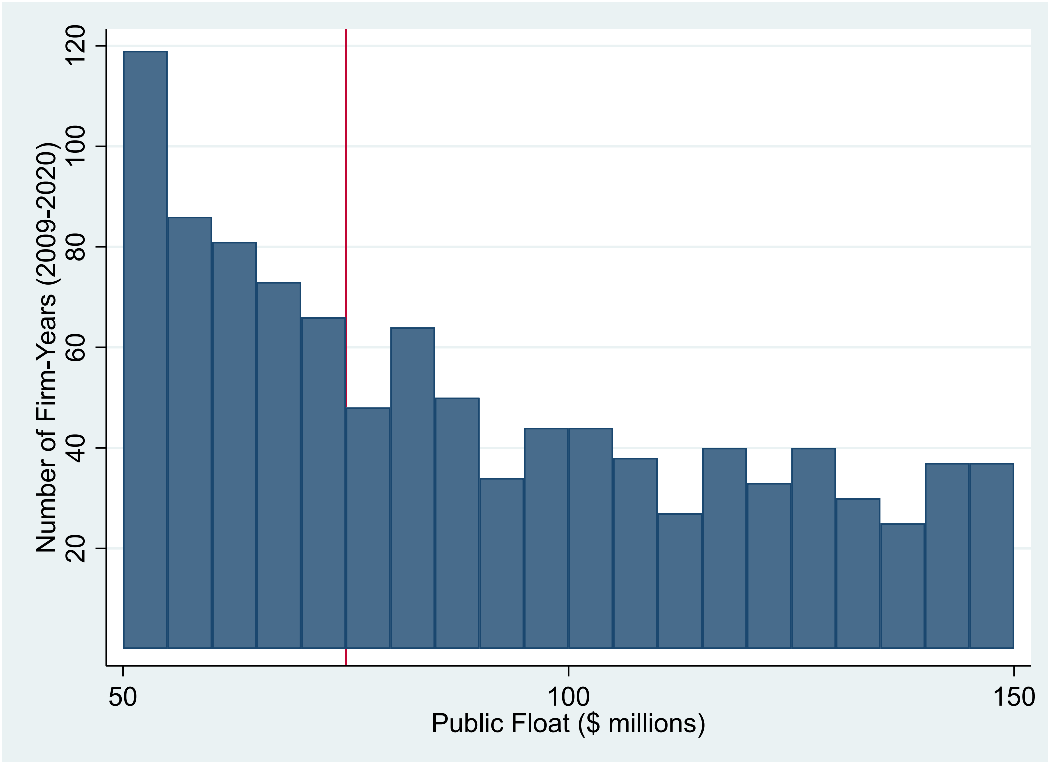
Note: This graph plots the median growth rate of public float from the current year to the next year, as a function of public float in the current year, for the SOX period (2003-2015). The growth rate is shown for each of the bins representing public float observations in the range $50 million to $150 million (using bins of $5 million width). The vertical red line represent public float of $75 million, and the vertical blue lines mark the limits of the excluded region (public float of $66 million to $83 million). The counterfactual density is computed using the fifth-order polynomial represented in Equation (11), using bins outside the excluded interval.

Figure 9: The Density of Public Float Observations in the SOX period (2003-2015), Including "Insensitive" Firm-Years



Note: This graph represents the number of firm-year level observations of public float during the SOX period (2003-2015) within each of the bins representing public float observations in the range $50 million to $150 million. The vertical red line represent public float of $75 million. The bin width is $5 million. Unlike in Figure 3, the analysis includes firm-years for which, in the prior year, the firm reported public float below $75 million but also reported being an AF (or had a prior public float report of $75 million or more). The underlying number of public float observations is 11,601.

Figure 10: The Density of Public Float Observations in the SEC Structured Data (2009-2020)



Note: This graph represents the number of firm-year level observations of public float during the 2009-2020 period, using the SEC's Structured Data on public float, within each of the bins representing public float observations in the range $50 million to $150 million. The vertical red line represent public float of $75 million. The bin width is $5 million. The underlying number of public float observations is 5,539.

Table 1: Descriptive Statistics for Public Float ($ millions)

Sample	Mean	Median	Standard Deviation	Number of Observations
Pre-SOX (1993-2002) sample with public float $50-150 million	89.94	85.09	28.45	10,247
SOX (2003-2015) sample with public float $50-150 million	91.16	86.60	28.98	10,093

Note: Public float is defined as the "aggregate worldwide market value of the voting and non-voting common equity held by its non-affiliates . . . as of the last business day of the issuer's most recently completed second fiscal quarter . . ." (17 CFR 240.12b-2). Public float is obtained using the web scraping approach described in the text, and is measured in millions of US$. The SOX (2003-2015) sample excludes firm-years for which, in the prior year, the firm reported public float below $75 million but also reported being an AF (or had a prior public float report of $75 million or more).

Table 2: Comparison of Descriptive Statistics for Public Float ($ millions) for this Paper's Dataset and the SEC Structured Dataset (Public Float $50-150 Million, 2009-2015)

	This Paper's Dataset	SEC Structured Dataset
Mean	88.17	87.00
Standard Deviation	28.78	29.53

Note: This table compares descriptive statistics for the dataset used in this paper and the SEC structured dataset for the years in which the coverage of the two datasets overlap (2009-2015). The comparison is restricted to observations with public float that falls within the range of $50 million to $150 million (as observed in this paper's dataset). The construction of this paper's dataset is described in detail in the text. The SEC's structured data is available at:
https://www.sec.gov/dera/data/financial-statement-data-sets.html

Table 3: Public Float Observations around the Enactment of SOX (2001-2005)

Year	Number of Public Float Observations in the Range $50-150 Million
2001	1,039
2002	979
2003	1,112
2004	1,197
2005	1,212

Note: This table reports the number of observations with public float that fall within the range of $50 million to $150 million for each year in a five-period (2001-2005) around the enactment of the Sarbanes-Oxley (SOX) legislation.

Table 4: Estimates of Bunching Parameters

		Bunching parameter ($\hat{b}$)	**Excess Mass ($\hat{B}$)**	**Missing Mass ($\hat{H}$)**	**Number of Observations (N)**
(1)	Pre-SOX Period (1993-2002)	-0.21	-28.45	37.08	10,247
		(0.38)	(53.12)	(32.05)	
(2)	SOX Period (2003-2015):	1.21**	151.01**	55.73	10,093
	Full Sample	(0.50)	(61.78)	(31.79)	
(3)	Early SOX Period (2003-	1.58**	146.06**	61.16**	7,298
	2009)	(0.62)	(57.56)	(27.95)	
(4)	Later SOX Period (2010-	0.15	4.95	5.44	2,795
	2015)	(0.62)	(19.97)	(16.94)	
(5)	SOX Period (2003-2015):	0.82**	106.57**	106.64**	10,093
	Using Pre-SOX Counterfactual	(0.39)	(50.31)	(43.85)	

Note: This table reports estimates of the bunching parameter and related measures. Bootstrapped standard errors, computed using 200 replications, are shown in parentheses.
*: significant at 10%; ** significant at 5%; *** significant at 1%.

about ECGI

The European Corporate Governance Institute has been established to improve *corporate governance through fostering independent scientific research and related activities.*

The ECGI will produce and disseminate high quality research while remaining close to the concerns and interests of corporate, financial and public policy makers. It will draw on the expertise of scholars from numerous countries and bring together a critical mass of expertise and interest to bear on this important subject.

The views expressed in this working paper are those of the authors, not those of the ECGI or its members.

Electronic Access to the Working Paper Series

The full set of ECGI working papers can be accessed through the Institute's Web-site (www.ecgi.global/content/working-papers) or SSRN:

Finance Paper Series	http://www.ssrn.com/link/ECGI-Fin.html
Law Paper Series	http://www.ssrn.com/link/ECGI-Law.html