Three Essays in Financial Economics

Adrien Alvero

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2022

Abstract

Three Essays in Financial Economics

Adrien Alvero

This dissertation studies three topics in financial economics. In the first chapter, *ESG Investing in Emerging Markets: Betting on Firm Fundamentals or Riding Investor Preferences?*, we examine the relation between firms' environmental, social, and governance (ESG) practices and the pricing of corporate bonds in emerging markets, which is an important yet understudied market for ESG-related issues. Firms with different ESG scores can have different costs of capital, either because ESG scores help forecast future cash flows – the *fundamental* channel – or because investors have non-pecuniary preferences for high-ESG-score assets – the *preference* channel. We identify the existence of a preference channel with a natural experiment – the historical opening of the Chinese onshore bond market – that leads to an increase in the proportion of international investors, who are more ESG-conscious. Consistent with theory, we find that the bond yield of companies with high ESG scores decreases more than that of companies with low ESG scores. By focusing on firms that also have bonds traded in the offshore market, which, as opposed to the onshore market, does not experience any change in regulation, we can control for issuer-time fixed effects in a triple difference design, hence reducing considerably the influence of the fundamental channel.

In the second chapter, *Watch what they do, not what they say: Estimating regulatory costs from revealed preferences*, we show that distortion in the size distribution of banks around regulatory thresholds can be used to identify costs of bank regulation. We build a structural model in which

banks can strategically bunch their assets below regulatory thresholds to avoid regulations. The resulting distortion in the size distribution of banks reveals the magnitude of regulatory costs. Using U.S. bank data, we estimate the regulatory costs imposed by the Dodd–Frank Act. Although the estimated regulatory costs are substantial, they are significantly lower than those in self-reported estimates by banks.

In the third chapter, *Fuzzy Bunching*, we introduce a new fuzzy bunching approach that is robust to noise. The existing bunching approach identifies the extent of bunching from a sharp spike in the probability density function. In many finance settings, however, the sharp spike could be diffused by data noise. The key idea behind our fuzzy bunching estimator is to identify bunching from the area of a bulge in the cumulative distribution function. The fuzzy bunching approach also avoids density estimation, which makes it easy to implement in sparse data. We provide the theoretical foundation of this approach and illustrate the advantages by using simulated and real data.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I have the deepest gratitude to all the members of my dissertation committee, Paul Tetlock, Kairong Xiao, Wei Jiang, Harry Mamaysky, and Tuomas Tomunen, for their guidance, mentorship, and insightful feedback during my Ph.D. journey. Paul honed my research and focused me on the big picture. He taught me the importance of a strong work ethic. Paul believed in me and my research and for that I am grateful. Kairong showed me the importance of perseverance, how to tell a story, and helped me to stay motivated. I am particularly grateful for being his co-author as well. Harry was always available to quickly answer my research questions and shared his invaluable lessons from his time working in industry to help guide me. Wei taught me how to well execute an empirical study early in the Ph.D., and her lessons in Financial Econometrics continually guide me when thinking about empirical design. While working as their Teaching Assistant, Paul, Harry, and Wei taught me how to engage a class, break down complex ideas, and find the joy in engaging and teaching students Finance. Tuomas and I spent hours discussing various economic and financial ideas. He challenged me and in turn refined my approach. I am thankful for his friendship and guidance through the years.

I also wish to thank Wang Renxuan, my co-author, former Ph.D. colleague, and closest friend, for showing me the importance of storytelling, for debating me and being a thought-partner no matter where in the world he is.

I would also like to thank all my professors during the Ph.D. program: Patrick Bolton, Emilie Breza, Kent Daniel, Lawrence Glosten, Geoffrey Heal, Robert Hodrick, Gur Huberman, Ralph Koijen, Lars Lochstoer, Xavier Sala-i-Martin, Tano Santos, Paolo Siconolfi, John Steinsson, Suresh

Sundaresan, Neng Wang, Motohiro Yogo, and Fanyin Zheng, for sparking my intellectual curiosity.

To my friends and Ph.D. colleagues: Aref Bolandnazar, Yifeng Guo, Tomas Mondino, Lira Mota, Daheng Yang, for all the great conversations that helped assimilate the literature and think about new research ideas.

Finally, I am indebted to my family. To my wife, Nina D'Souza, for always bringing me motivation when I needed it. To my son, Caleb, who joined us last year, for his contagious laugh. To my mother, Heidi, my father, Luis, and my brother, Cyril, for all their support and encouragement during my Ph.D. journey.

Dedication

To Caleb and Nina.

Chapter 1

ESG Investing in Emerging Markets: Betting on Firm Fundamentals or Riding Investor Preferences?[1]

1.1 Introduction

ESG investing – the practice of incorporating environmental, social, and governance scores into the investment process – has witnessed tremendous growth over the past decade. This growth is not expected to wane anytime soon, with total ESG assets under management projected to reach $50 trillion by 2025, or one-third of projected global assets.[2] Despite the large amounts of capital being at play, the exact channels through which ESG scores are incorporated into asset prices, if at all, remain ambiguous.

One the one hand, investors can use ESG scores for fundamental analysis of a company, hence improving the accuracy of cash flow forecasts and risks associated with them. Thus, ESG information is reflected in the price through what we call a *fundamental* channel. On the other hand, ESG scores can impact asset prices through what we call a *preference* channel, because some investors have non-pecuniary preferences for high-ESG-score companies. These investors are thus willing to pay a price premium to hold a high-ESG-score company in their portfolio.

Identifying the existence of a preference-induced ESG premium has important implications for all market participants. For firms, the preference-induced ESG premium renders certain ESG-score-improving projects more profitable because part of the expenses can be passed through to their equity holders or lenders through a lower cost of capital. For investors, the preference channel

[1]This chapter is based on Alvero and Renxuan (2022). We thank Travers Child, David Erkens, Dalibor Eterovic, Viktar Fedaseyeu, Robert Hodrick, Wei Jiang, Che Jiahua, Tom Liu, Harry Mamaysky, Paul Tetlock, Tuomas Tomunen, Kairong Xiao, and participants at the CEIBS Faculty Research Workshop for helpful comments and discussions. We thank Zheyang Zhu for excellent research assistance.

[2]Bloomberg Intelligence report from Diab and Adams (2021).

implies that that unless ESG preferences increase, the expected returns on high-ESG assets should be lower, all else being equal. For ESG score providers, the preference channel highlights the importance of providing scores that accurately reflect ESG practices, as opposed to greenwashing, which is the practice of reporting unsubstantiated and/or deceiving claims of ESG-friendliness.

Researchers face three main identification challenges when trying to isolate the preference channel of the ESG premium. First, *expected returns* are not observable. Using *average realized returns* as a proxy is inaccurate given the short history of ESG investing, during which preferences for ESG have increased.[3] Second, several confounding fundamental factors affect ESG scores and expected returns simultaneously, making firms' cross-sectional comparisons always subject to an omitted variable bias. Third, investors' preferences are usually not observable, making it challenging to link these preferences to expected returns.

In this paper, we identify the existence of a preference-induced ESG premium in emerging markets' corporate bonds with a natural experiment – the historical opening of the Chinese onshore bond market in 2019 – that lead to an increase in the proportion of international investors, who are arguably more ESG conscious (Martin and Moser, 2016; Riedl and Smeets, 2017; Hartzmark and Sussman, 2019; Ilhan et al., 2021; Bauer et al., 2021).

We focus on emerging markets' corporate bonds for several reasons. First, ESG investing has been understudied in emerging markets, despite them having an enormous role to play in the fight for more ESG-friendly practices across the globe.[4] For instance, according to Crippa et al. (2020), China represented 30% of global fossil CO_2 emissions in 2019, which is higher than the United States, the United Kingdom, and the EU combined (22.1%). Second, bonds, as opposed to stocks, have a cash flow schedule that is known in advance. This enables the computation of the *yield to maturity*, which is a more precise measure of expected returns compared with realized returns over

[3]Using realized returns as the proxy for expected returns is problematic in a period of increasing taste for ESG. Pastor et al. (2021a) provide evidence that the outperformance of green stocks in the 2010s was mainly because of the increase in climate concerns. In other words, the increase in the price of high-ESG stocks reflect equilibrium transition dynamics rather than a return premium.

[4]When looking at research papers on the ESG premium, specifically in credit markets across issuers, we found that only 3 out of the 24 papers had some analysis including emerging markets, or 12.5% of the papers. This number further drops if we include ESG-related research on stocks and green bonds.

a short period (Campello et al., 2008). Third, the opening of the Chinese onshore bond market to international investors provides us with a natural experiment to identify the preference channel of the ESG premium.

We develop a theoretical framework to guide our empirical analysis and outline the challenges in identifying the preference channel of the ESG premium.[5] We highlight two testable implications of the model: 1) all else being equal, companies with higher ESG scores have lower credit spreads, and 2) an increase in the proportion of ESG-conscious investors causes the credit spread of high-ESG-score companies to decrease more than that of low-ESG-score companies.

We find preliminary evidence of the first testable implication in the cross-section of dollar-denominated emerging markets' corporate bonds. For this purpose, we regress secondary markets' bond spreads on company ESG scores with control variables measuring systematic, liquidity, and default risks. We use ESG scores from Sustainalytics and RepRisk that are available to investors in real time through J.P. Morgan bond indices, which contrast with other ESG scores, such as Refinitiv, that have been revised over time (Berg et al., 2020).

Using this cross-sectional regression methodology, we find an average ESG premium, which is the estimated coefficient associated to the ESG score variable, of 18 basis points (bp) over the April 2018–March 2022 period, originating mainly from the within-sector variation. In terms of magnitude, given that the ESG score scale goes from 0 to 5, all else being equal, a bond issued by the best ESG company, i.e., with a score of 5, has a spread that is approximately 90 bp (= 18 bp $\times$ 5) lower than the worst ESG-rated issuer, who has a score of 0. Distinguishing between the pre-COVID and post-COVID periods, we find that the premium has increased from 11 bp pre-COVID to 19 bp post-COVID, suggesting that either the proportion of ESG-conscious investors, and/or the taste for ESG assets, has increased. This is consistent with the sharp increase in flows toward sustainable funds starting in 2020 and continuing in 2021, as well as the interest from funds themselves to become more ESG aware following the introduction of the Sustainable Finance Disclosure Regulation (SFDR). The SFDR is a European initiative that came into law in March

[5] Our theoretical framework is mostly inspired by Baker et al. (2018) and Pastor et al. (2021b).

2021, that aims to provide a more transparent approach on how financial products disclose their ESG impact. Finally, we also find the presence of an ESG premium within China, a market that we will use to better identify the presence of a preference-driven ESG premium.

The results obtained from this cross-sectional regression are consistent with the existence of a preference-driven ESG premium, but the results do not fully distinguish between the fundamental and preference channel through which ESG scores impact equilibrium prices. Indeed, the cross-sectional approach might fail to control for all confounding factors simultaneously affecting ESG scores and credit spreads. In fact, ESG scores are correlated with several of our observed risk controls, including the bond's credit rating, beta, size, and past volatility, as we show empirically. Furthermore, our risk controls only explain between 13.3% and 20.2% of the cross-sectional variation of ESG scores, suggesting that there is considerable remaining orthogonal information in ESG scores that might be impacting both firms' fundamentals and investors' preferences.

To better isolate the impact of the preference channel from the fundamental channel, we propose an identification strategy that uses the opening of the Chinese onshore bond market as an instrument. The opening of the Chinese onshore bond market in 2019 lead international investors from the United States and Europe, who are more ESG-conscious, to increase their holdings in Chinese onshore corporate bonds. We test whether this increase in the proportion of ESG-conscious investors differently affected high-ESG bonds and low-ESG bonds, as predicted by the preference channel. Because the spread of high-ESG-score companies might not have evolved in parallel with the spread of low-ESG-score companies, we leverage the fact that several firms in China are issuing bonds both in *onshore* as well as *offshore* markets. Given that offshore markets did not experience the same regulatory changes, we can use offshore bonds to control for the issuer's time-varying credit risk in a triple difference framework with issuer-time fixed effects. We find that following the opening of the Chinese onshore market, a company with the highest possible ESG score would have seen its cost of borrowing decrease by 48 bp more than a company with the lowest ESG score.

We provide evidence to support the validity of our identification strategy. Our key assumption is that the proportion of ESG-conscious investors increased in the Chinese onshore bond market after its opening, more so than in the offshore market. To examine this assumption, we first show that foreign net capital inflows toward Chinese onshore corporate bonds increased almost fourfold from \$1.1 billions per quarter before the regulation change (2018Q1 to 2019Q3) to \$3.9 billions per quarter post regulation (2019Q3 to 2020Q2). Because international investors are more ESG conscious (Martin and Moser, 2016; Riedl and Smeets, 2017; Hartzmark and Sussman, 2019; Ilhan et al., 2021; Bauer et al., 2021), an increase in the proportion of international investors in China *de facto* increased the proportion of ESG-conscious investors. Second, several local reports from the studied time period demonstrated that onshore investors did not pay much attention to ESG investing before the opening of the onshore market.[6] Conversely, local reports and surveys published after the opening indicated that onshore investors paid more attention to ESG-related issues, and the reason for doing so was international investors' demand.[7] Third, we provide evidence that offshore bonds, which are used to control for issuers' time-varying credit risks, did not experience changes in their investors' base during that same period. Finally, we are not able to reject the null hypothesis of a parallel trend between the onshore-minus-offshore difference in bonds' credit spreads of high-ESG companies and that of low-ESG companies in the months that preceded the opening of capital market.

Taken together, the existence of preference-driven ESG premium, which we provide evidence for, has important implications for all market participants. First, investors and funds should know that sustainable investing in emerging markets should yield lower expected returns going forward if the proportion of ESG-conscious investors and their ESG taste stay constant. However, as our results in the Chinese onshore market show, an increase in the proportion of ESG-conscious investors, or ESG taste, will decrease the spread of high-ESG-score companies more than low-ESG-score companies. This transition dynamic will provide existing investors in high-ESG-score com-

[6]The report from the Asset Management Association of China (AMAC) in 2017 uncovered that only 6% of the sampled asset managers took ESG into account when making investment decisions. Additionally, a 2018 CFA survey of financial professionals in China revealed that only 9% of the respondents included ESG in their credit analysis.

[7]The AMAC survey conducted in 2021 showed that approximately 50% of the respondent mutual funds incorporated green investing into their investment decisions. Furthermore, the international-investor-driven ESG demand is well acknowledged by several reports, ranging from local brokers (PingAn, 2020) to foreign brokers (J.P.Morgan, 2021), to international organizations (PRI, 2020)

panies with larger gains from the price appreciation that runs in line with spread compression. Second, debt issuers in emerging markets should strongly consider ESG-improving policies as part of their expenses, given the significant impact of ESG scores on their cost of debt, especially given that most of the decrease in the cost of debt occurs thanks to the improvement in ESG scores *within* sector. Finally, our paper has implications for ESG score providers. Our evidence suggests that companies that are awarded better scores by ESG score providers benefit from a lower cost of capital, especially as the demand for ESG increases. Thus, it is their duty to ensure that their scores properly reflect good ESG practices as opposed to simply greenwashing.

1.1.1 Related literature

Our paper is mainly related to the literature identifying the existence of a preference-induced ESG premium in the bond market. Recent papers have studied the existence of a *greenium,* which refers to the yield difference between a *green* bond, which is meant to raise money for environmental projects, and a conventional bond from the same issuer (Ehlers and Packer, 2017; Karpf and Mandel, 2018; Baker et al., 2018; Hachenberg and Schiereck, 2018; Zerbib, 2019; Lau et al., 2022). Finding a greenium is clean evidence of a preference channel because looking at two otherwise identical bonds from the same issuer ensures to compare bonds with identical risks, hence reducing considerably the influence of the fundamental channel. However, this approach might not capture the real ESG preference of investors because money is fungible. Indeed, ESG-conscious investors are probably not getting much utility from investing in the green bond of a company known for human rights violations. This is because ultimately, the money used to reimburse the debt of the company comes from the entire business. Our main distinction from this stream of the literature is to empirically identify the existence of a preference-driven ESG premium *across* issuers, instead of ESG investments *within* the same issuer.

Several papers have looked at the across-issuer ESG premium, mainly in developed fixed-income markets such as the United States and Europe. We show in Table 1.1 24 papers that analyze links between ESG-related attributes and firms' costs of debt. Among those papers that include emerging markets in their analysis, we can cite Ferriani (2022), who looks at global issuance during the COVID crisis, Hoepner et al. (2016), who investigate bank loans in 28 countries, and Kim et al. (2014), who analyze syndicated loans in 19 countries. Overall, the average ESG premium found in these paper is around 18.6 bp, which means that a company with a one standard deviation higher ESG score can borrow at a 18.6 bp lower interest rate on average.[8] This ESG premium, however, is the product of both the fundamental and preference channel. We contribute to this literature by presenting evidence of the preference channel behind the ESG premium.

Our work also relates to the literature on the relation between firms' ESG scores and their cost of equity capital (stocks' expected returns), though our results are not directly applicable here. Our results can overstate or understate the effect for equity markets. On the one hand, bond holders can have a larger influence on issuers, especially given their need to refinance their debt frequently (Oikonomou et al., 2014). The payoff profiles are also different; downside risks are more of interest for bond holders compared with upside potential. An extensive amount of empirical research has typically explored the direct relation between ESG scores and firms' equity realized returns or cost of capital. The results are mixed. Hong and Kacperczyk (2009) find that "sin" stocks (stocks with low ESG scores) outperform non-sin stocks. Similarly, Bolton and Kacperczyk (2021) find that stocks of firms with higher CO_2 emissions earn higher average returns and interpret their findings through investors demanding a premium to hedge carbon risk. Other studies find opposite results – firms with high ESG scores (green assets) have higher expected returns than those with low ESG scores (brown assets). In particular, Edmans (2011) and Gompers et al. (2003) find that firms with higher employee satisfaction and strong shareholder rights, respectively, have higher stock returns. Kempf and Osthoff (2007) document that stocks of firms with higher ESG scores in the 1992–2004 period also outperformed.

Our paper is related to the theoretical literature aiming to explain the documented empirical relation between ESG scores and asset returns. Pedersen et al. (2021) develop a theory in which a firm's ESG score can both predict the firm's fundamentals and affect investor preference. Their

[8]This number is an imperfect approximation given the difficulty in aggregating the different methodologies and ESG scores used in the papers.

model implies that an ESG-efficient frontier and green assets can either out- or underperform brown assets, depending on the compositions of different types of ESG-aware investors in the market. Pastor et al. (2021b) develop a model in which investors have different levels of tastes for green assets and in which the dispersion in tastes create an ESG premium in equilibrium. Their model predicts higher expected returns for brown assets in the long run; however, green assets might outperform when investors' tastes change in a short-period sample. Both studies highlight the importance of heterogeneity in terms of investor preference (tastes) in pricing green vs. brown assets. Goldstein et al. (2022) propose a rational expectation equilibrium model with green and traditional investors, with interesting implications regarding the increase in the share of green investors. Provided that the share of green investors is relatively small, they show that an increase in the share of ESG-conscious investors increases the expected returns of high-ESG-score companies because there is lower price informativeness as the result of a more diverse investor base. Our work builds on this stream of the literature by providing an empirical estimate of the direction and magnitude of the impact of investor preferences on the expected return of assets.

Finally, this paper relates to the literature that shedding light on the large Chinese onshore bond market. Amstad and He (2019) provide a comprehensive overview of the large and yet understudied market. Ding et al. (2022) document the overpricing of the issuance of Chinese corporate bonds. Ang et al. (2018) analyze the pricing of the Chinese municipal bond market, which is considered to be corporate bonds in China. We contribute to the literature by studying the impact of the opening up of the market, which only began in 2017, on bond pricing.

1.2 Theoretical Framework and Testable Predictions

We develop a simple theoretical framework to guide our empirical research design. Our model shows that in equilibrium, all else being equal, 1) firms with higher ESG scores have lower yields and 2) an increase in the demand for ESG assets decreases the yield of high-ESG-score firms more than that of low-ESG-score firms. We focus on bond yield because it provides a less noisy measure of expected returns, especially over a short period. Proxying expected returns with realized returns is problematic, especially when studying ESG scores. First, ESG scores can correlate with events that occur at a very low frequency (e.g., natural disasters, social unrest, and government scandals, ...), so realized returns do not capture these events in a small sample. This is the traditional Peso problem. Second, the past few years have seen an increase in demand for ESG assets, creating equilibrium transition dynamics (Pastor et al., 2021a). Suppose high-ESG assets require lower expected returns in equilibrium, as in our model. As high-ESG assets require gradually less expected returns, their price increase, resulting in high-ESG-score firms posting higher realized returns.

1.2.1 Model specifications

We approximate the yield of corporate bonds as the sum of the bond's expected return and the expected loss, similar to Nozawa (2017). Let y be the vector of bond yields, μ the vector of expected returns, and l the vector of expected losses for all bonds in the economy. We have

$$y = \mu + l, \tag{1.1}$$

where for economy of notation, we assume the risk-free rate to be zero, $r_f = 0$. Similarly to Baker et al. (2018), we model expected returns μ from a standard asset pricing framework with two groups of investors: ESG-indifferent investors with total capital of I and ESG-conscious investors with a total amount of capital C. Each group of investors has identical risk aversion and identical expectations for the bond's return vector μ and the variance-covariance matrix Σ. Additionally, both ESG-indifferent and ESG-conscious investors have additional preferences for a vector of bond characteristics g and h, respectively. This feature of the model is motivated by the literature on demand-driven asset pricing, where investors have specific demand for certain characteristics (Koijen and Yogo, 2019). For instance, pension funds and insurance companies tend to prefer investment-grade bonds due to regulatory constraints (Ellul et al., 2011), while mutual bond funds prefer more liquid bonds (Choi et al., 2020). Finally, both groups face a mean-variance optimization problem, but ESG-conscious investors have an additional constraint requiring a minimum

ESG score for their portfolio. Formally,

$$
\begin{aligned}
\text{ESG-Indifferent Investors (I): } &\max_{w_I} w_I'\mu + w_I'g - \frac{\gamma}{2} w_I'\Sigma w_I \\
\text{ESG-Conscious Investors (C): } &\max_{w_E} w_C'\mu + w_I'h - \frac{\gamma}{2} w_C'\Sigma w_C \\
&s.t. w_C'z \geq \bar{z},
\end{aligned}
$$

where z is a vector of ESG scores and $\bar{z}$ is the minimum weighted average ESG score that ESG-conscious investors require. The additional constraint for ESG-conscious investors has micro-foundations in the asset management industry, where this type of constraint is frequent.[9] Denoting the Lagrangian multiplier for ESG-conscious investors by λ, and the vector of deviation of ESG score versus the minimum threshold $\bar{z}$ by $\tilde{z} = z - \bar{z}$, we have the following two first-order conditions:

$$w_I = \frac{1}{\gamma} \Sigma^{-1} (\mu + g) \tag{1.2}$$

$$w_C = \frac{1}{\gamma} \Sigma^{-1} (\mu + h + \lambda \tilde{z}) . \tag{1.3}$$

The market-clearing condition imposes that the vector of market weights w_m is

$$w_m = (1 - c) w_I + c w_C, \tag{1.4}$$

where c is the proportion of ESG-conscious investors $c = C/(C + I)$.

Finally, substituting equations (1.2) and (1.3) in the market-clearing condition equation (1.4), we get

$$\mu = \gamma \Sigma w_m - (1 - c) g - ch - c\lambda \tilde{z}.$$

[9]For instance, the Carmignac Portfolio Prospectus writes “Average exposure weighted score is above 3 making a positive contribution to society and the environment”.

We can apply this formula to the market portfolio $\mu_m = w_m' \mu$:

$$\mu_m = \gamma \sigma_m^2 - (1-c)\, w_m' g - c w_m' h - c\lambda w_m' \tilde{z}.$$

We notice that the market risk premium is higher if the market portfolio's average ESG deviations are negative, as ESG-conscious investors demand compensation for holding a portfolio that is below their required ESG level $\bar{z}$. Similar argument can be applied to the characteristics g and h. For simplicity, we assume that the market portfolio is ESG- and characteristic-neutral, such that the weighted-average ESG score exactly equals the threshold level $\bar{z}$, and $w_m' g = w_m' h = 0$. We can then express the expected return vector as

$$\mu = \beta \mu_m - (1-c)\, g - c h - c\lambda \tilde{z},$$

where $\beta = \frac{\Sigma w_m}{\sigma_m^2}$.

Substituting μ in the original yield equation (1.1), we can express the yield of any bond i as

$$y_i = \beta_i \mu_m - (1-c)\, g_i - c h_i - \lambda c \tilde{z}_i + l_i. \tag{1.5}$$

1.2.2 Testable implications

Given equation (1.5), we can write two testable implications from the model:

Testable Implication 1: *Given same default risk l, systematic risk β, and characteristics g and h, high-ESG bonds have lower yields than low-ESG bonds, with the difference proportional to λc.*

Testable Implication 2: *Given same default risk l, systematic risk β, and characteristics g and h, when the proportion of ESG-conscious investors increases, the equilibrium yield of high-ESG bonds decreases more than that of low-ESG bonds, with a difference proportional to $\lambda \tilde{z}_i$.*

The main challenge in testing either of these implications in the data is to control for l_i, β_i, related to the fundamental channel, and for g_i, and h_i, related to other preferences of investors. Before we present our empirical strategy to overcome these challenges, we briefly review alternative models with different implications.

1.2.3 Discussion on alternative models

There has been a growing theoretical literature linking expected returns to a company's ESG scores. In most models, the existence of ESG-conscious investors leads to the same testable implications of the model presented in section 1.2.1, where high-ESG-companies have lower expected returns. This is the case for Heinkel et al. (2001), Luo and Balvers (2017), Baker et al. (2018), Pastor et al. (2021b), and Berk and van Binsbergen (2021) for instance. There are, however, two recent papers we would like to mention because of their ability to generate different relationships between a firm's ESG score and its expected return.

The first paper is Pedersen et al. (2021), in which the authors develop an equilibrium asset pricing model with three types of investors, leading to a modified CAPM similar to that shown in the previous section. Their model exhibits a major difference with ours. Instead of having all investors use all information to forecast expected returns and variance-covariances, as in our model, they include a third category of investors, "ESG-uninterested" investors, who do not derive utility from holding high-ESG-score companies and do not use ESG information at all despite its forecasting power for firms' future profits. As a result, when ESG scores have a sufficiently high correlation with future cash flows, high-ESG-score companies can have higher expected return if market participants do not fully account for this signaling power of ESG scores. This is the case with a high-enough proportion of ESG-uninterested investors.

The second paper is Goldstein et al. (2022), which proposes a rational expectation equilibrium model with green and traditional investors. This model generates a non-monotonic relation between ESG score and expected return. On one hand, high-ESG-score companies, with high expected non-monetary output in their model, have a lower cost of capital because ESG-conscious investors are willing to pay more to hold them. On the other hand, high-ESG-score companies have a more diverse investor base, composed of both ESG-conscious and ESG-indifferent investors, which reduce the company's price informativeness and increase the required rate of return. Their model has interesting implications regarding the increase in the share of green investors, which refers to testing implication 2 in previous section. Provided the share of green investors is relatively small, they show that an increase in the share of ESG-conscious investors increases the expected return of high-ESG-score companies, because the lower price informativeness due to a more diverse investor base dominates the price premium that ESG-conscious investors are willing to pay. Conversely, when the share of ESG-conscious investors is large, an increase in the proportion of ESG-conscious investors decreases the expected return. Finally, their models offer additional testable predictions of the effect of the share of ESG-conscious investors on price informativeness, price volatility, and trading volumes. We do not test these predictions in this paper and choose to focus on the cost of capital.

Given the alternative predictions from these models, not being able to reject testable implications 1 and 2 in our data indicate the validity of the different models.

1.3 Empirical Framework

This section presents our empirical framework to test implications 1 and 2 of our model in emerging markets' corporate bonds.

1.3.1 Preliminary evidence of ESG premium across emerging markets

We first test implication 1 of our model across USD-denominated emerging markets corporate bonds. Similar to Crifo et al., 2017, Baker et al., 2018, Ghouma et al., 2018, and Ben Slimane et al., 2019, we regress secondary market bonds' spreads onto companies' ESG scores with control variables proxying for systematic, liquidity, and default risks. Specifically, we run the following regression for bond i, company s, at time t:

$$s_{i,s,t} = \sum_{\tau=1}^{T} \alpha'_\tau D_{i,s,\tau} + \gamma ESG_{s,t} + \delta' X_{i,s,t} + \varepsilon_{i,s,t}, \tag{1.6}$$

where $s_{i,s,t}$ indicates the bond spread with respect to the duration-matched U.S. treasury yield and γ indicates the coefficient of interest. The null hypothesis is $H_0 : \gamma \geq 0$, against the alternative of the existence of a preference-induced ESG premium $H_1 : \gamma < 0$, as in the model presented in the previous section. Our controls include time fixed effects, α'_τ, which are specific to issuers' and bonds' characteristics contained in the vector $D_{i,s,t}$, to control for liquidity and default risks. Specifically, $D_{i,s,t}$ includes 17 bonds' rating dummies (from AAA to CCC), coupon types (fixed, floating, step, variable, zero), a callable indicator, an ownership indicator (corporate or quasi-sovereign), payment-rank types (1st/2nd lien, Secured, Senior Unsecured, Senior/Junior Subordinated, Senior Unsecured, Unsecured), a dummy equal to 1 if the bond's amount outstanding is above \$500M, and sector dummy variables.[10] $ESG_{s,t}$ corresponds to the company's ESG score, a continuous variable ranging from 0 (worst score) to 5 (best score). Finally, $X_{i,s,t}$ contains an additional set of variables to control for liquidity risk (bid-ask spread), systematic risk (bond's beta to the EM corporate index), and default risk (duration, ex-ante default probability). All these control variables also capture alternative bond preferences other than ESG, e.g., liquidity (Choi et al., 2020) and credit ratings (Ellul et al., 2011).

Identification challenges

We acknowledge that rejecting $H_0 : \gamma \geq 0$ in regression (1.6) is consistent with the existence of a preference-induced ESG premium but is not conclusive evidence. There could exist several unobserved firm attributes correlated with both the ESG score of the company and its fundamental risks, leading to the classical omitted variable bias. This bias can be positive or negative. On

[10] We choose \$500M as the threshold for the bond's amount outstanding dummy variable because J.P. Morgan has another index, the CEMBI, which includes "large" bonds with minimum \$500M in amount outstanding. The CEMBI Broad contains all bonds with at least \$300M in amount outstanding. The sectors are Communications, Consumer Discretionary, Consumer Staples, Energy, Financials, Government, Health Care, Industrials, Materials, Technology, and Utilities.

one hand, high-ESG-score companies could be less exposed to governance scandals or climate regulations, thus being less risky and enjoying a lower yield. Similarly, a high-ESG score could attract better employees, which would add value to the firm, decreasing its cost of borrowing. Additionally, good governance would lead to better decisions and a more profitable firm, again decreasing the firm's yield. In all these cases, the bias would be negative, i.e., toward finding an ESG premium ($\hat{\gamma} < \gamma$). On the other hand, large ESG-related expenses to improve the ESG score, with management spending money on charities that are not profit-enhancing for instance, could decrease the profitability of the firm, increasing its yield. In this case, the bias would be positive, toward not finding a preference-driven ESG premium ($\hat{\gamma} > \gamma$). All these confounding factors might not be captured by our controls. Additionally, one can think of other endogeneity concerns. Large and profitable firms are more capable of improving their ESG score, which leads to a simultaneous equation problem, rendering γ not identifiable in any cross-sectional OLS regression specification.

Although there is no way to test the presence of an omitted variable bias, we can still investigate whether ESG scores correlate with ex-post measures of risk in our sample, before and after including our controls for systematic, liquidity, and default risks. At least if we were to find that ESG scores predict ex-post measures of risks beyond our control variables, our empirical framework would be proven invalid from the get go. First, we look at the predicting power of ESG scores for the probability of default. To do so, we regress a dummy variable equal to one if the bond default in any of the next 12 months on ESG scores and additional controls. We repeat this exercise for the downgrade probability of investment-grade bonds, where the dependent variable is a dummy equal to one if the bond drops from investment-grade to high-yield at any point in the next 12 months.

1.3.2 ESG premium following change in investors' ESG consciousness

We test the existence of a preference-driven ESG premium using the second testable implication of our model: a positive shock to investors ESG preferences causes the yield to decrease more for high-ESG-score companies versus low-ESG-score companies.

Our research design takes advantage of the unique setting of how Chinese bonds are traded and the historical opening of the Chinese bond market. The setting allows us to observe the yield of bonds issued by the same company but traded in different markets: local Chinese markets (onshore) and international markets (offshore). Two bonds issued by the same firm have similar credit risks but different investor base if one bond is traded onshore (CNY-denominated) and the other offshore (USD-denominated). Furthermore, the opening of the onshore market leads to a shift in investor composition in the onshore market, corresponding to an increase in the share of international investors.

This change in the investor base allows us to test the existence of an ESG premium, while controlling for the firm time-varying credit risks. We start by providing a brief overview of the institutional background on the Chinese bond market and its historical opening. We then present our triple-difference design and discuss the identifying assumptions.

Shock in the composition of investors in the Chinese onshore market

Despite the tremendous size of the market, the Chinese onshore bond market has largely remained isolated from international investors. However, the government has recently changed its regulations to open up the market, and a growing number of foreign investors have started to participate in the market. The regulation changes thus present a quasi-experiment to study the impact of investors demand on the ESG premium.

Important changes include the introduction of Bond Connect in July 2017, which facilitated access for international investors to the China Interbank Market through the connection between infrastructure providers in Hong Kong and mainland China ("northbound" connection). The main consequence of this change is that it helps foreign investors access the interbank market, which mostly contained enterprise and sovereign bonds.

The second event occurred in September 2019, which is our treatment date in the analysis. In September 2019, regulators announced the removal of lock-up periods, allowing hedging of foreign exchange for Qualified Foreign Institutional Investors (QFIIs) and increasing their investment quota from \$150million to \$300million. The State Administration of Foreign Exchange (SAFE) also announced that it would repeal the QFII investment quotas in the near future. At the same time, the J.P. Morgan index announced that it would add Chinese government bonds to its index for February 2020.[11] Finally, international rating agencies, such as Standard & Poor's, Moody's, and Fitch, which were monitoring Chinese onshore bonds but barred from releasing official ratings, were authorized to rate Chinese issuers in 2019. All these events have led to an increase in net capital inflows to corporate bonds from offshore investors, as shown in Figure 1.3, which reports quarterly changes in holdings of onshore enterprise bonds and medium-term notes held by foreign institutions.[12] The average of net capital inflows increased almost fourfold from $1.1B per quarter before the regulation change (2018Q1 to 2019Q3) to $3.9B per quarter post-regulation (2019Q3 to 2020Q2).

We argue that the opening of the capital market has lead to an increase in ESG-conscious investors in the Chinese onshore market relative to the offshore market for several reasons.

First, onshore investors did not pay much attention to ESG investing before the opening of the onshore market. For instance, the report from the Asset Management Association of China (AMAC) in 2017 uncovered that only 6% of the sampled asset managers took ESG into account when making investment decisions (AMAC, 2017). Additionally, a 2018 CFA survey of financial professionals in China revealed that only 9% of the respondents included ESG in their credit analysis (CFA, 2019). Thus, we can reasonably state that the proportion of ESG-conscious investors could not have *decreased* following the market opening, given the extremely low starting point.

Second, the increase in the proportion of offshore investors has likely increased the proportion of ESG-conscious investors. There is considerable literature documenting the ESG-consciousness of U.S. and European investors (Martin and Moser, 2016; Riedl and Smeets, 2017; Hartzmark and Sussman, 2019; Ilhan et al., 2021; Bauer et al., 2021). This fact is also directly documented in local Chinese surveys; when asked about where the ESG demand comes from, all three interviewees from the CFA survey indicated that the demand mainly comes from international investors, espe-

[11] Reuters article of September 4th, 2019: https://www.reuters.com/article/china-markets-bonds/update-2-jpmorgan-adds-china-to-emerging-bond-index-from-february-2020-idUSL5N25V3F4

[12] The data are obtained from the China Central Depository & Clearing (CCDC) and the Shanghai Clearing House (SHCH).

cially the United States and Europe (CFA, 2019). Furthermore, this international-investor-driven ESG demand is well acknowledged by several reports, ranging from local brokers (PingAn, 2020) to foreign brokers (J.P.Morgan, 2021), to international organizations (PRI, 2020).

Third, onshore investors have also become more ESG-conscious following the capital market opening. For example, the AMAC survey conducted in 2021 showed that approximately 50% of the respondent mutual funds incorporated green investing into their investment decisions.

Finally, we also provide evidence that offshore bonds did not experience dramatic changes in their investors' base during that period. Offshore bonds are Euroclearable and can be traded by all investors in the world without the need for local accounts. Nothing in this regard changed in 2019-2020. Additionally, we also look at an informal measure of institutional investors' positioning in offshore Chinese corporate bonds. This positioning is surveyed monthly by J.P. Morgan, which started in 2001 and covered around 140 institutional investors in emerging markets with a total of one trillion USD in assets in 2019. We explain in Appendix A.3 the details of the methodology used to compute this positioning. Figure 1.2 shows a relatively smooth trend with neutral positioning in corporate bonds and underweight in banks. Corporate bond positioning started at approximately -1% (i.e., underweight) in January 2019 and was approximately 0% during the months around September 2019, to finally end up at approximately 0.5% in July 2020. For banks, it stayed at approximately -2% throughout January 2019 to July 2020. More importantly, we did not see any sharp increase on the date of the Chinese market opening, as opposed to the net inflows pattern shown in figure 1.3.

Empirical strategy: triple difference with continuous treatment

Given this shock to the investor's base in the onshore market, a natural setup would be to look at the difference in difference between onshore high-ESG bonds versus onshore low-ESG bonds, before and after the Chinese onshore market opening. Unfortunately, such specification will lead to a biased estimate of the difference-in-difference coefficient if there are time-varying unobserved characteristics that affect high-ESG bonds differently from low-ESG bonds; in other words, the parallel trend assumption is violated. Such violation can be attributed to different reasons. Credit ratings are quite sticky and sometimes non-existent for certain bonds. Hence, given that one cannot timely observe the issuer's default probability, and if this default probability is correlated with the ESG rating, the spread of high- and low-ESG bonds should evolve differently in the absence of the shock. We propose below a setting that can control for the issuer-specific time-varying effects. Since we observe both offshore and onshore bonds, we can include issuer-time fixed effects which will control for the time-varying issuer-specific default probability. Let us denote $s_{i,s,m,t}$ the spread of bond i, of issuer s, in market m, with $m \in \{\text{offshore,onshore}\}$, at end-of-month t. We can run the following triple difference regression:

$$s_{i,s,m,t} = \alpha_i + \alpha_{s,t} + \alpha_{m,t} + \gamma ESG_s \times Onshore_i \times Post_t + \delta' X_{i,s,m,t} + \varepsilon_{i,s,m,t}, \tag{1.7}$$

where $Onshore_i$ is a dummy that takes the value one if the bond is onshore, $Post_t$ is a dummy that takes the value one if $t \geq$ Sep 2019, and ESG_s is the ESG score of the company at the time of opening in September 2019. Our coefficient of interest is γ, which measures the ESG premium resulting from the increase in ESG consciousness of investors. The important feature in this specification is the control for issuer-time fixed effects, $\alpha_{s,t}$, capturing unobserved changes in the default probability of the issuer. The bond fixed effects α_i control for the time-invariant heterogeneity in the bonds, such as bond covenants and coupon type. We also include market-time fixed effects, $\alpha_{m,t}$, to control for common variations specific to the offshore and onshore markets (e.g., different Fed and PBoC policy rates). Finally, to rule out other preference channels, we include in $X_{i,s,m,t}$ several controls that might affect the investment decision of international investors, including liquidity (Choi et al., 2020) and credit ratings (Ellul et al., 2011). Specifically, we include several interaction terms allowing for different trends between onshore-minus-offshore spreads of high- and low-ESG issuers. The triple interaction terms are $SOE_s \times Onshore_i \times Post_t$, with the dummy SOE_s controlling for state-owned enterprise (SOE) status of the issuer; $\log(Amount\ Issued_i) \times Onshore_i \times Post_t$, controlling for the size of the issuance; $CPN\ Type_i \times Onshore_i \times Post_t$, controlling for the coupon type (floating, fixed, step, or variable coupon); and $Rating\ category_{s,m,t} \times Onshore_i \times Post_t$, controlling for potential differences in the credit rating of onshore and offshore ratings of the issuer, although this is mostly controlled for by the issuer-time fixed effect $\alpha_{s,t}$.[13] We also include all single and double interaction terms that are not subsumed by the fixed effects for the completeness of the model. Finally, We also include the $Duration_{i,t}$ of the bond and $Duration_{i,t} \times Onshore_i$ to allow for different relations in the onshore market.

Identifying assumptions

Parallel trends

The standard identifying assumption for the triple difference estimator is the assumption of a parallel trend between the onshore-minus-offshore spread of high-ESG issuers and that of low-ESG issuers. In other words, in the absence of the onshore market opening, there should be no other factors differently affecting the path of onshore-minus-offshore spreads of high-ESG issuers versus low-ESG issuers. We test the validity of this assumption with a Granger causality test, estimating the following regression:

$$y_{i,s,m,t} = \alpha_i + \alpha_{s,t} + \alpha_{m,t} + \sum_{\tau} \gamma_\tau ESG_s \times Onshore_i \times D_t^\tau + \delta' X_{i,s,m,t} + \varepsilon_{i,s,m,t}, \tag{1.8}$$

where $\tau \in$ {February 2019, ...,February 2020} and D_t^τ is a dummy variable taking value one if $\tau = t$, and zero otherwise. The base month is January 2019, which is the first month of the year and further enough from the announcement date of September 2019. Support for the parallel trend assumption should come from not being able to reject the null hypothesis, $H_0 : \lambda_\tau = 0, \forall \tau <$ September 2019.

[13]We download the end-of-month time series of onshore and offshore ratings for each issuer from Bloomberg. We aggregate the ratings from Moody's, Fitch, and Standard & Poor's into one by taking the median if all three are available or the worst rating if less than three are available. The Bloomberg Tickers for onshore rating are RTG_MDY_LC_CURR_ISSUER_RATING, RTG_FITCH_LT_LC_ISSUER_DEFAULT, and RTG_SP_LT_LC_ISSUER_CREDIT, for Moody's, Standard & Poor's, and Fitch, respectively. For the offshore ratings, the tickers are RTG_MDY_FC_CURR_ISSUER_RATING, RTG_SP_LT_LC_ISSUER_CREDIT, and RTG_FITCH_LT_ISSUER_DEFAULT, for Moody's, Standard & Poor's, and Fitch, respectively. We also include a category "not rated" for issuers for which no ratings are available.

Anticipation of the shock

To measure the full causal effect, the regulatory opening should not be anticipated by market participants. We choose September 2019 because it corresponds to a new announcement by the Chinese government regarding quota changes. Of course, it is likely that some market participants anticipated the communication from Chinese regulators given the introduction of Bond Connect in July 2017, signaling the government's willingness to open its capital market to foreign investors. However, any anticipation would actually work against finding an ESG premium, as local investors would have already started rotating toward higher ESG bonds in expectation of increased demand from international investors. The Granger test of equation 1.8 allows us to test whether the difference started already before the change in regulation.

1.4 Data

Before presenting the results, we review the different data sources and construction methodology used in the analysis.

Emerging markets offshore bonds

Offshore USD-denominated bonds come from the EM Global Diversified Bond Index (EMBI) and Corporate EM Broad Diversified Bond Index (CEMBI), two flagship indices maintained by J.P. Morgan. The bonds included in the index have several features ensuring tradability for international investors. All bonds must be fixed-income instruments (fixed, floating, amortizing, or capitalizing) with at least six-month to maturity, not defaulted, internationally clearable, and with a minimum issue size of \$300 million (\$500 million for the EMBI index). We consider all corporate bonds, including bonds guaranteed by federal governments (quasi-sovereigns) but excluding those directly issued by national governments (sovereigns).

Price-related bond characteristics, such as spread, yield, and duration, are directly obtained through PricingDirect, a subsidiary of J.P. Morgan, which is a service that provides pricing based on broker-dealer quotes and everyday transactions.

Table 1.2 summarizes the statistics for the entire sample of emerging markets corporate bonds used in the analysis of the ESG premium across emerging markets (section 1.3.1). Panel a) exhibits the mean, standard deviation, and main percentiles of the variables of interest, which include the amount issued, ex-ante probability of default, spread, duration, duration-times-spread (DTS), Beta, ex-ante volatility, "default next month" dummy, and "downgrade next month" dummy. We compute betas with respect to the CEMBI Broad index on a rolling basis with weekly data over the past six months. Using a shorter period than the usual three years in the literature allows us to have more bonds for the analysis. We compute the issuers' ex-ante default probability from Bloomberg, which uses several financial ratios and the firms' estimated volatility to compute a distance-to-default, ultimately converted into a probability of default for the next five years at the issuer level.[14] In panel b), we show the distribution of bonds across rating categories, sectors, and regions, where the percentage refers to the number of bonds divided by the total number of bonds (as opposed to market capitalization). Finally, panel c) shows the number of bonds and issuers we have per region. In summary, our sample comprises over 1,046 issuers and 3,250 bonds. The distribution of payment rank of each bond is as follows: 87.9% of bonds are senior unsecured, 4.1% subordinated, 4% 1st lien, 3.4% secured, 0.3% unsecured, 0.2% 2nd lien, and 0.1% junior subordinated. In terms of coupons, 90.2% of the bonds are fixed-coupon bonds, 7.1% variable, 2.5% floating, and 0.2% step- and zero-coupon. We winsorize at 1% all bond-specific variables that can take extreme values, such as spread, ex-ante probability of default, duration, DTS, and beta.

ESG scores

We use the company-level ESG scores computed by J.P. Morgan to construct its two flagship emerging markets' ESG bond indices: the JESG CEMBI Broad Diversified index and the JESG EMBI Global Diversified index. The ESG scores correspond to an equally weighted average of

[14] The Bloomberg Ticker is "BB_5Y_DEFAULT_PROB". For more details on the methodology, please refer to the Bloomberg Issuer Default Risk Model described in Bondioli et al. (2021).

the ESG scores of RepRisk and Sustainalytics, two major providers of ESG scores globally. We describe in Appendix A.1.1 and A.1.2 more details on how Sustainalytics and RepRisk construct their scores.

There are several advantages in using this combined Sustainalytics-RepRisk score, as opposed to the single scores obtained from other providers, such as MSCI or Refinitiv. The first advantage is the high coverage of Sustainalytics and RepRisk. Overall, 99.7% of the market capitalization of the indices is covered by at least one ESG provider. This high coverage is explained by the fact that J.P. Morgan directly notifies Sustainalytics and RepRisk about the new issues included in the index, incentivizing the ESG providers to cover the company in question. There is now over $40B dedicated asset under management benchmarked to these indices, and J.P. Morgan regularly engages with its clients about companies being included in or excluded from the index. Large asset managers, such as Blackrock, BNP Paribas, or MFS Investment Management, have dedicated funds benchmarked to one of these two indices. The second advantage is the longer history of the indices. The JESG indices were launched in April 2018. For comparison, the only index competitor for emerging markets was launched by a partnership between Bloomberg and MSCI in May 2021.[15] Finally, and most importantly, the existence of this index ensures that these scores have not been restated or revised ex-post, which is not always the case, as shown in Berg et al. (2020). These scores used in this paper have been distributed to investors every month since April 2018.

We show in figure 1.1, panel a), the distribution of ESG scores sorted in five categories: $[0, 1)$, $[1, 2)$, $[2, 3)$, $[3, 4)$, and $[4, 5]$. The first observation is that ESG scores are approximately normally distributed, with 40% scores in the $[2, 3)$ category and less than 5% in the tail categories $[0, 1)$ and $[4, 5]$. There is, however some negative skewness, with 32% in the $[2, 3)$ category compared to 19% in the $[3, 4)$ category. In Panel b), we plot the median ESG score deviation (versus the overall median) by sector, as illustrated by the filled gray bars. We also add the 25th and 75th percentiles, as illustrated by the error bars. The best sector is Technology, with an excess score of

[15]It is the Bloomberg Barclays MSCI EM USD Aggregate ESG Weighted Index.

0.7 versus the median, and quite low dispersion, from 0.25 to 1.15. The worst sector is Transport, with a negative excess ESG score of 0.6, followed by Government, with -0.5 excess score. The main composition of the Government is government-owned development banks that tend to finance activities in metals and mining with low ESG scores. Despite these differences, the ESG scores are far from being fully explained by the sectors. We see large and similar dispersion within all sectors, and the medians are not much different after all. Materials, Industrials, Energy, Consumers, and Real Estate all have median excess scores around zero.

Chinese onshore bonds

We collect onshore bond prices, yield-to-maturity, duration, amount outstanding, and volume traded from THS-iFind (THS), a major commercial data vendor for market practitioners in China. To ensure quality and consistency, we complement and compare the THS data with WIND, another popular data provider, used in Ding et al. (2022).[16] Furthermore, following the market structure of Chinese onshore bond markets, we include enterprise bonds, corporate bonds, financial bonds, medium-terms notes, and commercial papers.[17] Our original sample goes from January 2015 until September 2021, but we focus on a short period window around the regulation change in September 2019. Chinese onshore bonds are traded on exchanges in Shanghai or Shenzhen or on the interbank market (IB). Investors active in the IB are mostly institutional investors, while retail investors have only access to exchange markets. Foreign investors who gain access to the Chinese onshore bond market through BondConnect can only access IB. In contrast, foreign investors who obtain a QFII license can access both the IB and exchange markets. Therefore, we collect bond information on both IB and exchange markets (Shanghai and Shenzhen) when applicable, to maximize the matching of onshore and offshore bonds. Bonds that trade both on IB and in exchanges have different identifiers in our database. We provide additional institutional background about the

[16]The reason we chose THS over WIND as our main data source is the better coverage of THS. For example, THS reports 6,647 enterprise bonds and the corresponding number for WIND is 5,152. Furthermore, THS reports 10,567 financial bond tickers, with 4,672 tickers being additional issuance and subsequent offering. Meanwhile, WIND classification reports 2,718 financial bonds.

[17]We exclude municipal bonds, asset-backed securities, or private placement notes.

types of onshore bonds included in this study in Appendix A.2. The price and returns of bonds traded on IB are based on actual transactions, reported by dealers to the National Association of Financial Market Institutional Investors (NAFMII).[18] Based on these reported transactions, THS collects the dirty prices of the last transaction each day as daily prices. Furthermore, THS also reports the low, high, last, average clean price, accrued interest, corresponding yield to maturity, volume traded for each bond each day, and computes price-related measures, such as bond yields and duration. If no transaction occurs for a particular bond, THS uses the price of the previous transaction. We compute the bond spread as the difference between the corporate bond's yield and the duration-equal yield of the Chinese local sovereign bond yield curve. We estimate the Chinese government curve at the end of each month by using the Nelson and Siegel (1987) model.

Matching Chinese onshore and offshore bonds

We match onshore and offshore bonds into one issuer by using the Bloomberg issuer identifier from each bond ISIN. We use the concept called "CAST_PARENT_ID", which identifies the ultimate parent of the company, and thus, allows us to match bonds that are issued by the same company but through different special vehicle subsidiaries. For instance, the China Evergrande Group directly issues USD-denominated bonds. There are other bonds issued by Evergrande Real Estate Group Ltd, a subsidiary. Using the company ticker would not match the bonds issued by these two entities, despite being the same company ultimately liable for all cash flow payments.

Once all bonds are sorted into issuers, we further require that offshore and onshore have at least $150 million amount issued, that the bonds within an issuer have amounts issued that are not more than 10 times apart between onshore and offshore bonds, that the difference in spread is at most 1% as of the treatment date in September 2019, that the time-to-maturity is at least 1.5 years, and the difference between onshore and offshore bonds' time-to-maturity is not more than four years.[19] We additionally require that as of the treatment date, onshore bonds have been traded

[18] This association is similar in its nature to FINRA in the United States.

[19] We use $150M as our threshold for the amount issued because it corresponds to the minimum amount required for bonds to be included in a flagship J.P. Morgan Asian index. We also require a maximum of 1% spread difference to avoid selecting bonds that could have large differences in their covenants. This concerns less than 10% of the sample

at least twice during the past six months and their bond yield correlations with offshore bonds had been at least 50% in the past 12 months. The results are robust to slightly different filtering parameters. Having too stringent restrictions reduces the number of issuers and, thus, the number of tests we can conduct, such as looking at within-sector ESG effect. After the merging process, we have a sample of 56 issuers with 323 onshore bonds and 197 offshore bonds for which we measure spread and yields from January 2019 to February 2020. We exclude the COVID-crisis because of the large spike in spreads across both onshore and offshore markets, with the lack of liquidity being a major driver for these movements. Table 1.3 presents the summary statistics of our merged sample. The average ESG score is 1.9, which is slightly worse than that in all emerging markets, which has an average of 2.2. The average yield is 3.8% for onshore bonds and 3.2% for offshore bonds, reflecting the higher interest rate in CNY than USD. Onshore bonds are, on average, larger, with approximately $2 billion amount issued compared to $670 million for onshore bonds. Onshore ratings are, on average, 5.9 (corresponding to A) for onshore and 6.8 for offshore bonds (corresponding to A-).[20] In terms of bond characteristics, 78.5% of bonds have fixed coupon bonds, 11.2% have variable coupons (a period of fixed coupon followed by floating), 9.7% have floating coupons, and 0.6% have step coupons (coupons that increase at threshold dates). In terms of sector distribution, 63.6% of issuers are Financials, 11.4% are Industrials, 9.1% are Utilities, 6.8% are Consumer Discretionary, 4.5% are Energy, 2.3% are Materials, and 2.3% are Government-linked. Additionally, 25% of the issuers are SOEs, and 75% are private. Finally, 97.8% of the bonds are traded on the IB market, while 2.2% are traded on exchange markets (Shanghai and Shenzhen).

of matched bonds. The median spread difference between matched onshore and offshore bonds is approximately 0.25%. This difference is not necessarily equal to zero, given the absence of credit repo specials markets and single name default swap markets (Hill and Jia, 2021). Additionally, it can be difficult to perfectly hedge the exchange rate risk and duration risks between USD- and CNY-denominated bonds, as their cash flow schedule and duration can vary; liquidity on interest rate swap markets is low (Hill and Jia, 2021).

[20]The rating scale is as follows: 1 (AAA), 2 (AA+), 3 (AA), 4 (AA-), 5 (A+), 6 (A), 7 (A-), ... 16 (B-). We use the median rating of all three rating agencies (S&P, Fitch, and Moody's) when all are available, and the minimum if less than three ratings are available.

1.5 Results

1.5.1 ESG score and expected return

Table 1.4 reports the result of regression (1.6). We run in total six specifications; in the first two columns we estimate the ESG premium for the entire period from April 2018 to March 2022, without and with sector-time fixed effects, respectively. Across all emerging markets (panel a)), the estimated coefficients are between -18 bp and -19 bp independently of the sector fixed effects, which are highly significant, with two-way clustered standard error at time and issuer level around 3.7bp. The similar magnitude with and without sector-fixed effects suggests that the premium comes mainly from within-sector variation. In terms of magnitude, given that the ESG score scale goes from zero to five, all else being equal, a bond issued by the best ESG company has an average credit spread of approximately 90 bp (=18 bp $\times$ 5) lower than the worst ESG-rated issuer within the same sector. Although this number is large, it would be virtually impossible for a company to change its ESG rating that much. The average monthly change in ESG score is -0.009, with a standard deviation of 0.18. The 1st and 99th percentiles equal -0.70 and 0.48, respectively, which are far from five. To have a better point of comparison, we can look at the top 25% firms that improved their ESG cost the most in the past year. Their improvement is 0.20, which would translate into a decrease of 3.6bp.

We then test how the ESG premium varies between the pre-COVID and post-COVID periods. We find that the premium has increased from 11-12 bp pre-COVID to 20 bp post-COVID, suggesting that either the proportion of ESG-conscious investors or the taste for ESG assets has increased. This is consistent with the sharp increase in flows toward sustainable funds starting in 2020 and continuing in 2021, as well as the interest from funds themselves to become more ESG-aware following the introduction of the SFDR. Additionally, given the ESG index was launched in April 2018, it is likely that investors only started paying serious attention to these scores in 2019.

Given our paper's focus on China, we also estimate the ESG premium within China in panel b). We find a smaller but statistically significant ESG premium of approximately 10 bp for the whole period, 7 bp pre-COVID and 11 bp post-COVID. Again, the effect originates from the within-sector variation, given the similar coefficients with and without sector fixed effects.

Comparing our results with the literature

To put our estimate of the ESG premium in perspective, we survey the literature that precisely analyzes across-issuer variation of ESG-related attributes and cost of borrowing in fixed-income markets. We thus exclude studies analyzing stock market returns as well as those analyzing green bonds within the same firm. Table 1.1 presents quantitative estimates of the ESG premium (a negative number corresponds to the existence of an ESG premium). Over 75% of the studies use similar specification (i.e., some form of cross-sectional OLS regressions of spreads onto ESG-related measures and controls). Only three papers out of the 24 have some analysis including emerging markets; Ferriani (2022) looks at global issuance during the COVID-crisis, Hoepner et al. (2016) investigate bank loans in 28 countries, and Kim et al. (2014) analyze syndicated loans in 19 countries. All other papers are based on U.S., Canada, or European data. The average ESG premium across all 24 papers is 18.6 bp (corresponding to a coefficient of -18.6bp) and the median across all papers is 11.5bp, with a standard deviation of 30bp. These estimates are remarkably similar to our findings above. These might not identify the preference channel, however, suffering from the same biases as ours.

ESG rating and ex-post/ex-ante risk characteristics

The results based on the previous regression cannot properly identify the ESG premium, because they are subject to the biases exposed in section 1.3.1. As shown in proposition 1.2.2, it is crucial to control for the systematic risk and the default risk given their potential correlation between ESG scores. There is, however, no guarantee that our control variables fully capture these risks. Several of our observed controls for risk are correlated with ESG scores, suggesting a clear link between firms' fundamentals and ESG scores; in Table 1.6, we report the coefficients from a regression of ESG scores on issuer characteristics.[21] We find that the ESG scores negatively correlate with firms' credit ratings (high-ESG-score companies tend to be of better credit quality), positively correlate with bonds' beta, and negatively correlate with bond size and bonds' past volatility. Overall, issuer characteristics traditionally linked to systematic, default, and liquidity risks explain approximately 13.3% of the variation in ESG scores, which is slightly higher for the investment-grade segment, with R-squared of 20.2%. As a result, the ESG scores contain considerable orthogonal information.

We can, however, test whether ESG scores are correlated to measurable ex-post risks in our sample. In Table 1.5, we test through a linear probability model whether ESG scores help forecast the next 12-month issuers' default probability in the high-yield segment (panel a) and the next 12-month issuer's downgrade probability in the investment-grade segment (panel b). We find that an ESG score higher by one unit corresponds to a decrease of 0.06 percentage bonds of the default probability, although statistically not significant (t-stat of 1.5). When adding controls for credit rating, duration, ex-ante default probability, bond size, and sectors, the coefficient gets closer to zero and stays statistically not significant. Similarly, for the downgrade probability, we do not find any predicting power from ESG scores in our sample. These results at least do not allow us to reject the hypothesis that our model controls for the fundamental channel and isolates the preference channel of the ESG premium. This is definitely not a conclusive test, especially given the short period from April 2018 to March 2022.

Despite not showing predicting power for ex-post risks, which would invalidate this cross-sectional specification from the get go, ESG scores are clearly correlated with ex-ante measures of risks, and, again, there is no guarantee that we have controlled for all relevant variables. We now focus on the results of a better empirical framework to identify the preference channel of the ESG premium.

[21] We aggregate all bond characteristics at the issuer level by taking the market value weighted average of each bond characteristic each month for each issuer.

1.5.2 ESG investors preferences and expected return

Table 1.7 shows the main result of equation (1.7). The first column shows that our coefficient of interest is -9.6 bp and statistically significant. This is without controlling for sector fixed effects. This value implies that a company with the highest possible ESG score would see its cost of borrowing decrease by 48 bp more than a company with the worst ESG. However, such a difference in ESG scores, from 0 to 5, is not realistic. Looking at the summary statistics of our onshore-offshore matched sample in Table 1.3, the 95th percentile ESG score is 3.1 compared to 0.5 for the 5th percentile. This difference of 2.6 corresponds to a spread difference of 25 bp (=2.6 $\times$ 9.6), which is still high but far from the theoretical 48 bp value from the best to worst ESG company.

As mentioned before, we also distinguish between the across-sector and within-sector effect in columns two, three, four, and five. Here again, we find that all of the effect is due to within-sector differences in ESG scores. In the second column, the ESG premium coefficient increases in magnitude to 11.4bp, again statistically significant at 95% confidence interval. When regressing only on the median-sector ESG score, the coefficient becomes statistically insignificant. Conversely, the coefficient becomes strongly significant when using the residual sector ESG score (ESG score minus sector median).

Finally, in Figure 1.4, we report the results of the Granger test exposed in section 1.3.2. There are no statistically significant differences between the onshore-minus-offshore spread of high-ESG and low-ESG issuers during any of the months' pre-treatment (from February 2019 to August 2019), both without controls for sectors, panel a), and with controls for sectors, panel b).

1.5.3 Discussion of the implications for all market participants

Taken together, the results presented in sections 1.5.1 and 1.5.2 have several important implications for all market participants. First, investors and funds ought to know that sustainable investing in emerging markets should yield lower expected returns going forward if the levels of ESG consciousness and ESG taste stay constant. However, as our results in the Chinese onshore market show, an increase in the proportion of ESG-conscious investors or ESG taste will decrease the spread of high-ESG-score companies more than low-ESG-score companies. This transition dynamic will provide existing investors in high-ESG companies larger gains from the price appreciation that goes in pair with spread compression.

The second main implication of our results concerns debt issuers. Firms in emerging markets should strongly consider ESG-improving policies as part of their expenses, given the significant impact of ESG scores on their cost of debt, and especially given that most of the decrease in the cost of debt is due to within-sector ESG score variation. For instance, firms can perform a cost-benefit analysis to improve their ESG scores. Our analysis provides firms with an estimate of the decline in the cost of borrowing arising from one unit score improvement. Of course, firms should account for any modification to their credit risk resulting from additional expenses. This is relatively easy to measure because credit spreads are strongly linked to credit ratings. Additionally, firms should measure the opportunity cost to allocate resources to ESG improvements. From a researcher's viewpoint, there would be tremendous value in quantifying the cost for a firm to improve its ESG scores. To the best of our knowledge, we are not aware of any paper that has quantified the firms' costs of improving ESG performance in a systematic manner. We find some numbers in the sustainable statements of firms, but it remains an open question to link these costs to the actual improvement in the score. The answer to this question is beyond the scope of this paper, and we leave it for future research.

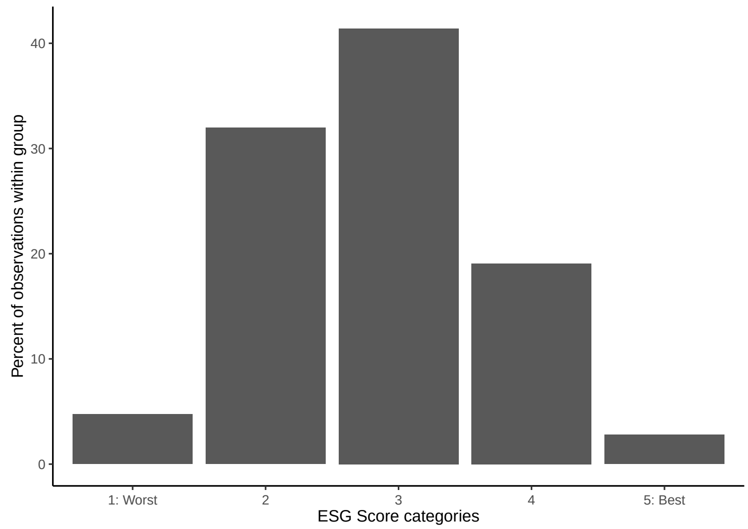
Finally, our paper has implications for ESG score providers. We show that companies that are awarded better scores by ESG score providers benefit from a lower cost of capital, especially as the demand for ESG increases. It is thus their duty to ensure that their scores properly reflect good ESG practices as opposed to simply greenwashing.

1.6 Conclusion

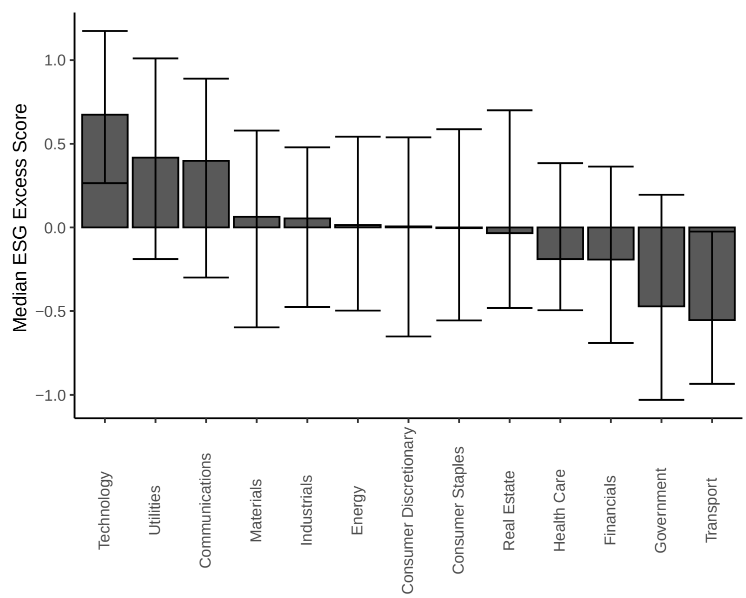
In this paper, we propose two tests derived from a simple theoretical framework to determine whether ESG non-pecuniary preferences are priced in emerging markets' corporate bonds. In our first test, we regress secondary market bond spreads on company ESG scores, controlling for measures of systematic, liquidity, and default risks. We find a statistically significant ESG premium over the period 2018–2022, originating mainly from within-sector variations of ESG scores. All else being equal, a bond issued by the best ESG company has an average spread approximately 90 bp lower than the worst ESG-rated issuer. In our second test, which better addresses the challenges in identifying the non-pecuniary source of the ESG premium, we test whether an increase in the proportion of ESG-conscious investors decreases the cost of borrowing more for high-ESG than low-ESG bonds, as predicted by theory. We use the opening of the Chinese onshore corporate bond market to international investors, who are more ESG-conscious, as a proxy for an increase in ESG preferences. We reduce considerably the influence of the fundamental channel by including issuer-time fixed effects in a triple difference framework, leveraging Chinese companies that also have bonds traded in the offshore market, which did not experience any regulatory changes. Following the opening of the Chinese onshore capital market, a company with the highest possible ESG score would have seen its cost of borrowing decrease by 48 bp more than a company with the worst ESG score. Overall, both results support the hypothesis that firms with a high ESG score benefit from a lower cost of borrowing because of investors' non-pecuniary preferences. The main limitations of this paper are twofold. First, further research is necessary to improve the identification of the ESG premium in the cross-section of emerging markets in order to fully shut down the fundamental channel. Second, additional research is required to better quantify the impact of preference changes on the cost of debt. Fitting a structural model to the opening of the Chinese onshore market can, for instance, present a promising line of future research.

1.7 Figures and Tables

Figure 1.1: ESG scores



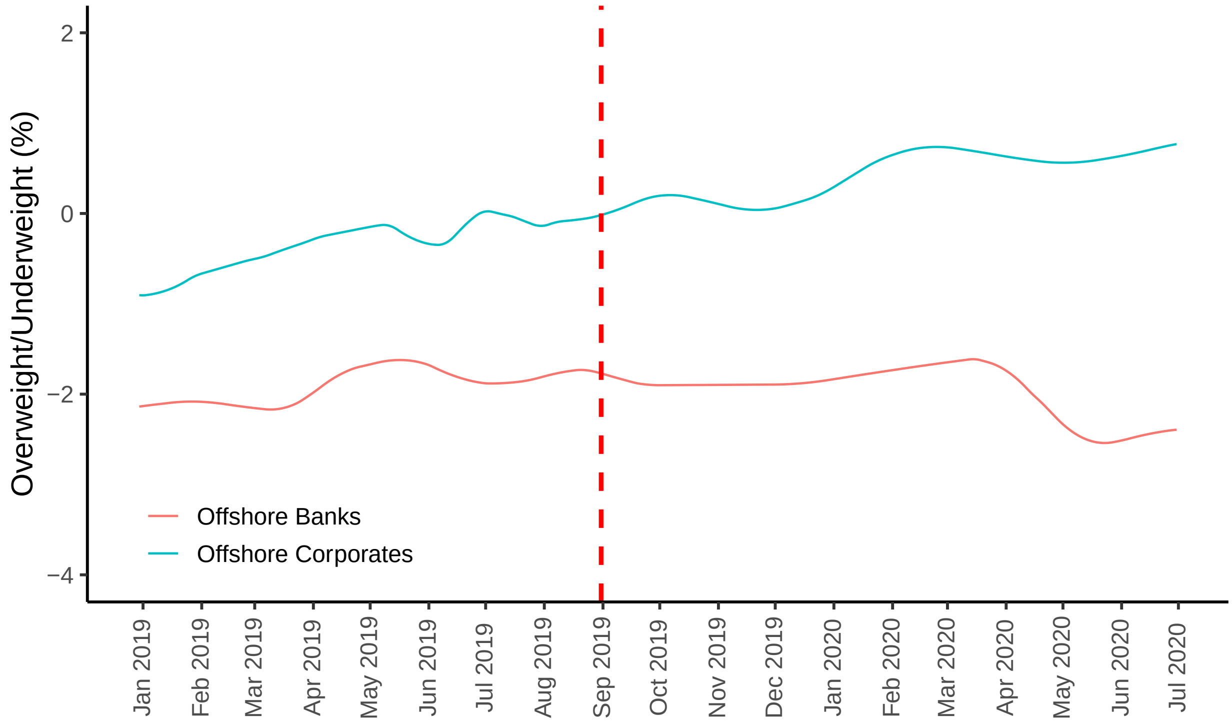
(a) Distribution of ESG scores



(b) Median ESG scores by sector

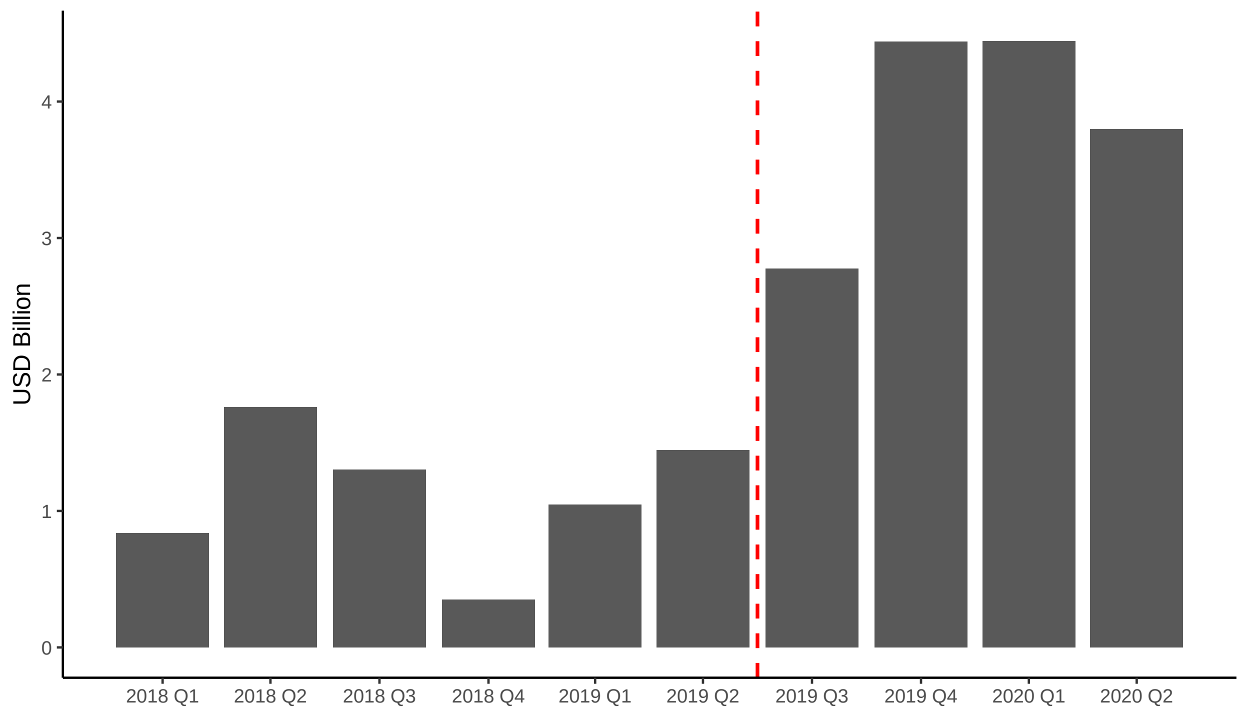
Note: This figure shows the distribution of companies' ESG scores. Panel a) provides the percentage of ESG scores in each category, where category "1" contains all scores from 0 to 1, "2" contains scores from 1 to 2, ... until category "5", which contains all ESG scores between 4 and 5. In Panel b), we plot the median ESG score by sector as well as its 25th and 75th percentiles with error bars.

Figure 1.2: Chinese offshore positioning



Note: This figure shows the positioning of institutional investors in Chinese offshore corporate bonds around the opening of Chinese onshore capital markets in September 2019. This is obtained from the EM Client Survey that covers approximately 140 institutional investors managing over $1 trillion assets. More details on the methodology can be found in Appendix A.3.

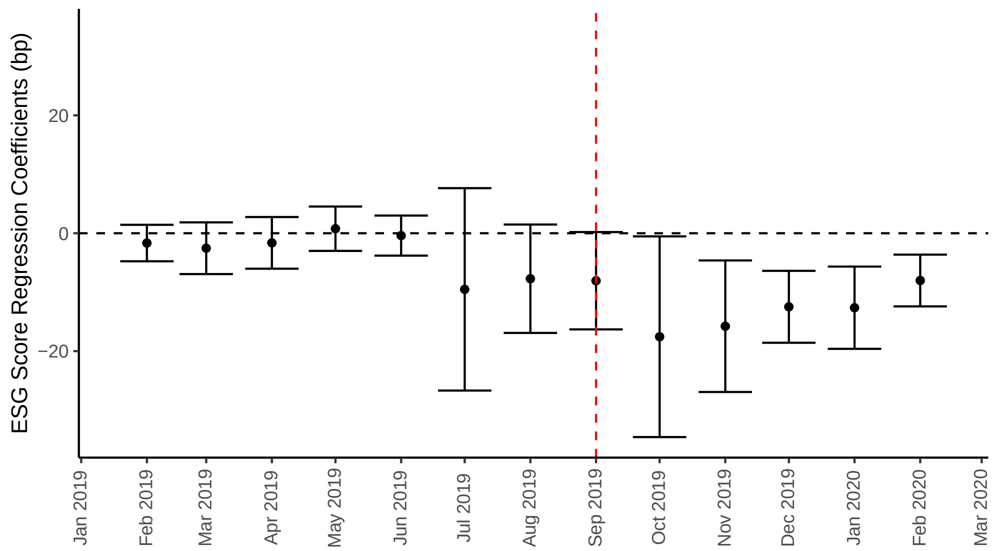
Figure 1.3: Quarterly net foreign flows into enterprise bonds and medium-term notes



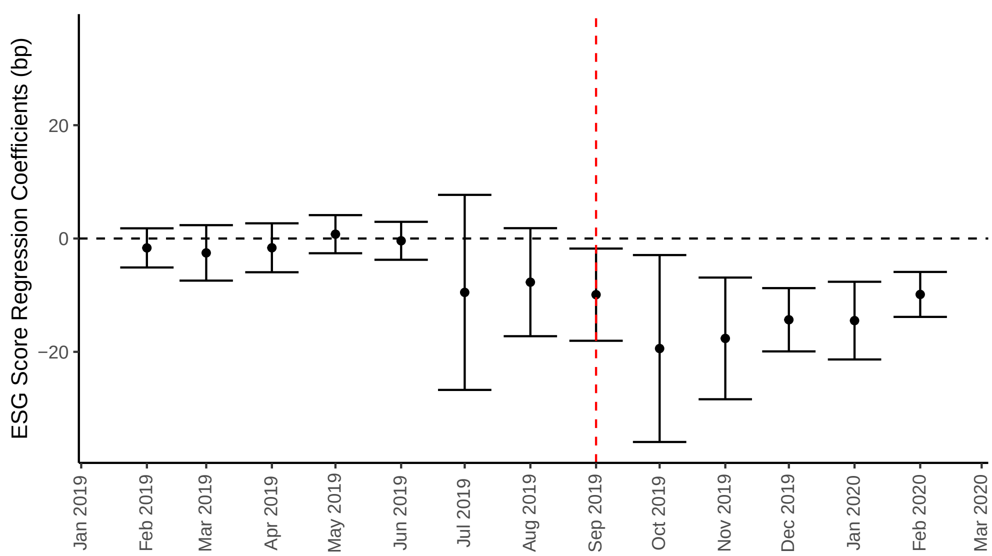
Note: This figure shows the total net inflows in USD billion from foreign investors during each quarter. We use the average exchange rate (USD:CNY) during the quarter to convert CNY into USD. The data are obtained from China Central Depository & Clearing (CCDC) and the Shanghai Clearing House (SHCH). The red dashed line indicates the Chinese onshore capital market opening described in section 1.3.2.

Figure 1.4: Bond yield and spread - dynamic coefficients associated with companies' ESG scores

(a) Without controlling for sectors



(b) With controls for sectors



Note: This figure shows the dynamic effect of the Chinese capital market opening by regressing bond spreads on $ESG_{s,t} \times Onshore_{i,t}$ multiplied by month dummies, as follows:

$$s_{i,s,m,t} = \alpha_i + \alpha_{s,t} + \alpha_{m,t} + \sum_{\tau} \gamma_\tau ESG_s \times Onshore_i \times D_t^\tau + \delta' X_{i,s,m,t} + \varepsilon_{i,s,m,t},$$

where $\tau \in$ {February 2019, ...,February 2020} and D_t^τ is a dummy variable taking value one if $\tau = t$, and zero otherwise. The base month is January 2019. Other variables in $X_{i,s,m,t}$ are described in section 1.3.2.

Table 1.1: ESG and bond spreads in the literature

Authors	E, S, or G?	Scope	# of bonds/loans	Period	Method	ESG Data Source[d]	ESG Premium (in bp)
Ferriani (2022)	ESG	Global	1,078	2020	OLS	MS, RF, RS, SA	-13
Halling et al. (2021)	ES	US	5,227	2002 - 2020	OLS	MS	-19
Zhang (2021)	ESG	US	11,157[b]	2002 - 2019	OLS	RF, SA	-6
Seltzer et al. (2020)	E	US	4,235	2009 - 2017	OLS, DiD	SA, carb. em.	-21
Amiraslani et al. (2018)	ESG	US	1,989	2005 - 2013	OLS, DiD	MS	-9
Ghouma et al. (2018)	G	CA	1,632	2013	OLS	Globe and Mail Report	-9
La Rosa et al. (2018)	ESG	EU	1,228[b]	2005 - 2012	OLS	A4	-26
Hasan et al. (2017)	S	US	32,425[b]	1990 - 2012	OLS, DiD	NRCRD	-4
Hoepner et al. (2016)	E	28 countries	470	2005 - 2012	OLS	Oekom	-64[e]
Ge and Liu (2015)	ESG	US	4,260	1992 - 2009	OLS, IV	KLD	-10
Stellner et al. (2015)	ESG	EU	872	2006 - 2012	OLS	A4, BBG	-10
Chava (2014)	E	US	5,879	1992 - 2007	OLS	KLD	-25
Kim et al. (2014)	ESG	19 countries	1,568[b]	2003 - 2007	OLS	SA	-20
Oikonomou et al. (2014)	ESG	US	3,240	1993 - 2008	OLS	KLD	-10[f]
Izzo and Magnanelli (2012)	ESG	4 countries	1,641[b]	2005 - 2009	OLS	RS	10
Nandy and Lodh (2012)	E	US	3,153	1991 - 2006	OLS	KLD	-13
Goss and Roberts (2011)	ESG	US	3,996	1991 - 2006	OLS	KLD	-13
Ye and Zhang (2011)	S	CN	2,833[b]	2007 - 2008	OLS, IV	Charity/Sales	U-shape[g]
Bauer and Hann (2010)	E	US	2,242	1995 - 2006	OLS	KLD	-6
Menz (2010)	ESG	EU	498	2004 - 2007	OLS	RS	+[h]
Derwall and Koedijk (2009)	ESG	US[a]	/	1987 - 2003	OLS, matching	US SIF	-130
Sharfman and Fernando (2008)	E	US	267[c]	1999 - 2002	OLS	KLD, TRI	24
Graham and Maher (2006)	E	US	357	1995 - 1998	OLS	EPA	-1
Klock et al. (2005)	G	US	1,877[b]	1990 - 2000	OLS	Governance Index	-34

[a] Derwall and Koedijk (2009) study US fixed income funds.
[b] It is the number of observations instead of the number of bonds.
[c] It is the number of firms instead of the number of bonds.
[d] MS stands for MSCI ESG. RF stands for Refinitive. RS stands for RobecoSAM. SA stands for Sustainanlytics. A4 stands for ASSET4. BBG stands for Bloomberg.
[e] It is the country ESG effect. Hoepner et al. (2016) do not find firm ESG effect.
[f] It is the average effect of aggregate strengths and aggregate concerns.
[g] Ye and Zhang (2011) find that the ESG effect on the cost of debt is U-shaped, with an optimal level of ESG investment.
[h] Menz (2010) finds that the risk premium for socially responsible firms was ceterius paribus higher than for non-socially responsible companies. The original paper is vague in its quantitative analysis.

Note: This table presents the results of our survey of research papers that study the relation between company ESG and expected return in fixed income markets. We exclude studies analyzing stock market returns as well as those analyzing green bonds. We focus on studies investigating across-firm variations in ESG scores, since this is the main objective of this paper.

Table 1.2: Summary statistics of offshore bonds across emerging markets

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
ESG score (0-5)	87,888	2.2	1.0	0.04	1.5	2.2	2.9	4.9
Par Amt (USD Million)	87,888	703.2	434.2	300	500	504.5	800	2,885
Rating #	87,888	9.6	3.2	1	7	10	12	17
Ex-ante prob. of default (%)	87,888	3.4	4.6	0.2	1.3	2.1	3.7	35.6
Spread (bp)	87,888	285.1	261.5	37	119	195	369	1,642
Duration	87,888	4.8	3.7	0.1	2.5	3.9	6.1	17.6
Duration-times-spread (DTS)	87,888	16.1	16.9	0.5	4.0	10.2	21.4	77.0
Beta	87,888	0.7	0.7	−0.1	0.2	0.6	1.1	3.0
Ex-ante annualized volatility (%)	87,888	6.6	8.2	0.4	1.9	3.8	7.5	47.6
Default next month (%)	85,695	0.01	0.9	0.0	0.0	0.0	0.0	100.0
Downgrade next month (%)	86,398	0.2	4.8	0.0	0.0	0.0	0.0	100.0

(a) Bond characteristics

Sector	Share (%)	Region	Share (%)	Rating	Share (%)
Financials	40.3	China	34.4	1: AAA-A	28.3
Energy	13.7	LATAM	25.0	2: BBB	37.9
Materials	10.6	CEEMEA	21.6	3: BB	20.7
Utilities	10.1	Asia (ex-China)	19.1	4: B	12.0
Communications	6.1			5: C and below	1.1
Industrials	5.4				
Consumer Discretionary	4.5				
Consumer Staples	4.1				
Government	3.4				
Technology	0.9				
Health Care	0.7				

(b) Bond distribution

Region	# of issuers	# of bonds
All	1,046	3,250
China	273	1,159
CEEMEA	269	711
LATAM	305	765
Asia (ex-China)	200	616

(c) Number of bonds and issuers by regions

Note: This table shows the summary statistics of our sample of offshore corporate bonds and quasi-sovereign bonds across all emerging markets, which is used for the analysis of the ESG premium presented in section 1.3.1. Details regarding the construction methodology of these variables can be found in section 1.4.

Table 1.3: Summary statistics of filtered matched sample of onshore and offshore bonds

	N	Mean	St. dev.	P5	P25	Median	P75	P95
Onshore								
# of issuers	56							
# of bonds	323							
# of bonds per issuer		5.77	7.44	1.00	2.00	4.00	7.00	13.25
ESG Score	4,522	1.93	0.77	0.53	1.56	1.85	2.58	3.10
Yield (%)	4,522	3.79	0.77	2.94	3.36	3.67	4.08	5.06
Spread (bp)	4,522	108.96	72.67	35.91	70.58	95.74	126.49	234.11
Duration	4,522	3.08	1.75	1.41	1.87	2.46	3.63	6.96
Credit Rating	2,896	5.91	1.42	5.00	5.00	5.00	7.00	9.00
Amt. issued ($ Million)	4,522	2,062.06	3,209.20	157.23	314.47	471.70	2,830.19	8,176.10
Coupon (%)	4,522	4.48	0.77	3.25	3.98	4.50	4.97	5.61
Offshore								
# of issuers	56							
# of bonds	197							
# of bonds per issuer		3.52	4.66	1.00	1.00	2.00	3.00	12.75
ESG Score	2,758	2.00	0.70	0.53	1.64	1.85	2.36	3.14
Yield (%)	2,758	3.21	1.12	1.99	2.54	3.00	3.57	5.27
Spread (bp)	2,758	141.03	99.43	64.00	91.00	118.00	151.00	351.30
Duration	2,758	3.03	2.13	0.07	1.75	2.76	4.15	6.85
Credit Rating	2,585	6.80	1.77	5.00	6.00	6.00	8.00	10.00
Amt. issued ($ Million)	2,758	669.00	377.02	300.00	500.00	500.00	750.00	1,400.00
Coupon (%)	2,758	3.85	1.04	2.45	3.15	3.62	4.38	6.00
Correlation								
Cor. btn. spreads	0.73							

Note: This table exhibits the summary statistics of our matched filtered sample of onshore and offshore bonds. The methodology to obtain this sample is exposed in section 1.4.

Table 1.4: Cross-sectional regression of spreads on ESG company score

	Spread					
	Whole sample		Pre-Covid		Post-Covid	
ESG Score	-18.86^{***}	-18.42^{***}	-12.27^{***}	-11.81^{***}	-20.05^{***}	-19.59^{***}
	(3.71)	(3.67)	(3.38)	(3.32)	(4.44)	(4.44)
Sector-time FE		X		X		X
Controls	X	X	X	X	X	X
Observations	87,888	87,888	38,551	38,551	49,337	49,337
Adjusted R^2	0.70	0.71	0.74	0.74	0.71	0.72

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

(a) Across all emerging markets

	Spread					
	Whole sample		Pre-Covid		Post-Covid	
ESG Score	-8.94^{**}	-10.92^{***}	-6.86^{**}	-7.58^{**}	-9.16^{*}	-11.53^{**}
	(3.49)	(3.87)	(2.95)	(3.25)	(4.68)	(5.15)
Sector-time FE		X		X		X
Controls	X	X	X	X	X	X
Observations	30,260	30,260	12,547	12,547	17,713	17,713
Adjusted R^2	0.84	0.84	0.84	0.84	0.85	0.85

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

(b) China

Note: This table presents the ESG premium in basis point, which corresponds to the coefficient γ of equation (1.6) presented in section 1.3.1. The whole sample covers the period April 2018 to March 2022. The pre-COVID period stops in March 2020, and the post-COVID period starts in April 2020. Panel a) includes all countries in emerging markets, while panel b) only includes China. We report robust standard errors clustered two-way at the issuer and time levels.

Table 1.5: ESG scores and ex-post risks

	Probability of default (Percent)	
ESG Score	−0.060	−0.015
	(0.041)	(0.038)
Ex-ante default prob.		1.989*
		(1.123)
Rating		0.155***
		(0.037)
Duration		−0.025**
		(0.011)
log(Par)		0.099*
		(0.057)
Cluster S.E.	Issuer+Time	Issuer+Time
Observations	16,210	16,210
Adjusted R^2	0.001	0.009

*p<0.1; **p<0.05; ***p<0.01

(a) Default probability

	Probability of downgrade (Percent)	
ESG Score	0.306	0.436
	(0.469)	(0.490)
Ex-ante default prob.		52.461
		(34.563)
Rating		1.969***
		(0.312)
Duration		−0.370***
		(0.104)
‘log(Par)‘		0.545
		(0.591)
Time FE	*X*	*X*
Controls		*X*
Cluster S.E.	Issuer+Time	Issuer+Time
Observations	13,627	13,627
Adjusted R^2	0.005	0.052

*p<0.1; **p<0.05; ***p<0.01

(b) Downgrade probability

Note: This table shows the coefficients from the following linear regression:

$$Y_{s,t\to t+12} = \alpha_t + \beta x_{s,t} + \varepsilon_{s,t\to t+12}$$

where $Y_{s,t\to t+12}$ is a dummy variable that takes 1 if the issuer defaulted (panel a) or got downgraded from investment-grade to high-yield (panel b) at any point in the next 12 months, and zero otherwise. $x_{s,t}$ contains the company ESG score in the first column, with additional controls (including sector-fixed effects) in the second column. We report robust standard errors clustered two-way at the issuer and time levels.

Table 1.6: ESG scores and ex-ante risks

	ESG Score		
	All	High Yield	Investment Grade
Ex-ante default prob.	−0.329	0.245	−0.931
	(0.693)	(0.684)	(1.412)
Rating	−0.016*	−0.036**	−0.017
	(0.009)	(0.017)	(0.022)
Duration	0.012	0.0004	0.006
	(0.009)	(0.014)	(0.011)
Beta	0.107***	−0.009	0.181***
	(0.040)	(0.040)	(0.067)
Vol	−0.053***	−0.020	−0.013
	(0.015)	(0.017)	(0.022)
log(Par)	−0.119***	−0.019	−0.173***
	(0.030)	(0.042)	(0.039)
Sector FE	*X*	*X*	*X*
Cluster S.E.	Issuer+Time	Issuer+Time	Issuer+Time
Observations	35,587	16,619	18,968
Adjusted R^2	0.133	0.123	0.202

*p<0.1; **p<0.05; ***p<0.01

Note: This table shows the coefficients from a linear regression of company ESG score on company's fundamental variables, over the entire sample, the high-yield segment, and the investment-grade segment in columns 1, 2, and 3, respectively. We report robust standard errors clustered two-way at the issuer and time levels.

Table 1.7: Investors' preferences and ESG premium: Triple-difference regression

	Dependent variable:				
	Spread				
ESG Score X Onshore X Post	-9.58^{**}	-11.44^{**}			
	(3.79)	(3.88)			
Sector ESG Score X Onshore X Post			10.34		1.70
			(9.24)		(11.22)
Residual Sector ESG Score X Onshore X Post				-9.78^{**}	-9.67^{**}
				(3.46)	(3.68)
Sector-Onshore-Post FE		*X*			
Controls	*X*	*X*	*X*	*X*	*X*
Bond FE	*X*	*X*	*X*	*X*	*X*
Issuer-Time FE	*X*	*X*	*X*	*X*	*X*
Market-Time FE	*X*	*X*	*X*	*X*	*X*
Observations	7,280	7,280	7,280	7,280	7,280
Adjusted R^2	0.96	0.96	0.96	0.96	0.96

*p<0.1; **p<0.05; ***p<0.01

Note: This table shows the results from regression (1.7):

$$s_{i,s,m,t} = \alpha_i + \alpha_{s,t} + \alpha_{m,t} + \gamma ESG_s \times Onshore_i \times Post_t + \delta' X_{i,s,m,t} + \varepsilon_{i,s,m,t}$$

where $Onshore_i$ is a dummy that takes value one if the bond is onshore, $Post_t$ is a dummy that takes one if $t \geq$ Sep 2019, and ESG_s is the ESG score of the company at the time of opening in September 2019. The "Sector ESG Score" corresponds to the monthly sector-specific median across all emerging markets' ESG scores. The "Residual Sector ESG Score" corresponds to the company ESG score minus the sector median. We report robust standard errors clustered two-way at the issuer and time levels.

Chapter 2

Watch what they do, not what they say: Estimating regulatory costs from revealed preferences[1]

2.1 Introduction

Many regulatory agencies are required—by statute or executive order—to perform a cost-benefit analysis (CBA) on a proposed rule.[2] A CBA on financial regulation usually compares the social benefits of regulations, such as those arising from a reduced probability of a financial crisis, with the regulatory costs, which are mainly borne by financial institutions that must comply with them. CBA could form the basis for judicial review and Congressional oversight of regulatory actions.[3] Although there is a growing body of research on quantifying regulatory benefits to the public, quantifying regulatory costs borne by financial institutions is often viewed as a mundane task interesting only to regulators, and as a result, it has received less academic scrutiny.

Quantifying the costs of financial regulation, however, is far from trivial. Financial regulators generally do not have enough information to gauge the regulatory impacts on complex financial

[1]This chapter is based on Alvero et al. (2022). We thank Juliane Begenau, Utpal Bhattacharya, Olivier Darmouni, Will Diamond, Xavier Gabaix, Vidhan Goyal, Ernest Liu, Jens Dick-Nielsen, Yiming Ma, Claudia Robles-Garcia, Philipp Schnabl, Rajdeep Sengupta, Philip Strahan, Adi Sunderam, and participants at the Western Finance Association Annual Meeting, the SFS Cavalcade, the Financial Intermediation Research Society Annual Meeting, the Federal Reserve Stress Testing Research Conference, the FDIC/JFSR Bank Research Conference, the Swiss National Bank Workshop, the Financial Intermediation Discussion Group, Columbia Finance Lunch, the Hong Kong University of Science and Technology, the Norges Bank Conference, and the Midwest Finance Association Annual Meeting for helpful comments and discussions. We thank Yiyuan Zhang for his excellent research assistance. This paper reflects the views of the authors and does not reflect the views of the IMF, its Executive Board, or IMF management.

[2]The current set of requirements includes Executive Order 12866 and Office of Management and Budget (OMB) Circular A-4, the Regulatory Flexibility Act (RFA), and the Unfunded Mandates Reform Act (UMRA). These requirements vary in terms of the agencies and rules they cover, and the types of analyses that are required (Carey, 2014). Independent financial regulators are exempt from Executive Order 12866 requirements but are subject to other requirements to assess the effects of their rules. See CRS Report R42821, *Independent Regulatory Agencies, Cost-Benefit Analysis, and Presidential Review of Regulations.*

[3]Judicial review is a right imbued by statute, but some elements of analysis can be shielded from judicial review.

institutions. Instead, they often have to rely on self-reported estimates from financial institutions to guide their policies. However, these self-reported estimates may be inflated by financial institutions to justify regulatory relief. For instance, Parker (2018a,b) suggests that a highly influential estimate of regulatory costs cited by the 2016 House Concurrent Budget Resolution comes from studies "funded by organizations having a strong financial or institutional stake in the outcome of their studies" and the methodologies used to estimate regulatory costs are "fundamentally flawed." Nevertheless, these self-reported estimates are widely cited in policy debates and have promoted efforts to roll back financial regulations enacted after the 2008 financial crisis.[4] The lack of independent and rigorous assessment of regulatory costs highlights the need for academic research in this area (Posner and Weyl, 2013; Coates, 2014).

This paper proposes a revealed preference approach to infer regulatory costs from the regulatory distortion on the size distribution of financial institutions. To illustrate our approach, consider the Dodd–Frank Act of 2010, which imposes additional regulations on banks when their assets cross the \$10 billion and \$50 billion thresholds.[5] As shown in Figures 2.1a and 2.1b, banks shrink their assets to avoid these regulations, creating excess densities around these thresholds. Such distortions are not present in the pre-Dodd–Frank period, or around other round numbers that are not regulatory thresholds, such as \$20 billion or \$40 billion, as shown in Figure 2.2a and Figure 2.2b. We show that the extent of such regulatory distortions can reveal the regulatory costs faced by banks.

To formalize this idea, we develop a structural model of the size distribution of banks. In the model, regulatory costs are modeled as a tax on banks' profits in the spirit of Posner (1971). The tax rate jumps discretely at regulatory thresholds, reflecting the regulations that come into effect once banks' assets exceed these regulatory thresholds.[6] The discrete jumps in the tax rate create

[4]For instance, in January 2017, the Trump administration issued Executive Order 13771 that mandates "for every one new regulation issued, at least two prior regulations be identified for elimination" (this order was revoked by Executive Order 13992 in January 2021). In May 2018, Congress passed the Economic Growth, Regulatory Relief, and Consumer Protection Act, which scales back many financial regulations enacted after the crisis.

[5]We provide a more detailed discussion on the regulations triggered by each threshold in Section 2.2.

[6]This formulation also allows for the possibility that certain regulations may generate private value for banks that comply with them. In this case, the tax rate should be interpreted as the net regulatory burden. See Section 2.3.1 for a more detailed discussion.

an incentive to bunch. Banks that are just above the regulatory threshold shrink their size to avoid regulation. Banks that are far above the regulatory threshold do not bunch because the costs of bunching outweigh the costs of regulation.

We use the model to guide our estimation of the regulatory costs. In the absence of regulatory distortions, the size distribution of banks should not display any excess density around a particular threshold. In the presence of regulation, however, banks bunch to avoid regulation, creating excess densities around regulatory thresholds. The magnitude of the excess densities then reveals the regulatory costs borne by banks. Using this idea, we derive a maximum likelihood estimator and apply it to U.S. bank size data to estimate the regulatory costs imposed by the Dodd–Frank Act. The estimation shows that the regulation costs triggered at the \$10 billion threshold are equivalent to a 0.41% tax on banks' average annual profits. The additional regulatory costs triggered at the \$50 billion threshold are equivalent to a 0.11% tax. The estimated regulatory costs are economically significant. For a bank with \$50 billion assets, the regulatory costs amount to \$4.16 million per year,[7] which is equivalent to the salary expenses of hiring 52 additional compliance officers, assuming the average annual compensation for a compliance officer is around \$80,000 (Feldman et al., 2013).

Practitioners often point to the low franchise values in the post-crisis period as evidence of high regulatory costs imposed by the Dodd–Frank Act. Although this argument is appealing, it remains unclear how much regulations can explain the decline in bank values (Sarin and Summers, 2016; Chousakos and Gorton, 2017). Our structural model can study this question quantitatively by simulating bank values with and without the estimated regulatory costs. We find the estimated regulatory costs only explain a small fraction of the decrease in bank franchise value in the post-crisis period. Our result suggests non-regulatory factors, such as ultra-low interest rate environment (Calomiris and Nissim, 2014; Whited, Wu, and Xiao, 2021), market reassessment of risks (Sarin

[7]The regulatory costs introduced by the Dodd–Frank Act for a \$50 billion bank are equivalent to $0.41\% + 0.11\% = 0.52\%$. We add up the regulatory tax triggered at \$10 billion and \$50 billion to calculate the total Dodd–Frank burden for a \$50 billion bank because it is subject to the regulations triggered at both \$50 billion and \$10 billion. We then calculate the dollar value of regulatory costs per year by multiplying the regulatory tax rate with the average profits, $\$50,000M \times 1.6\% \times 0.52\% = \$4.16M$, where 1.6% is the ratio of average annual profits to assets.

and Summers, 2016), and the removal of too-big-to-fail subsidies (Atkeson, d'Avernas, Eisfeldt, and Weill, 2019; Berndt, Duffie, and Zhu, 2020) may also have contributed to the decline in bank values.

The Dodd–Frank Act also has distributional implications in the cross-section of banks. One may expect the market shares of big banks (>$50 billion) would shrink after the Dodd–Frank Act because they face the most-stringent regulations. Interestingly, the counterfactual simulation suggests that the market share of big banks would expand after the Dodd–Frank Act. The reason for the expansion of big banks is twofold. First, medium banks ($10 billion to $50 billion) engage in regulatory avoidance, which reduces their average asset size. Second, heightened regulatory costs depress bank value across the size spectrum, which reduces the entry of small banks (<$10 billion). This result suggests a possible unintended consequence of regulation on the bank industry's market structure.[8]

Regulation not only imposes compliance costs for banks but also leads to indirect costs for the rest of the economy. As banks shrink their size to avoid regulation, bank-dependent firms could be adversely affected if banks shrink lending.[9] Regulation could also affect aggregate credit supply from the extensive margin of bank entry and exit.[10] Quantifying the indirect costs of regulation requires solving the full market equilibrium. To this end, we calibrate parameters that govern the equilibrium supply and demand of bank credit using either values in the prior literature or the corresponding moments in the data. We then simulate counterfactual economies with and without the direct regulatory costs estimated from our maximum likelihood estimator. We quantify the indirect costs of regulation as the lost output of bank-dependent firms due to the regulation. We find that the indirect costs of Dodd–Frank Act appear to be modest: the indirect regulatory costs are equivalent to a 0.02% tax on the output of bank-dependent firms.

[8]This result does not imply that the Dodd–Frank Act has worsened the too-big-to-fail problem because we do not explicitly model how regulation affects big banks' risk-taking and their reliance on the implicit government guarantee.

[9]See Bouwman et al. (2018) for evidence that banks near the Dodd–Frank threshold decrease their lending. Note banks could shrink their assets by reducing securities instead of loans. We can interpret securities as a form of bank credit supply to the economy.

[10]It is possible that firms can substitute to shadow banks so firms' overall credit access is not affected. We discuss these issues in Section 2.4.4.

We compare our estimated regulatory costs with those from other methodologies. We show that reduced-form methods such as difference-in-differences and regression-discontinuity design are likely to underestimate the direct regulatory costs because banks can strategically avoid regulation. Furthermore, these methods cannot capture implicit regulatory costs that are not measured in banks' financial statements. Indeed, these reduced-form methods usually find little evidence of changes in regulatory costs after the Dodd–Frank Act. In contrast, estimates based on self-reported surveys from banks are usually much larger than our estimates, consistent with anecdotal evidence that these estimates may be inflated to lobby regulators for regulatory relief (Hinkes-Jones, 2017; Parker, 2018b).

We conduct several extensions and robustness checks on our results. First, we extend our analysis to the 2018 Economic Growth, Regulatory Relief, and Consumer Protection Act of 2018, which rolled back many regulatory requirements imposed by the Dodd–Frank Act. In the data, the excess densities around the Dodd–Frank thresholds have decreased significantly after the 2018 regulatory relief, suggesting the incentive to avoid regulation has weakened since then. Indeed, the estimated regulatory costs in the post-relief period are significantly smaller than those in the post-Dodd–Frank period. We also find that the number of banks in the steady state increases after the 2018 regulatory relief, consistent with an increase in bank entry observed in the data.

Second, our maximum likelihood estimator assumes the undistorted bank assets follow a power law distribution. Although this assumption is motivated by the data, we assess the robustness of our estimation to alternative distributions. To this end, we re-estimate the regulatory costs using an alternative distribution assumption for the maximum likelihood estimator and find the results are robust. Our estimates are not sensitive to the distribution assumption because the regulatory costs are mainly identified from the local abnormal densities around the threshold and are insensitive to the global property of the distribution.

Third, our baseline estimation uses all the post-Dodd–Frank years to estimate the regulatory cost. However, if there are convex adjustment costs, the bank size distribution may take time to adjust to its new steady state. The still-in-progress changes in the size distribution may bias our estimates downwards. To address this concern, we drop the first several years post the Dodd-Frank Act and re-run the estimation. We find that the estimated regulatory costs are quite similar to the baseline estimates, which suggests that bank size adjustment is relatively fast.

Finally, we conduct several placebo tests by applying our estimator to the pre-Dodd–Frank sample and find no regulatory costs triggered by the thresholds in this sample. We also conduct placebo tests on round numbers that are not regulatory thresholds in the post-Dodd–Frank period. Again our estimator correctly indicates null results. We also show that our results are robust if we allow bank profit margin to be size-dependent or use alternative measures of bank regulatory assets.

Our paper relates to the literature on cost-benefit analysis (CBA) of regulation. There is extensive literature on regulatory CBA in environmental economics and industrial organization (Harberger, 1964; Viscusi and Aldy, 2003). However, analogous literature has been conspicuously missing in financial economics until recently (Posner and Weyl, 2013). The existing research focuses on quantifying the benefits of financial regulation.[11] However, quantifying regulatory costs has received less attention so far. A contribution of our paper is a revealed preference approach to quantify regulatory costs.[12] Our revealed preference approach is less prone to potential bias in self-reported surveys, and complements reduced-form methods to estimate regulatory costs such as difference-in-differences and regression discontinuity, for which endogenous selection around the threshold is an impediment for identification.

Our approach is related to the bunching literature in public finance and labor economics (Saez, 2010; Chetty, Friedman, Olsen, and Pistaferri, 2011; Kleven and Waseem, 2013).[13] The classic bunching approach is to take a known discontinuity in a tax schedule and then estimate preference

[11]Coates (2014) surveys the literature on the estimation of the GDP losses due to financial crisis and how much bank regulation can reduce such losses. Posner and Weyl (2013) propose and estimate a parameter called the statistical cost of a crisis (SCC), which allows regulators to translate the reduced probability of a crisis into a dollar value with certainty.

[12]Some researchers have developed indices of regulatory costs based on textual analysis of the regulatory provisions or compliance-related spending (Al-Ubaydli and McLaughlin, 2017; Calomiris, Mamaysky, and Yang, 2020; Simkovic and Zhang, 2020). These indices are helpful for understanding the cross-section variations in regulatory exposure, but they are difficult to be translated into monetary values in a CBA.

[13]See also DeFusco and Paciorek (2017), Dagostino (2018), Bachas, Liu, and Yannelis (2019), Alvero and Xiao (2020), Bachas, Kim, and Yannelis (2020), and Antill (2020) for applications of the bunching approach in finance.

parameters, such as labor supply elasticity. In contrast, this paper takes banks' preferences as given and then backs out a tax equivalent of regulatory costs. In other words, our approach is a bunching estimator in reverse. In addition, our paper differs from the prior bunching studies by taking a full structural approach, which allows us to estimate the indirect regulatory cost and conduct a rich set of counterfactuals. By doing so, our paper adds to a growing body of literature that employs structural techniques to study financial markets (Koijen and Yogo, 2015; Egan, Hortaçsu, and Matvos, 2017; Buchak, Matvos, Piskorski, and Seru, 2018b; Benetton, 2021; Nelson, 2018; Robles-Garcia, 2019; Darmouni, 2020; Xiao, 2020; Corbae and D'Erasmo, 2020a; Begenau and Landvoigt, 2021).

This paper also relates to the vast literature on financial regulation. The aftermath of the 2008–2009 financial crisis witnessed one of the most active periods of financial regulation in U.S. history. The unprecedented wave of financial regulation has stimulated a fast-growing body of research. The existing literature shows from a variety of perspectives how the new regulations affect the financial system (Buchak, Matvos, Piskorski, and Seru, 2018b; Acharya, Berger, and Roman, 2018; Cortés, Demyanyk, Li, Loutskina, and Strahan, 2020; Fuster, Plosser, and Vickery, 2018; Carletti, Goldstein, and Leonello, 2020; Kashyap, Tsomocos, and Vardoulakis, 2020). We contribute to this literature by quantifying the costs of Dodd–Frank regulations on banks. This question is crucial because the narrative of repressive regulatory costs has made many Dodd–Frank regulations targets of repeal. We find that the self-reported estimates from the finance industry are usually much larger than the estimates from the revealed preference approach, consistent with the anecdotal evidence that regulatory costs are often "hyped" by the finance industry (Hinkes-Jones, 2017, p.1). Furthermore, we study the impacts of the Dodd–Frank regulations on bank size, productivity, and profit distribution. By doing so, our paper contributes to the literature on the distortionary effects of size-based regulation, which has mainly focused on the labor market so far (Garicano, Lelarge, and Van Reenen, 2016; Gourio and Roys, 2014; Ando, 2021).

2.2 Institutional background

We first discuss the cost-benefit analysis of financial regulation. Then we discuss the Dodd–Frank Act and its impact on the size distribution of banks.

2.2.1 Cost-benefit analysis (CBA) of regulation

CBA entails an economic or statistical assessment of the social benefits of the regulation and the compliance costs borne by the regulated parties.[14] The goal of CBA is to advance regulators' ability to increase welfare and allow the public to detect and push back against regulations that fail to do so. CBA often forms the basis for judicial review and Congressional oversight of regulatory actions.

CBA of financial regulations has emerged as an important point of policy debate since the passage of the Dodd–Frank Act. Proponents argue that financial regulators should be held accountable for quantified CBA to ensure their discretion on rulemaking is not abused (Posner and Weyl, 2013). However, others caution about the inherent difficulties in CBA of financial regulations because of the complexity of financial institutions and the lack of reliable data (Coates, 2014; Cochrane, 2014).

2.2.2 The Dodd–Frank Act

The Dodd–Frank Wall Street Reform and Consumer Protection Act of 2010 (Dodd–Frank) is the centerpiece of post-crisis financial reform. The Dodd–Frank Act takes a tiered regulatory approach: banks are classified into size categories based on several regulatory thresholds. Banks in larger size categories are subject to stricter regulations. Specifically, banks whose assets exceed the $10 billion threshold are required to (1) conduct annual stress tests, (2) comply with the Durbin Amendment, which puts a cap on the fees charged to merchants for debit card transactions, (3) report to the Consumer Financial Protection Bureau (CFBP), a government agency created as part

[14] Appendix Section B.1 discusses the major statutes that require financial regulators to conduct CBA and a case study of a judicial review of a financial regulation based on the CBA.

of Dodd–Frank, and (4) create risk committees with independent directors. Banks whose assets exceed $50 billion are subject to additional risk-based capital and liquidity requirements, stress tests, and annual resolution plans. A detailed summary of the law can be found in Huntington (2010).

The tiered regulation creates discontinuities in the regulatory burden at the regulatory thresholds. In response to such discontinuities, banks around the thresholds strategically downsize their assets to avoid regulation. As shown by the red solid line in Figures 2.1a and 2.1b, the cumulative distribution function of bank size displays abnormal bulges around the $10 billion and $50 billion thresholds after the passage of the Dodd–Frank Act, suggesting that banks bunch their assets to avoid regulation.[15] This pattern is consistent with Bouwman, Hu, and Johnson (2018), who find that banks around the Dodd–Frank thresholds substantially reduce their assets.[16] Such a pattern is not present in the pre-Dodd–Frank period, as shown by the blue dashed line in Figures 2.1a and 2.1b.

One may worry that banks may simply cluster around round numbers and do not necessarily try to avoid regulation. To address this concern, we examine round numbers that are not regulatory thresholds, such as $20 billion or $40 billion. We find no excess density around these round numbers, as shown in Figures 2.2a and 2.2b. This result suggests that the excess densities around $10 billion and $50 billion are not driven by banks' clustering at round numbers. Instead, they are the outcomes of banks' strategic response to the regulations imposed by the Dodd–Frank Act.

2.2.3 Bank size determination for regulatory purposes

The Dodd–Frank Act does not provide a uniform methodology to determine bank size for regulatory purposes. Instead, the implementation of different provisions was left to separate rulemaking processes, which led to slightly different methodologies to determine bank size for regulatory purposes. For instance, the annual stress test uses the average assets in the past four quarters;[17] the

[15]Online Appendix Figure B.1 shows the histograms of bank size around the regulatory thresholds.

[16]This pattern is also consistent with Eisenbach et al. (2021), who find regulators spent considerable more work hours on banks above $10 billion in the post-crisis period.

[17]See `https://www.ffiec.gov/PDF/FFIEC_forms/FFIEC016_20171006_i_draft.pdf`.

Durbin Amendment uses the end of the calendar year assets;[18] and the CFPB requirement uses the minimum of the last four consecutive quarters with the exception that all institutions above \$10 billion as of the June 30, 2011 would be required to report to CFPB, and would only stop to be if they reported total assets below \$10 billion for four consecutive quarters.[19] In this paper, we use the simplest measure—the quarter-end assets—to construct the size distribution, and conduct different robustness checks with other methodologies to determine bank size. We will also allow a measurement error term in the estimation to account for these intricacies in the regulation.

2.3 Model

This section introduces our theoretical framework. We first describe individual banks' optimal size choices in the presence of the size-based regulation in a partial equilibrium setting in which the distribution of productivity and the lending rate are given. We derive a sufficient statistic formula for the direct regulatory costs. Then, we endogenize the productivity distribution and the lending rate in a general equilibrium framework with firms, which allows us to quantify the indirect costs of bank regulation on the rest of the economy.

2.3.1 Bank size choice and direct cost of regulation

The economy is populated by heterogeneous banks indexed by their productivity z, following a probability density function $g(z)$. Although we refer to z as productivity, one should interpret z broadly as any non-regulatory factors that affect banks' preferred size. For example, the value of z might vary with factors such as banks' market power and deposit base. Higher values of z correspond to larger potential bank size in the absence of regulatory distortion. A bank can raise funding from depositors at a cost $r(q|z)$, where q is the log quantity of funds. Banks' funding cost is increasing in q and is decreasing in z, that is, $r_q > 0$ and $r_z < 0$. As a result, a more productive bank can raise more funding for funding cost r. A simple example of the funding cost function is $r(q|z) = \frac{1}{\theta}(q - z)$, where $\theta \equiv \frac{1}{r_q}$ is the semi-elasticity of funding supply. Banks then lend to firms

[18] See `https://www.federalreserve.gov/paymentsystems/regii-faqs.html`.
[19] See `https://www.fdic.gov/news/press-releases/2011/pr11179a.html`.

with a rate R. For now, the distribution of the productivity z and the lending rate R are taken as exogenous. We will endogenize these variables in Section 2.3.2.

Banks face a size-based regulation that classifies banks into $I+1$ categories based on I size thresholds, $\underline{q_i}$, where $i = 1, ..., I$. If a bank's assets cross threshold $\underline{q_i}$, it will incur an additional regulatory cost that is equivalent to τ_i fraction of its profits.[20] Our approach of measuring regulatory costs as a tax is in a spirit similar to Posner (1971). One can translate the regulatory costs to an annual monetary value by multiplying them with the annual profits. One can also interpret τ_i as the fraction of bank value loss due to the present value of all future regulatory costs. This formulation of regulatory costs can be applied in many other settings where the specifics of regulations may differ.

Facing the size-based regulation, banks choose size q to maximize profits

$$\max_q \pi(q|z) = \max_q (R - r(q|z)) \exp(q) \cdot \prod_{i=1}^{I} (1 - \tau_i \mathbb{1}_{q \geq \underline{q_i}}). \tag{2.1}$$

As shown in the above formula, the regulatory costs increase discretely by τ_i once a bank crosses the i's threshold. τ_i captures the regulations that come into effect at the corresponding threshold. Note that τ_i does not include regulations that all banks need to comply with or regulations triggered by other thresholds. Neither does τ_i capture the potential indirect costs of regulation on the rest of the economy, which we will discuss in Section 2.3.3.

Certain regulations may generate private value for banks in compliance with them. For instance, regulatory disclosure can reduce the cost of capital for banks above the threshold. In this case, the tax rate reflects a net regulatory cost, the difference between the reporting burden and the private value from a lower cost of capital. We do not attempt to separate further the reporting burden and the private value to the regulated parties because it is sufficient to estimate the net costs for CBA. Regulators can compare the estimated net costs borne by financial institutions with the social benefits of regulation.

[20] In the data, bank profits may fluctuate due to random shocks while the regulatory burden is largely independent of these random fluctuations. So one should interpret the regulatory cost as a τ_i fraction of the expected profits rather than realized profits.

We first solve for the optimal undistorted log assets when there is no regulation by setting all τ_i to zero:

$$q_0(z) \equiv \arg\max_q (R - r(q|z)) \exp(q). \tag{2.2}$$

The optimal undistorted size is obtained by equalizing the profit margin to the inverse semi-elasticity of funding supply, $R - r(q_0(z)|z) = r_q(q_0(z)|z)$. If the funding supply function is $r(q|z) = \frac{1}{\theta}(q - z)$, then the optimal undistorted size is given by

$$q_0(z) = z + \theta R - 1, \tag{2.3}$$

where θ is the constant semi-elasticity of funding supply.

Now we solve for banks' optimal size when the regulation is present.[21] Regulation creates a discrete jump in regulatory costs as shown in the red dotted line in Figure 2.3a. Banks can avoid the regulation by bunching below the threshold. However, bunching is costly because banks give up profits they could have earned if they operated on their undistorted scale. The cost of bunching increases with the productivity z, as more-productive banks need to give up more assets to bunch, as shown by the blue dashed line in Figure 2.3a. As shown in Figure 2.3b, banks just above the regulatory threshold find it more profitable to bunch because they only need to shrink by a little bit. Banks whose undistorted size far exceeds the regulatory threshold find it more profitable to operate at their undistorted scale. Formally, we can derive the optimal asset choice as a function of productivity:

$$q^*(z) = \begin{cases} \underline{q}_i & z \in \left[\underline{z}_i, \overline{z}_i\right] \\ q_0(z) & z \notin \cup \left[\underline{z}_i, \overline{z}_i\right] \end{cases}, \tag{2.4}$$

where $\underline{z}_i$ is the productivity of a bank whose undistorted assets are equal to the regulatory threshold

$$\underline{q}_i = q_0(\underline{z}_i), \tag{2.5}$$

[21] We assume the regulatory thresholds are distant enough from each other so that banks only consider whether to avoid the nearest threshold.

and $\overline{z}_i$ is the productivity of a marginal bank that is indifferent between bunching at the regulatory threshold or paying the regulatory costs,

$$\overline{q_i} \equiv q_0(\overline{z}_i). \tag{2.6}$$

The indifference condition of the marginal bank is given by:

$$(R - r(\overline{q_i}|\overline{z}_i)) \exp(\overline{q_i})(1 - \tau_i) = (R - r(\underline{q_i}|\overline{z}_i)) \exp(\underline{q_i}). \tag{2.7}$$

Banks with productivity between $\underline{z}_i$ and $\overline{z}_i$ bunch at the regulatory threshold, $\underline{q}_i$, while banks outside the bunching interval operate in their optimal scale.[22] Using the above indifference condition of the marginal bank, equation (2.7), we can derive a sufficient statistic formula (Chetty, 2009) for the regulatory costs.

Proposition 1: *The direct regulatory costs τ_i that come into effect at threshold i is given by the following sufficient statistic formula:*

$$\tau_i \simeq 1 - \left(\overline{q_i} - \underline{q_i} + 1\right) \exp\left(\underline{q_i} - \overline{q_i}\right), \tag{2.8}$$

where the approximation is exact when the funding supply function has a constant semi-elasticity.

Proof: See Appendix B.2.

Equation (2.8) shows that the regulatory cost τ_i only depends on the difference between the marginal bank's log assets, $\overline{q_i}$, and the regulatory threshold, $\underline{q_i}$. It does not depend on the lending and deposit rates, R and r. The intuition is the following. Banks trade off the costs of regulation and the costs of bunching, which are expressed as a percentage of banks' profits. Although a higher lending rate (or a lower deposit rate) increases banks' profits, it does not change the relative magnitude of regulatory costs and the costs of bunching. In other words, if our goal is to estimate

[22] In our current model, banks choose their size to avoid regulation. It is equivalent to recast the model as banks choosing their funding rates to avoid regulation. Specifically, because the funding supply is upward sloping, banks can pay a low rate so that their size becomes smaller.

the tax-equivalent costs of regulation borne by banks, we only need the size distribution of banks as the data input. However, if we would like to translate the regulatory costs to a dollar value, we would need the lending and deposit rates to compute banks' profits.

It is worth noting that the sufficient statistic formula (2.8) still works when banks have heterogeneous semi-elasticities.[23] Intuitively, because the regulatory cost is modeled as a proportional tax, the semi-elasticity parameter is canceled out from the indifference condition, equation (2.7). Heterogeneous semi-elasticities matter only when we calculate the dollar value of regulatory costs. Such heterogeneity can be captured by multiplying the proportional tax rate with bank-specific profits.

2.3.2 Firms

We now endogenize the lending rate by introducing the firm sector, which borrows capital from banks and produces output using a Cobb-Douglas production function:

$$\max_{K} \Pi = AK^{\alpha} - RK, \tag{2.9}$$

where $Y = AK^{\alpha}$ is the output, K is the capital, and A is the total factor productivity.

The aggregate supply of capital is given by summing the credit supply among banks in the economy:

$$K^{s}(R) \equiv N \int \exp(q^{*}(z|R))g(z)dz, \tag{2.10}$$

where N is the number of banks, $g(z)$ is the probability density function of banks' productivity, and $q^{*}(z|R)$ is banks' optimal size choice defined in equation (2.4).

The equilibrium lending rate R is determined by the market-clearing condition in the lending market:

$$K^{s}(R) = \left(\frac{R}{A\alpha}\right)^{\frac{1}{\alpha-1}}. \tag{2.11}$$

[23] See Appendix B.3 for detailed derivation.

2.3.3 Bank entry and the distribution of productivity

We endogenize the equilibrium distribution of banks' productivity $g(z)$ by introducing growth, entry, and exit. Banks' productivity z evolves following a Brownian motion: $dz_t = \mu_z dt + \sigma_z dB_t$. The value function v of a bank with a current productivity z_0 is defined by

$$v\left(z_0\right) \equiv \mathbb{E}\left[\int_0^{\infty} e^{-(\rho+\lambda)t} \pi\left(q^*(z_t)|z_t\right) dt \mid z_0\right], \tag{2.12}$$

where ρ is the discount rate, λ is the exogenous exit rate, and π is the profits resulting from banks' optimal size choice as defined by equation (2.1).

New banks endogenously enter the economy. The entry rate is given by the following condition

$$m = \overline{m} \exp\left(\eta\left(\int v(z)\psi(z)dz - c_e\right)\right), \tag{2.13}$$

where c_e is the entry costs, $\psi(z)$ is the distribution of potential entrants' productivity, η is the entry elasticity, and $\overline{m}$ is the long-run entry mass when the expected value of entry equals entry costs. For simplicity, we assume the new entrants have the same starting productivity of z_n. The distribution of the productivity evolves according to the following Kolmogorov forward equation:

$$\frac{\partial g\left(z,t\right)}{\partial t} = -\frac{\partial}{\partial z}\left[\mu_z g\left(z,t\right)\right] + \frac{1}{2}\frac{\partial^2}{\partial z^2}\left[\sigma_z^2 g\left(z,t\right)\right] - \lambda g\left(z,t\right) + \frac{m}{N}\psi\left(z\right). \tag{2.14}$$

A stationary equilibrium exists in this economy, which is defined by bank value function $v(z)$, the probability density function of productivity $g(z)$, the number of banks N, the equilibrium lending rate R, and the aggregate capital K such that:

1. Incumbent banks optimally choose their credit supply given by equation (2.4).
2. Potential entrants optimally choose to enter the economy according to equation (2.13).
3. Firms optimally choose their credit demand given by equation (2.9).
4. Aggregate credit supply equals aggregate credit demand.
5. The distribution of banks reaches steady states $\frac{\partial g(z,t)}{\partial t} = 0, \forall z$.

Assuming the semi-elasticity of funding supply is constant, we can show that bank assets in the stationary equilibrium follow a power-law distribution.[24] The intuition is that bank assets follow proportional random growth in the absence of regulatory distortion, which generates a power-law distribution in the stationary equilibrium (Gabaix, 2016).

2.3.4 Indirect cost of regulation

We define the indirect costs of regulation in the stationary equilibrium, $\tau_{indirect}$, as the percentage output change between the regulated and unregulated economy:

$$\tau_{indirect} = \frac{Y(0) - Y(\tau)}{Y(0)}. \tag{2.15}$$

where $Y(\tau)$ and $Y(0)$ are the equilibrium firm output in the regulated and unregulated economies, respectively:

$$\begin{aligned} Y(\tau) &= A\left(N(\tau)\int \exp(q^*(z|R(\tau),\tau))g(z|\tau)dz\right)^{\alpha}, \\ Y(0) &= A\left(N(0)\int \exp(q^*(z|R(0),0))g(z|0)dz\right)^{\alpha}. \end{aligned} \tag{2.16}$$

The above equation shows that the indirect costs of regulation can arise from both the intensive and extensive margins. First, regulation reduces the credit supply from incumbent banks, q^* (intensive margin). Second, regulation can affect banks' entry incentive, and consequently, the total number of banks, N (extensive margin). Note that the sign of the second effect is ambiguous. On the one hand, as regulation reduces the incumbent banks' credit supply, the incentive to enter may increase. On the other hand, because new entrants may eventually grow bigger and face the regulation, heightened regulation may reduce their incentive to enter. Therefore, whether regulation increases or decreases bank entry is an empirical question.

[24] See Appendix B.4 for the proof of the stationarity and the derivation of the stationary distribution.

2.4 Estimation

We estimate the model in two stages. First, we estimate the direct regulatory costs using a maximum likelihood estimator. Second, we estimate the indirect regulatory costs by comparing the total output of bank-dependent firms in counterfactual economies with and without regulation. Separating the estimation into two stages clarifies what assumptions and data are needed to quantify the direct and indirect costs of regulation. Estimating the direct costs of regulation only requires the data on the size distribution of banks as shown in Section 2.3.1. However, estimating the indirect costs requires stronger assumptions and more parameters on the production functions of banks and firms. We calibrate these additional parameters to either the values in the prior literature or the corresponding moments in the data. We also discuss how these assumptions can affect the estimates of indirect costs in Section 2.4.4.

Note that we elect not to use the changes in bank values or bank entries before and after the Dodd–Frank Act to estimate regulatory costs. Although these two moments are related to the regulatory burden on banks, they can be driven by other factors such as a low-interest-rate environment, market perception of risks, and the size of the too-big-to-fail subsidy. Instead, we use the distortion in the size distribution of banks around the Dodd–Frank thresholds, which is unlikely to be driven by other factors as discussed in Section 2.2.2. Using the estimated regulatory costs from the bunching patterns, we can then quantify how much of the reduction in market valuation of bank shares and the entry rate can be explained by the regulation.

2.4.1 Data

We combine two main data sources: the Consolidated Reports of Condition and Income ("Call Reports," for commercial banks that are not part of a bank holding company) and the Consolidated Financial Statements for Holding Companies ("FRY-9C reports," for bank holding companies). Since the Dodd–Frank Act applies to the highest holding entity, we only consider the total consolidated assets for the estimation.[25] We thus refer to a standalone commercial bank and a bank

[25]See article 12 CFR Â§ 252.13.

holding company as a bank. The sample period is 2001–2019. We exclude banks with assets less than \$1 billion from the sample.

Table 2.1 provides the summary statistics for our sample. The sample covers around 40,000 bank-quarter observations. The average asset size is \$28 billion. The mean and standard deviation of the annual asset growth rate are 7.7% and 8.7%, respectively. The average profits per dollar of assets is 1.6%. We also report a set of regulation-related administrative expense items, including legal, data processing, advisory, printing and supplies, auditing, communications, and labor expenses. Note that we do not use these expense items in our structural estimation because these expense items do not capture all regulatory costs. Instead, we will use these expense items when we compare our structural method with other methods of estimating regulatory costs in Section 2.4.5. In our sample, the average administrative expenses are 0.2 cents per dollar of assets. The number of employees per million of assets is 0.2. The average salaries are 1.6 cents per dollar of assets.

2.4.2 Stage 1: Direct regulatory costs

We now illustrate how to estimate regulatory costs from the distortion in the size distribution. Following Garicano et al. (2016), we assume log assets are observed with a structural error, $a = q + u$, which follows a normal distributions: $u \sim N(0, \sigma^2)$. This structural error accounts for any empirical departures from the model. Specifically, the theoretical model implies a sharp bunching exactly at regulatory thresholds. However, the bunching pattern is more diffused in the data because (1) the empirical measure of bank size may not correspond to the exact definition of bank size for all the regulatory provisions, and (2) banks face random deposit flows and fluctuations in asset values.

We further assume the undistorted assets follow a power-law distribution: $\exp(q) \sim c \exp(q)^{-\beta}$. This assumption is consistent with the equilibrium property of our model and can be verified in the data. Specifically, Figure 2.4a plots the log frequency versus the log size of banks in the pre-Dodd–Frank sample. The relationship is close to a straight line, which suggests that the distribution can be well approximated by a power law (Gabaix, 2016).[26] We will examine the robustness of the results to this assumption in Section 2.5. We also assume that the semi-elasticity of funding supply is a constant locally around the threshold so that equation (2.8) holds exactly. We will examine the magnitude of the approximation error in Section 2.5.4.

We estimate the regulatory cost parameter using a maximum likelihood estimator:

$$\max_{\Theta} \mathcal{L}(\Theta) = \sum_{j=1}^{J} \ln f\left(a_j|\Theta\right), \tag{2.17}$$

where f is the likelihood function derived in Appendix B.5. a is the observed log assets. $\Theta = (\tau, \beta, \sigma)$ is the set of unknown parameters. τ is the direct regulatory cost. σ is the standard deviation of the structural error. β determines the curvature of the distribution of undistorted assets. Note that the scale parameter c of the power-law distribution enters the likelihood as a constant. Therefore, it is unidentified by the data.

Figure 2.5 illustrates the intuition of the estimation. We simulate the size distribution of banks around the regulatory threshold for different regulatory costs and standard deviations of the structural error. A larger regulatory cost τ leads to a larger abnormal mass around the regulatory threshold. A larger standard deviation of the structural error σ makes the abnormal mass more diffused around the threshold. The power-law exponent, β, is pinned down by the global shape of the distribution, that is, observations far away from the regulatory threshold. The goal of the maximum likelihood estimation is to find a set of parameters that can generate a similar distortion to what is observed in the data.

We conduct the estimation for each regulatory threshold separately using sub-samples of banks around the thresholds. Doing so is sensible because the thresholds are far apart, and the bunching ranges are unlikely to overlap. We use banks in the \$3 billion to \$40 billion intervals for the \$10 billion threshold, and banks above \$40 billion for the \$50 billion threshold.[27] The sample period

[26]Online Appendix Table B.1 conducts goodness-of-fit tests for the distribution assumption and find the power-law distribution fits well to the data.

[27]The results are robust to alternative sample ranges as shown in the Online Appendix Table B.2.

is from 2010Q3 to 2018Q2.[28]

We present the result of the maximum likelihood estimation in Table 2.2. The point estimate for the annual regulatory costs triggered by the \$10 billion threshold is 0.41% of average annual profits. The estimate is statistically significant with a standard error of 0.066%. Note that this estimate captures the additional regulatory cost imposed by the Dodd–Frank Act that is triggered at the \$10 billion threshold. It does not include the regulations already in place before the Dodd–Frank Act. Given this estimate, we calculate the undistorted assets of the marginal bank, $\exp(\overline{q})$, to be around \$11 billion using equation (2.8). Next, we examine the \$50 billion threshold in the second panel of Table 2.2. The point estimate of additional regulatory costs crossing the \$50 billion threshold is 0.11%. Adding the 0.41% regulatory costs triggered by crossing the \$10 billion threshold, the total regulatory cost imposed by the Dodd–Frank Act on banks above \$50 billion is 0.52%. We also calculate that the undistorted assets of the marginal bank at this threshold are around \$52 billion.

To put this number into perspective, we can calculate the dollar costs of the regulatory burden imposed by the Dodd–Frank Act on a bank with \$50 billion assets. Given the profits are around 1.6% of total assets (Table 2.1), our estimate implies a dollar value of regulatory costs around \$4.16 million per year. Our estimate is equivalent to the annual expense of hiring 52 additional compliance officers, assuming the average annual compensation for a compliance officer is around \$80,000 (Feldman et al., 2013). This estimate is also related to Kisin and Manela (2016), who estimate the shadow cost of bank capital requirements before the 2008 financial crisis to be around 0.4% of banks' annual profits. Although the exact regulations are different, the estimates are similar in order of magnitude.

2.4.3 Stage 2: Indirect regulatory costs

So far, we have estimated the direct costs of regulation on banks. We now calibrate other model parameters that affect the indirect costs of regulation on firms that borrow from banks. We calibrate

[28] The estimation is robust to adding pre Dodd–Frank data, as shown in Online Appendix Table B.3.

these parameters using the corresponding moments in the data or the values in the prior literature.

Table 2.3 presents the parameters used for counterfactual policy experiments. We first discuss the directly specified parameters. We set the regulatory costs triggered at \$10 billion and \$50 billion thresholds to 0.406 and 0.106 according to our estimates in Table 2.2. We normalize the mass of entry in the long-run $\overline{m}$ to 1. The curvature of the production function α is set to 0.3, as in Cooper, Haltiwanger, and Willis (2007). This parameter determines the aggregate elasticity of credit demand to the lending rate. We set the elasticity of entry η to a large number, 100, so that the equilibrium approximates the competitive entry in the long run (Moll, 2018).

We next discuss parameters calibrated to our data, for which we also report the targeted data moments and their model counterparts. The drift of productivity, μ_z, is calibrated to match the average asset growth rates in the pre-Dodd–Frank sample, which is 7.7%. The diffusion of productivity, σ_z, is calibrated to match the standard deviation of asset growth, 8.7%. We match the average and the standard deviation of asset growth rates exactly because there is a one-to-one correspondence between assets and productivity in the absence of regulation in the model. The exit rate λ is calibrated to match the value observed in the data, 4.4%.[29] We assume the funding supply function is $r(q|z) = \frac{1}{\theta}(q - z)$ and calibrate the semi-elasticity of funding supply θ to match the average profit margin, which is 1.62% in the data. The subjective discount rate of banks is calibrated to match the average ratio of the market value of assets to the book value in the pre-Dodd–Frank data. The productivity of new entrants, z_n, is calibrated to match the size of new entrants observed in the data, which is \$0.25 billion. The total factor productivity A is calibrated to match an average lending rate of 5%. Intuitively, higher total factor productivity drives up the demand for credit and the equilibrium lending rate. The entry cost c_e is calibrated to match the average bank size in 2010Q3. Overall, the model does a reasonable job matching most of the targeted moments.

Next, we examine the quantitative fit of the model for untargeted moments in Table 2.4. We group banks in three size categories: small (<\$10 billion), median (\$10 billion to \$50 billion), and big (>\$50 billion). We calculate the share of banks and assets in each size category in both the

[29] This value includes exits due to bank failures and due to mergers and acquisitions. In Online Appendix Table B.4, we evaluate the robustness of our results to alternative calibration of the exit rates.

data and the model. We find that the model generates a highly skewed size distribution in the data. For instance, big banks account for 4% of banks but 80% of assets in the data. The corresponding moments in the model are 5% and 87%, respectively.

We further examine whether the model can match the bunching pattern around the regulatory thresholds. We construct an imbalance ratio, defined as the number of banks in a narrow interval above the threshold divided by the number of banks in a narrow interval below the threshold (e.g., \$10 billion-\$11 billion vs. \$9 billion-\$10 billion).[30] A reduction in the imbalance ratio indicates that the number of banks above the threshold becomes more scarce relative to that below the threshold. In the data, the imbalance ratio at \$10 billion has decreased from 1.09 to 0.42 from the pre- to the post-Dodd–Frank sample. The model generates a similar decrease in the imbalance ratio. Similar pattern can be found for the \$50 billion threshold. Finally, the model also matches the imbalance ratios at other round numbers that are not regulatory thresholds, which do not increase systemically after the Dodd–Frank Act.

Figure 2.4 compares the model-predicted stationary probability density function with that in the data. The model fits the data quite well over the entire distribution. We further zoom in to the region around the regulatory thresholds in Figure 2.6. Note that the probability density function is quite noisy in small samples. Consequently, we plot the cumulative distribution function instead. We find that the model generates similar departures from the power-law distribution to the data.

We use the calibrated model to quantify the indirect cost of regulation. We first simulate a baseline economy without the Dodd–Frank Act in column 1 of Table 2.5. Then we simulate an economy with the Dodd–Frank Act and present the percentage change with respect to the baseline economy in column 2 of Table 2.5. We first examine how the Dodd–Frank Act affects the total number of banks in the stationary equilibrium. In theory, the Dodd–Frank Act can have two countervailing forces on bank entry. On the one hand, the regulation increases the profits of potential entrants because incumbent banks around the regulatory thresholds reduce their lending to avoid the regulation. On the other hand, the prospect of facing tightened regulation in the future for

[30] We use a \$1 billion window for \$10, \$20, and \$30 billion thresholds, and \$2 billion window for \$40 and \$50 billion thresholds to be consistent with the bunching range estimated in Table 2.2.

the potential entrants reduces their incentive to enter. The simulation suggests that the later force dominates. The total mass of banks decreases by 0.18% after the introduction of the Dodd–Frank Act. This prediction is consistent with the falling entry rates after the Dodd–Frank Act observed in the data, as shown in Figure 2.7a.

Practitioners often argue that high regulatory costs imposed by the Dodd–Frank Act leads to the decline in bank franchise values, as shown in Figure 2.7b. Although the decline in franchise values is consistent with heightened regulatory costs, it could be driven by many other factors such as ultra-low interest rate environment (Calomiris and Nissim, 2014; Whited, Wu, and Xiao, 2021), market reassessment of risks (Sarin and Summers, 2016), and the removal of too-big-to-fail subsidies (Atkeson, d'Avernas, Eisfeldt, and Weill, 2019; Berndt, Duffie, and Zhu, 2020). As a result, we abstain from using the change in the bank franchise value to estimate regulatory costs. Instead, we use the regulatory costs estimated from bunching to evaluate how much the decline in bank values can be attribute to the Dodd–Frank Act. We find the estimated regulatory costs lead to a 0.22% decrease in bank values, which can only account for 3% of the overall decline in bank franchise value in the post-crisis period. Our estimate suggests non-regulatory factors may have contributed to the decline in bank values.

Next, we investigate the distributional effects of the Dodd–Frank Act in the cross-section of banks. This question is motivated by the fact that big banks are subject to tighter regulation according to the Dodd–Frank Act. We find that big banks indeed suffer a greater decrease in profits than medium and small banks, as shown in Panel (b) of Table 2.5. But, surprisingly, the heightened regulation on big banks does not translate into smaller market shares. Instead, the share of big banks increases in the stationary equilibrium with the Dodd–Frank Act. The reason is twofold. First, medium banks (\$10 billion to \$50 billion) engage in regulatory avoidance, which reduces their average asset size, as shown in Panel (c). Second, heightened regulatory costs depress bank valuation across the size spectrum, which reduces the entry of small banks (<\$10 billion), as shown in Panel (d). This result does not imply that the Dodd–Frank Act has worsened the too-big-to-fail problem because we do not explicitly model how regulation affects big banks' risk-taking and their reliance on the implicit government guarantee. However, it does suggest a possible unintended consequence of regulation on the bank industry's market structure.

Finally, we examine the indirect costs of regulation on bank-dependent firms. We find that the Dodd–Frank Act increases the equilibrium lending rate by 0.046%, and decreases the lending quantity by 0.065%. Overall, the Dodd–Frank Act decreases the total output of bank-dependent firms by around 0.020%. These results suggest that the indirect regulatory costs are modest. One may worry that our estimates of the indirect regulatory costs may be sensitive to the parameters that we use to simulate the counterfactuals. In Table B.4 of the Online Appendix, we simulate the baseline and counterfactual economies using alternative values of the discount rate, exit rate, and entry cost. We find the estimated indirect regulatory costs with alternative parameters are in the same order of magnitude as our baseline estimate.

2.4.4 Discussion of indirect regulatory costs

Although the above exercise provides a novel quantification of the indirect costs of the Dodd–Frank Act, it is also subject to several caveats. First, we assume that banks have the same business model and lend to the same representative firm. Although this assumption simplifies the model, it also prevents us from studying the heterogeneity of indirect costs. It is possible that the impact of regulation is disproportionately felt by certain firms, such as those that borrow from the bunching banks. Second, in the above exercise, we assume that entrants can choose either to be a bank or not to enter the market at all. In reality, entrants can also choose to be a shadow bank. This choice could be important because banks experienced heightened competition from shadow banks since the financial crisis and subsequent shifts in the regulatory regime (Buchak et al., 2018b). Introducing shadow banks could affect the estimates in two potential ways. On the one hand, since firms can switch from banks to shadow banks, introducing shadow banks could lower the estimated indirect costs of regulation. On the other hand, the migration of activity to less regulated shadow banks could increase the fragility of the banking system, thus increase the indirect costs. Quantifying these indirect costs requires a more complex model and data on shadow banks, which we leave for future research. Finally, we define the indirect regulatory costs as the changes in outputs in two stationary equilibria, one with the Dodd–Frank Act and one without, following Garicano et al. (2016). In reality, the banking sector may take time to reach the new equilibrium. During the transition period, the indirect regulatory cost could differ from that in the stationary equilibrium.

2.4.5 Comparisons with existing methods

In this section, we compare our structural estimates with the existing methods of estimating regulatory costs. Note that the cost estimates surveyed in this section are about the direct compliance costs for banks. Therefore, they should be compared with the direct regulatory costs estimated in Section 2.4.2, which are 0.41%-0.52% of banks' profits.

Survey

One of the most commonly used methods of estimating regulatory costs is conducting surveys. We provide a summary of surveys on the regulatory costs of the Dodd–Frank Act for banks in Table 2.6. Many surveys only provide a qualitative indication of high regulatory costs but no quantitative estimates. Among surveys that do provide quantitative estimates, the magnitudes differ greatly. For instance, a survey conducted by the *Bank Director Magazine* suggests that the annual regulatory costs are around 9.9% of banks' annual profits. In contrast, a survey conducted by the American Action Forum estimates that the regulatory costs are around 1.8%. The survey estimates are usually much larger than the estimates obtained from our revealed preference approach.

Difference-in-differences

Some researchers also use the difference-in-differences approach to estimate regulatory costs. For instance, Hinkes-Jones (2017) estimates regulatory costs by comparing legal and administrative expenses of affected banks with those of unaffected banks. This methodology faces two challenges. The first challenge is measurement. Banks are not required to report legal or other selected noninterest expenses that do not exceed defined thresholds that have changed over time. Nevertheless, some banks report them even if they fall short of the thresholds. These issues may create difficulties for researchers that attempt to draw conclusions from trends in these expense items. Furthermore, not all regulatory costs are reflected as expenses in banks' income statements. For instance, an increased capital requirement is costly but does not directly lead to higher expenses. The second challenge is endogeneity. Because banks can endogenously bunch below the threshold and avoid the regulatory costs, a difference-in-differences regression may lead to downward biases.

To illustrate the difficulty of using the difference-in-differences approach to estimate regulatory costs, we apply this methodology to estimate the regulatory costs imposed by the Dodd–Frank Act on banks with more than \$10 billion in assets. The regression model is as follows:

$$Expenses_{i,t} = \alpha_i + \alpha_t + \beta Treat_i \times Post_t + \gamma X_{i,t} + \varepsilon_{i,t}, \tag{2.18}$$

where $Expenses$ includes regulation-related expenses such as legal, data processing, advisory, printing, stationery and supplies, auditing, and communication costs. α_i and α_t are bank and time fixed effects, respectively. $X_{i,t}$ includes the log number of branches. The expenses are normalized by assets. We also include the number of employees and total salaries to account for the possibility that regulations may force banks to hire more compliance officers. The sample includes banks with assets between \$3 billion and \$40 billion as of 2010Q2. The sample period is from 2003Q1 to 2018Q2. $Treat$ is a dummy equal to 1 if the bank's total assets is above \$10 billion as of 2010Q2, and 0 otherwise. $Post$ is a dummy that equals 1 for years after the Dodd–Frank Act (after 2010Q3).

Table 2.7 reports the results. None of the expenses increases significantly for the treated banks. Some expenses, such as the communication expenses, actually have the wrong sign. This result is confirmed by plotting the rolling four-quarter average annualized expenses for the treated and control groups in Figure 2.8, which shows no significant increase for the treated group after the Dodd–Frank Act. However, these results should not be interpreted as evidence that the Dodd–Frank Act imposes no costs on banks because of the measurement issue and endogeneity concern discussed earlier.

Regression discontinuity

Another methodology that can be used to estimate regulatory costs is regression discontinuity design. This methodology is subject to the measurement and endogeneity concerns as discussed in Section 2.4.5. Nevertheless, we apply this approach to banks around the \$10 billion threshold using the following regression model:

$$Expenses_{i,t} = \beta_0 + \beta_1 \mathbf{1}_{\left\{Q_{i,t} \geq \underline{Q}\right\}} + \beta_2 f\left(Q_{i,t} - \underline{Q}\right) + \beta_3 \mathbf{1}_{\left\{Q_{i,t} \geq \underline{Q}\right\}} f\left(Q_{i,t} - \underline{Q}\right) + \varepsilon_{i,t}, \tag{2.19}$$

where $Expenses_{i,t}$ include the same set of regulation-related expenses as in Section 2.4.5; $Q_{i,t}$ is the assets; $f(\cdot)$ is a polynomial function of degree one. β_1 is the increase in the regulation-related expenses when a bank's assets exceed the regulatory threshold, $\underline{Q}$. The sample period is from 2010Q3 to 2018Q2. Table 2.8 presents the estimated coefficients, and Figure 2.9 shows the results graphically. Again, most expenses do not exhibit significant changes above and below the thresholds. There is a small increase in the auditing costs for treated banks, but a small decrease in the communication costs.

Summary

The existing methods lead to inconsistent evidence on the regulatory costs of the Dodd–Frank Act. Surveys on banks typically suggest extremely large regulatory costs, consistent with the anecdotal evidence that survey estimates may be inflated to lobby regulators for regulatory relief (Hinkes-Jones, 2017; Parker, 2018b). In contrast, reduced-form methods such as difference-in-differences and regression discontinuity designs find virtually no evidence of additional regulatory costs due to the Dodd–Frank Act, with the caveats of the endogeneity and measurement issues. Our structural approach finds a modest level of regulatory costs.

We view our approach as complementary to the existing approaches to evaluate the impacts of regulation. The reduced-form approaches do not require researchers to model the data generating process. Instead, they require at least some observations to be randomly assigned. However, for big and highly anticipated regulatory changes, regulated parties are likely to engage in regulatory avoidance so observations are rarely randomly assigned. These selection issues pose challenges for the reduced-form methods. While such selection issues could be addressed by finding valid instruments, it does limit the scope of the analysis to settings in which instruments are available. In comparison, our structural approach allows us to analyze equilibrium responses to big and highly anticipated regulatory changes. With additional assumptions on the data generating process, we can conduct counterfactual policy experiments and analyze welfare implications.

2.5 Extension and Robustness

Our revealed preference approach addresses some challenges faced by the existing methods, such as the lack of reliable data and the endogeneity concern. Nevertheless, this approach does require some structural assumptions, and the robustness of the results to these assumptions should be duly assessed. In this section, we conduct a few extensions and robustness checks on our results.

2.5.1 Regulatory relief in 2018

In 2018, the U.S. Congress passed the Economic Growth, Regulatory Relief, and Consumer Protection Act, partially reversing some regulations of the Dodd–Frank Act. For instance, banks above $10 billion are no longer required to conduct stress tests except for those above $250 billion assets.[31] This regulatory relief presents a good validity check for our approach because the estimated regulatory costs should decrease in the post-relief period. Indeed, we find that the estimated regulatory costs at the $10 billion threshold decreases by around 50% relative to the pre-relief value, while the estimated regulatory costs at the $50 billion threshold almost completely disappeared, as shown in Table 2.9.[32] These estimates are consistent with the visual evidence in

[31] See "Amendments to the Stress Testing Rule for National Banks and Federal Savings Associations".

[32] Because there is only one year of data after the regulatory relief, we fix the exponent of the power law distribution and the structural error volatility to the baseline estimation value as these parameters are quite persistent.

Figure 2.10 that the abnormal densities at the Dodd–Frank regulatory thresholds have significantly reduced after the relief. Note that there is still considerable amount of bunching after the regulatory relief because some regulations such as the Durbin Amendment is still in place after the relief.

Then, we examine the effect of the 2018 regulatory relief on the indirect cost of regulation in column 3 of Table 2.5. We find that the regulatory relief alleviates the reduction in the number of banks. This result is consistent with the recovery in bank entries after 2018, as shown in Figure 2.7a. Similarly, the total output partially recovers but is still below the pre-Dodd–Frank Act equilibrium. Finally, the 2018 regulatory relief also partially reverses the distributional effects of the Dodd–Frank Act.

2.5.2 Alternative distribution assumption

In our baseline estimation, we assume that the undistorted assets follow a power law distribution. This assumption is consistent with the equilibrium property of our model and the data. Nevertheless, we evaluate the sensitivity of our results to this assumption. To this end, we re-estimate the regulatory costs assuming the undistorted assets follow a log-normal distribution.[33] Table 2.10 presents the results. We find that the estimated regulatory costs are similar using this alternative distribution assumption. Our estimates are insensitive to the distribution assumption because regulatory costs are mainly identified by the local abnormal densities.

2.5.3 Placebo tests

To further examine the robustness of our estimation, we conduct several placebo tests. First, we apply our maximum likelihood estimator to the pre-Dodd–Frank sample in Panels (a) and (b) of Table 2.11. We set the standard deviation of the structural error σ to the value from our baseline estimations because it is difficult to identify this parameter when there is no abnormal density created by regulation.[34] Nevertheless, the results are not sensitive to the value of this parameter.

[33] The derivation of the maximum likelihood estimator under log-normal distribution can be found in Appendix B.6.

[34] Intuitively, this parameter measures the extent to which the abnormal densities are diffused around the threshold, as shown in Figure 2.5. Therefore, it is difficult to identify this parameter if there is no abnormal density in the first place.

The results in Panels (a) and (b) show that banks incur no additional regulatory cost when they cross the \$10 billion or \$50 billion thresholds before the Dodd–Frank Act. We also apply our maximum likelihood estimator to round numbers that are not regulatory thresholds, such as \$20 billion and \$40 billion. Again our estimator correctly indicates null results as shown in Panels (c) and (d).

2.5.4 Approximation error

Proposition 1 shows that the sufficient statistic formula (2.8) is exact when the semi-elasticity does not depend on bank size choice, q. However, if the semi-elasticity is size-dependent, then the sufficient statistic formula (2.8) leads to an approximation error of $\frac{1}{2(R-r(\overline{q}|\overline{z}))} r_{qq} \left(\underline{q} - \overline{q}\right)^2 \exp\left(\underline{q} - \overline{q}\right)$.[35] We evaluate the magnitude of this approximation error using additional moments from the data. Specifically, the marginal buncher for the \$10 billion threshold has a size of \$11 billion according to Table 2.2. We calibrate the change in the inverse semi-elasticity, r_{qq}, to -0.2% using the data around the marginal buncher.[36] The approximation error at the \$10 billion threshold is 0.039%, which is small relative to the baseline estimate of the regulatory cost, 0.406%. This exercise suggests that the approximation error is economically small because the semi-elasticity does not seem to change substantially in the small bunching range.

2.5.5 Transition dynamics

Our baseline estimation assumes that banks' size distribution has already reached the stationary equilibrium one year after the Dodd–Frank Act. However, if there are convex adjustment costs, the bank size distribution may take more time to adjust to its new steady state. As a result, the still-in-progress changes in the size distribution may bias our estimates downwards. To address this concern, we first plot bank size distribution year by year in Figure B.2. We can see that the bunching pattern starts to appear immediately after the introduction of the Dodd–Frank Act. This

[35] See Appendix B.2 for derivation.

[36] We use a bin of \$1 billion centering around \$10 billion and \$11 billion in the pre-Dodd Frank sample to estimate the average profit margins at these two sizes, which are 2.3% and 2.1%, respectively. This implies that the change in the inverse semi-elasticity, r_{qq}, for the marginal buncher is -0.2%.

bunching pattern suggests that bank size adjustment is quite fast, possibly because banks have liquid assets such as marketable securities and short-term liabilities such as wholesale funding, which allows the adjustment to happen quickly.

To further address the concern on the still-in-progress changes in the size distribution, in Table 2.12, we drop the first one, two, and three years post the Dodd-Frank Act and re-run the estimation. We find that the estimated regulatory costs are quite similar to the baseline estimates when the whole post-Dodd–Frank sample is included. This result again suggests that bank size adjustment is relatively fast.

In addition to the lagged response, a related issue is anticipatory responses. If banks face convex adjustment costs and cannot adjust their size quickly, they may change their size before the regulation comes into effect, which may bias the estimates if the pre-regulation data are used to estimate the counterfactual distribution. To address the possible anticipatory responses, we elect to use only the post-Dodd–Frank sample in our baseline estimation. Nevertheless, the results are robust if the pre-Dodd–Frank sample is included, as shown in Table B.3.

2.5.6 Alternative measurement of bank regulatory assets

As discussed in section 2.2.3, different provisions of the Dodd–Frank Act use slightly different methodologies to measure bank size to determine banks' regulatory status. Because there is no uniform way to measure size for all the provisions, we use the simplest approach—quarter-end assets—to measure bank size. In this section, we evaluate the sensitivity of the estimated regulatory cost to three alternative measures of bank assets: 1) average of last four quarters, 2) end of last calendar year, 3) minimum of the last four quarters in Table B.5. We find very similar estimates of the regulatory costs as our baseline results. The reason is twofold: (1) these measures only differ slightly from each other in the data, and (2) we explicitly account for measurement error in the maximum likelihood estimation. Indeed, the differences between alternative measures mainly show up in the estimated standard deviation of the measurement error, σ, rather than the regulatory costs.

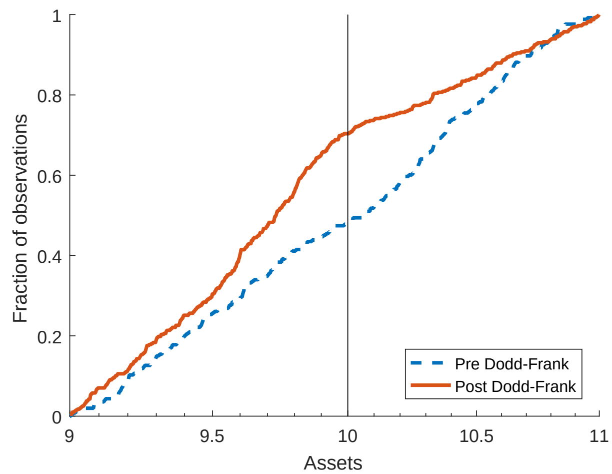
2.6 Conclusion

In this paper, we propose a revealed preference approach to estimate regulatory costs borne by financial institutions. By focusing on banks' actions rather than their self-reported estimates, our approach circumvents the information obstacle faced by regulators. We use our approach to estimate the costs of the Dodd–Frank Act. We find that the regulation costs triggered at the \$10 billion threshold are equivalent to a 0.41% tax on banks' average annual profits. The regulatory costs triggered at the \$50 billion threshold are equivalent to a 0.11% tax. In total, the regulatory costs introduced by the Dodd–Frank Act for a \$50 billion bank amount to \$4.16 million per year. We also examine the impacts of regulation on bank franchise value, entry and exits, and the output of bank-dependent firms. Although our estimated regulatory costs are substantial, they are significantly lower than those claimed by banks. Our estimates of regulatory costs can be used in conjunction with estimates of regulatory benefits in CBA. Overall, this paper shows that regulators can extract valuable information by analyzing financial institutions' endogenous response to regulations. This revealed preference approach can be used in many settings in which the regulatory changes are big and highly anticipated.

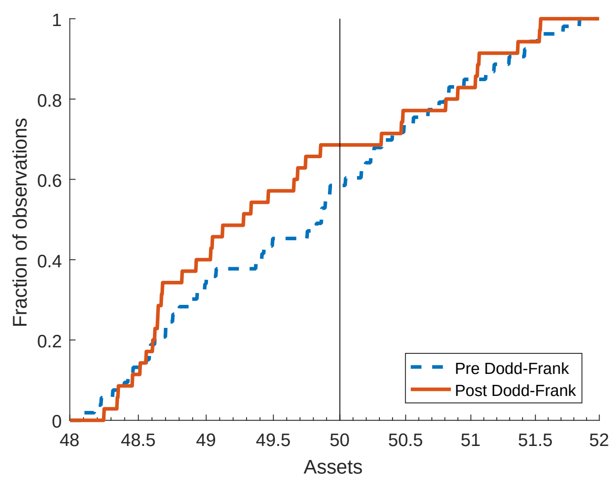
2.7 Figures and Tables

Figure 2.1: The Size Distribution of Banks Around Regulatory Thresholds Before and After Dodd–Frank

(a) $10 billion



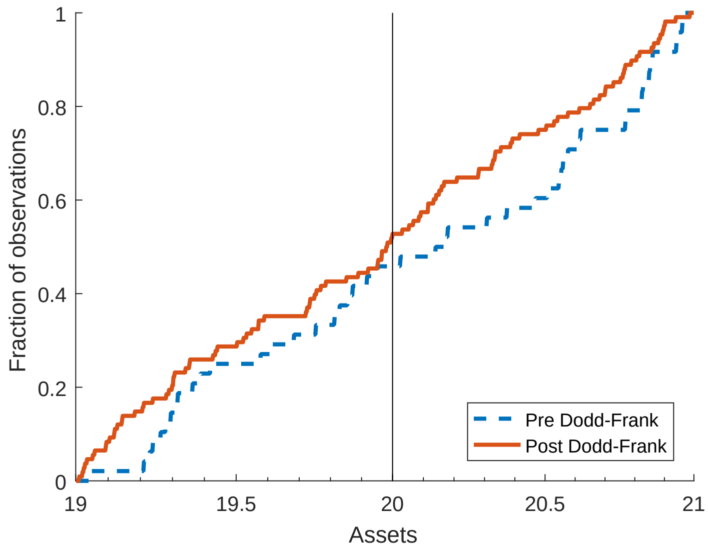
(b) $50 billion



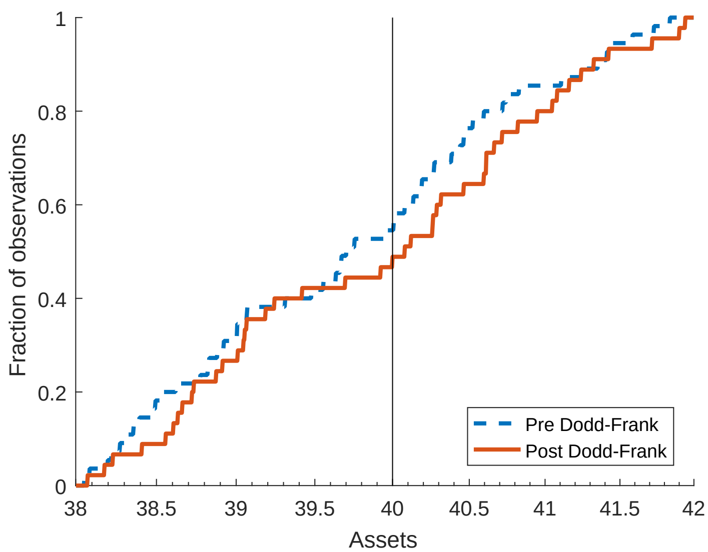
Note: This figure shows the empirical cumulative distribution of bank size conditional on the narrow intervals around the $10 billion and $50 billion regulatory thresholds. Bank size is measured by quarter-end assets. The blue dashed line and the red solid line represent the pre Dodd–Frank period (2001Q1 to 2007Q3) and post Dodd–Frank period (2010Q3 to 2018Q2), respectively.

Figure 2.2: The Size Distribution of Banks Around Non-regulatory Thresholds Before and After Dodd–Frank

(a) $20 billion



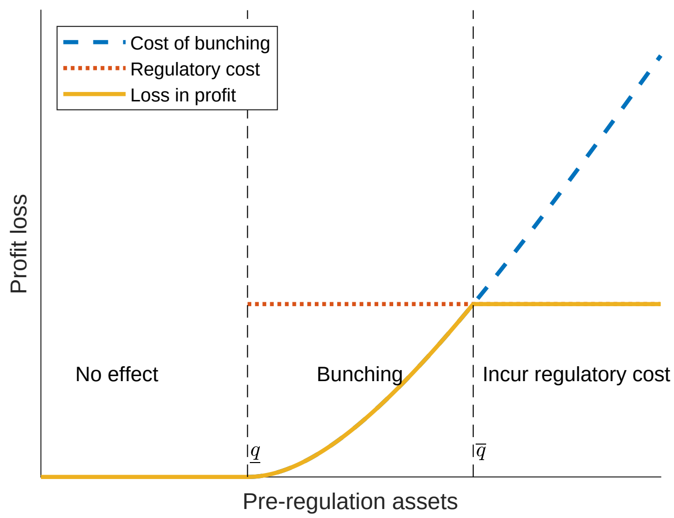
(b) $40 billion



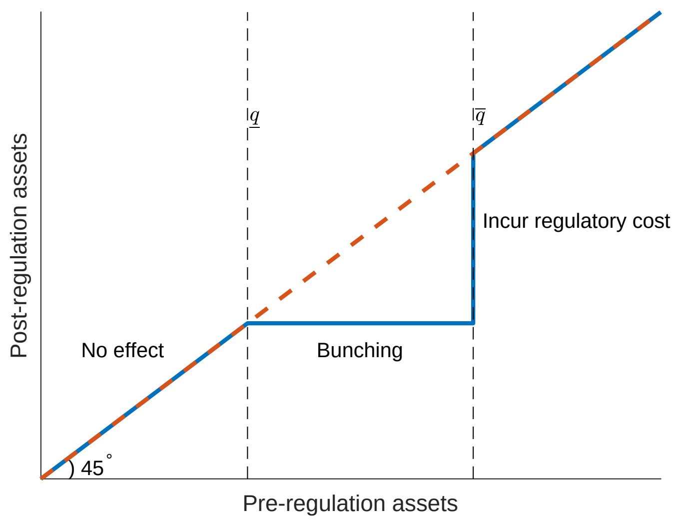
Note: This figure shows the empirical cumulative distribution of bank size conditional on the narrow intervals around $20 billion and $40 billion. Bank size is measured by quarter-end assets. The blue dashed line and the red solid line represent the pre Dodd–Frank period (2001Q1 to 2007Q3) and post Dodd–Frank period (2010Q3 to 2018Q2), respectively.

Figure 2.3: Effects of Threshold-Based Regulation on Bank Assets and Profits

(a) Effects on profits



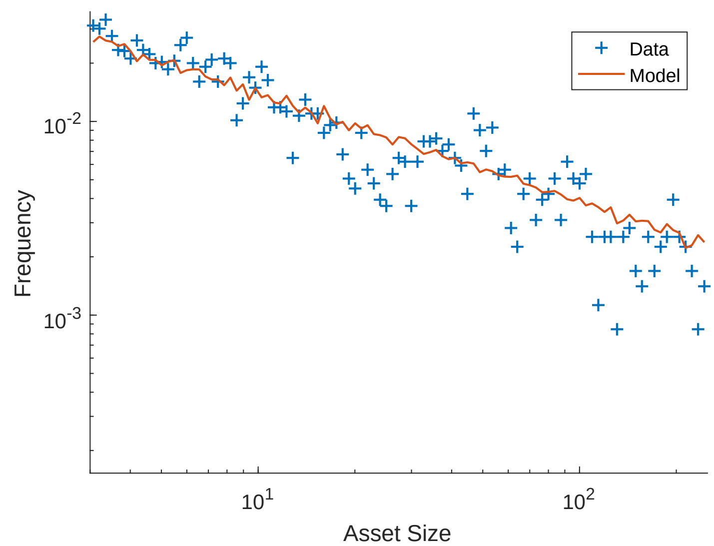
(b) Effects on assets



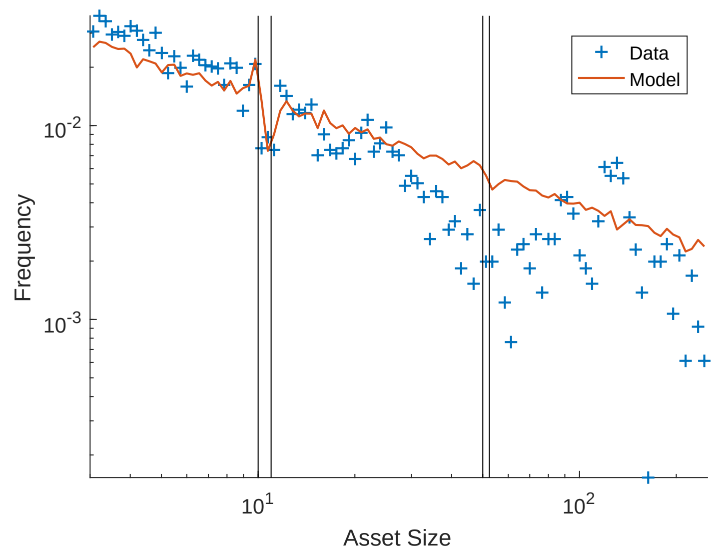
Note: The top panel shows regulatory costs and costs of bunching as functions of pre-regulation assets. The bottom panel shows the post-regulation assets as a function of pre-regulation assets.

Figure 2.4: The Size Distribution of Banks: Data vs. Model

(a) Pre Dodd–Frank Act

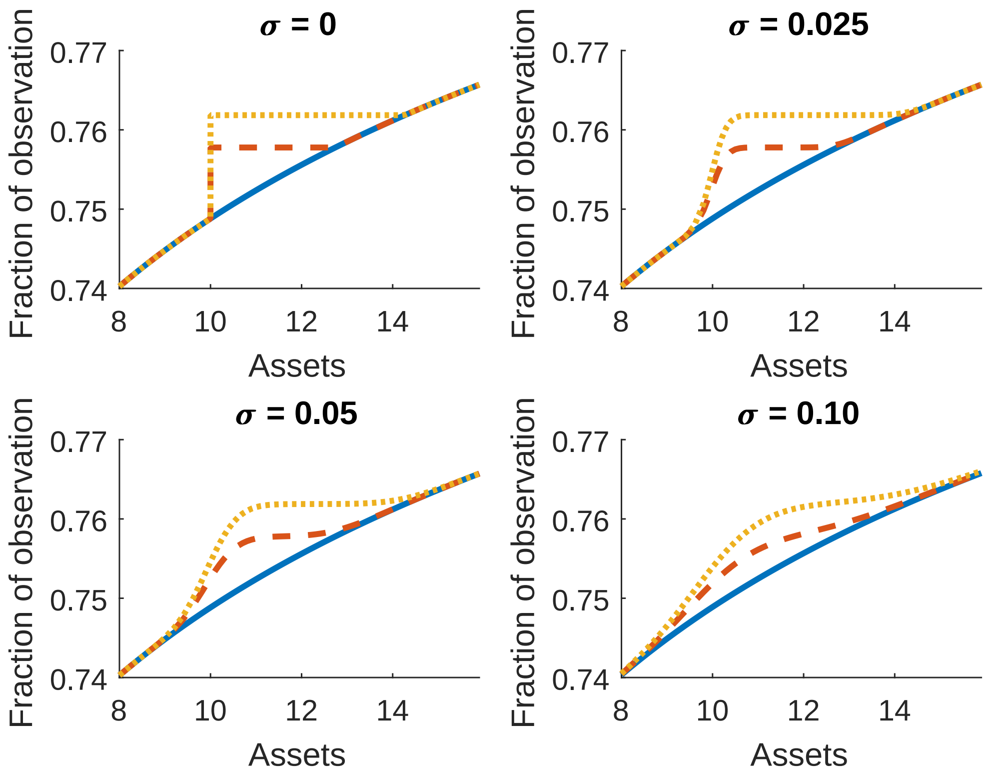


(b) Post Dodd–Frank Act



Note: This figure shows the probability density of bank size in the model and in the data before and after the Dodd–Frank Act, respectively. Both axes are in log scale. The vertical lines indicate the bunching range around the $10 and $50 billion thresholds. Bank size is measured by quarter-end assets. The sample period is from 2010Q3 to 2018Q2 for the post Dodd–Frank period, and from 2001Q1 to 2007Q3 for the pre Dodd–Frank period.

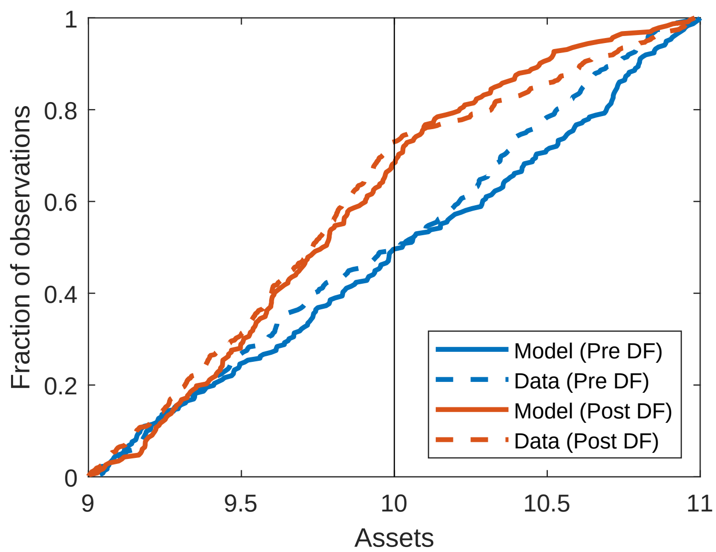
Figure 2.5: Theoretical Size Distribution Around Regulatory Thresholds



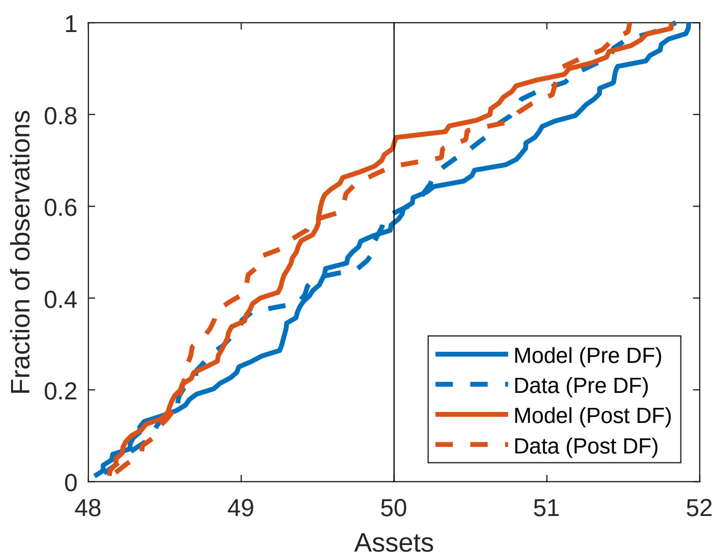
Note: This figure shows theoretical cumulative distributions of bank size under different values of regulatory costs τ and standard deviation of the structural errors σ. The solid, dashed, and dotted lines represent the case of $\tau = 0$, $\tau = 0.025$, and $\tau = 0.05$, respectively.

Figure 2.6: Simulated Size Distribution Around Regulatory Thresholds

(a) $10 billion



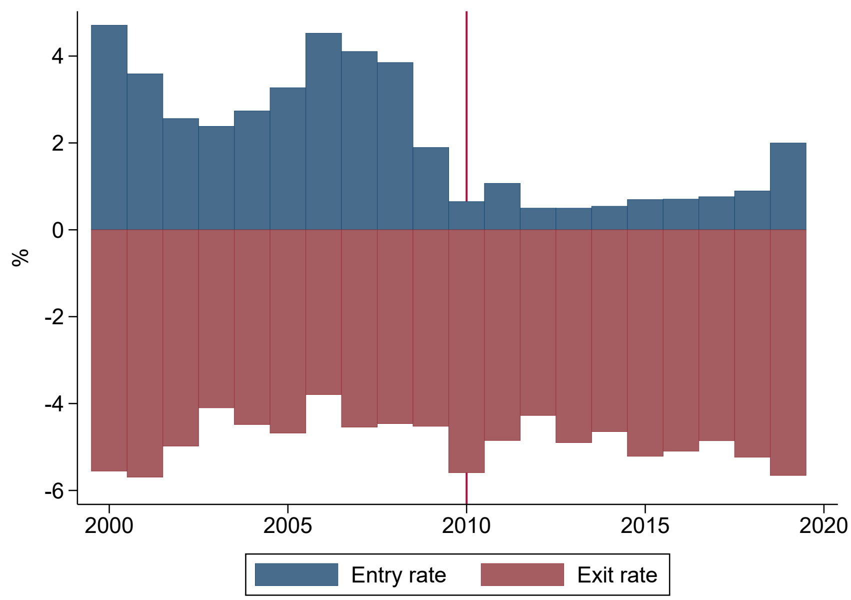
(b) $50 billion



Note: This figure shows the model-simulated and empirical distributions of bank size conditional on the narrow intervals around the $10 billion and the $50 billion thresholds before and after the Dodd–Frank Act, respectively.

Figure 2.7: Entry, Exit, and Bank Valuation

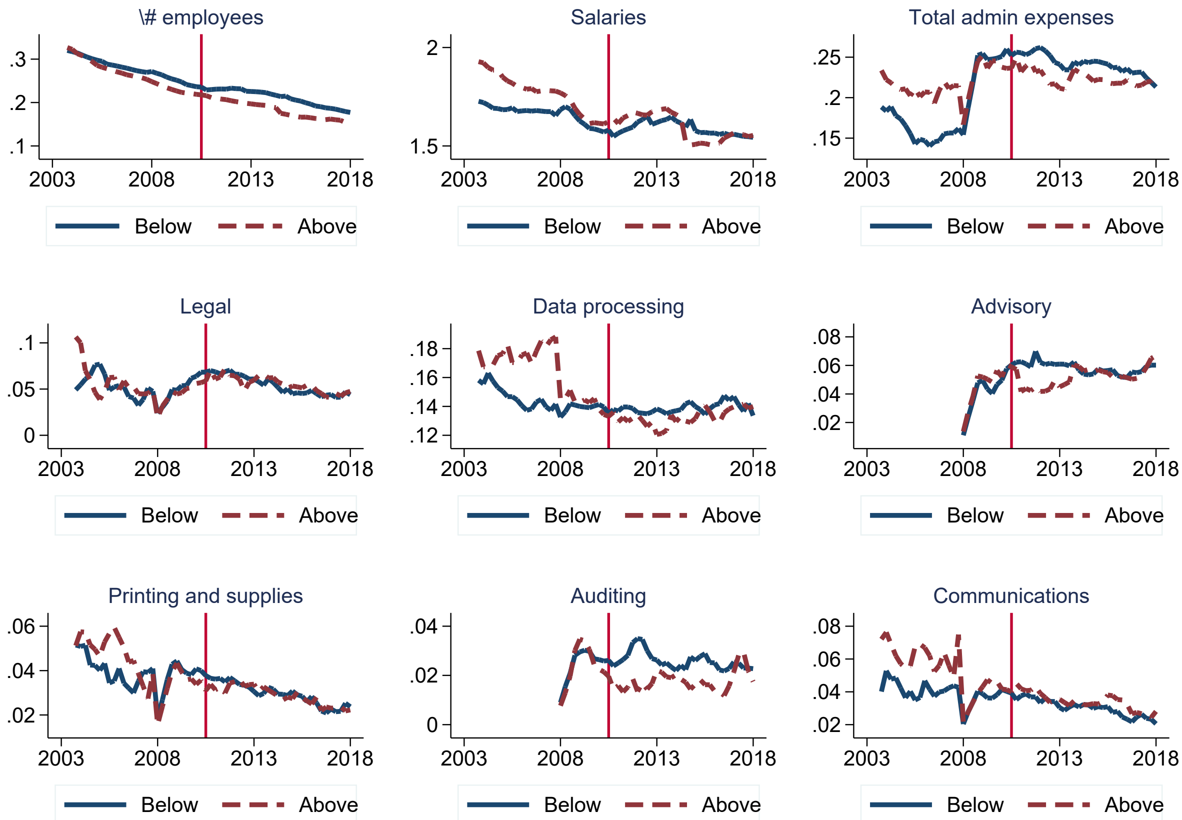
(a) Entry and exit rates



(b) Market-to-book ratio

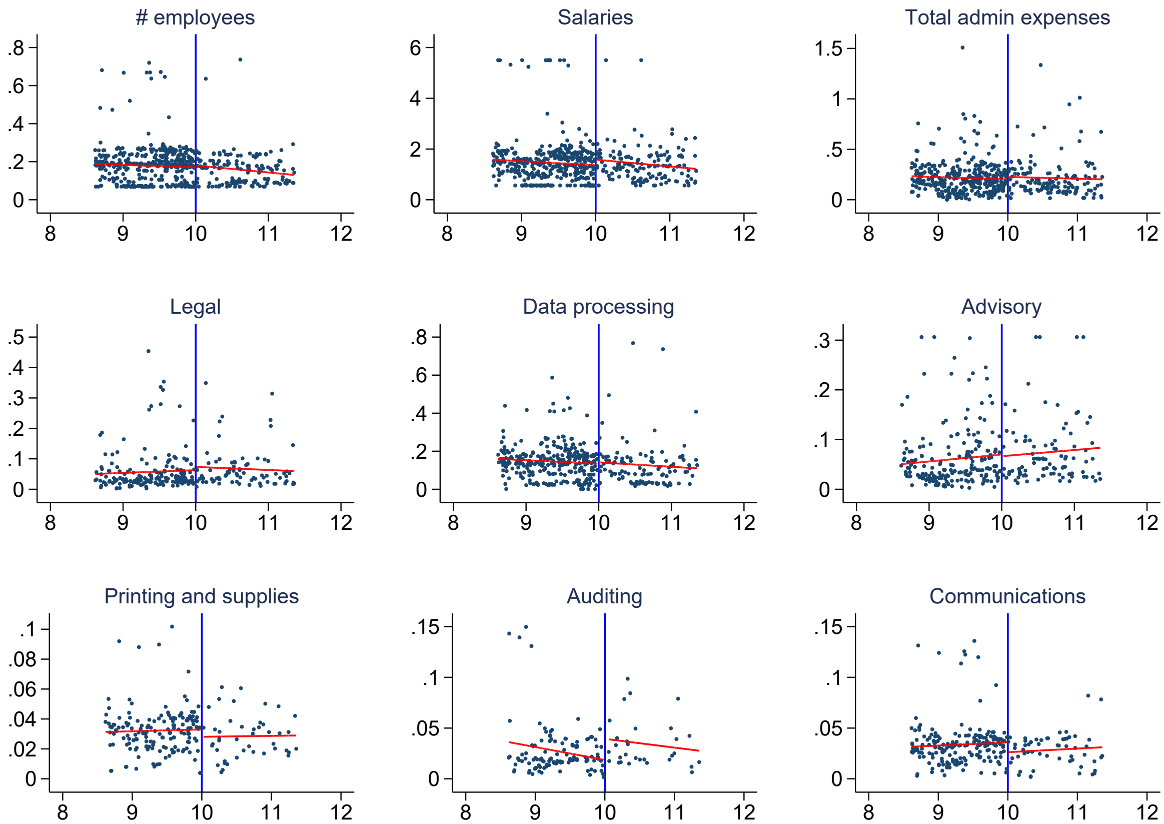
Note: Figure 2.7a shows entry and exit rates of U.S. banks before and after the Dodd–Frank Act. The entry (exit) rate is calculated by dividing the number of banks that enter the market (drop out from the market) in a year by the total number of banks in that year. Figure 2.7b shows the average ratio of market value of assets to the book value of assets for U.S. publicly traded banks.

Figure 2.8: Regulation-related Expenses for Banks Above and Below $10 Billion



Note: This figure presents the regulation-related expenses over time. Banks are classified as "below" ("above") if their total consolidated assets are between $3 billion and $10 billion ($10 billion and $40 billion) at the start of the Dodd–Frank Act in 2010. The dependent variables are normalized by assets.

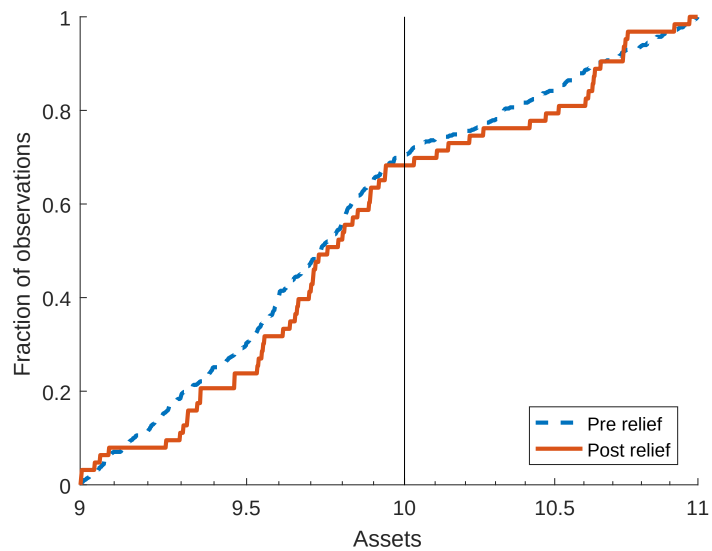
Figure 2.9: Regulation-related Expenses Around the Regulatory Threshold



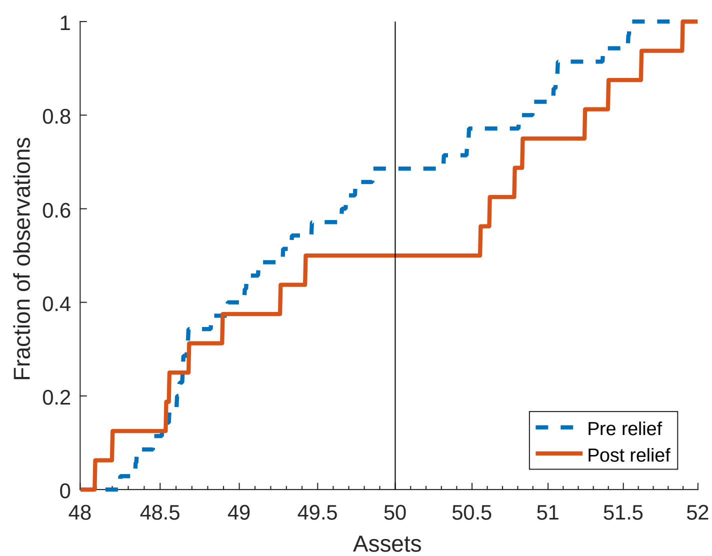
Note: This figure presents the scatter plot of regulation-related expenses against bank size around the $10 billion threshold. The sample includes all the bank-quarter observations from 2010Q3 to 2018Q2. The red lines represent the predicted values using the regression of equation (2.19). The dependent variables are normalized by assets.

Figure 2.10: The Size Distribution of Banks Around Regulatory Thresholds Before and After 2018 Regulatory Relief

(a) $10 billion



(b) $50 billion



Note: This figure shows the empirical cumulative distribution of bank size conditional on the narrow intervals around the $10 billion and $50 billion regulatory thresholds. The blue dashed line and the red solid line represent the cumulative distribution functions before and after the regulatory relief in 2018Q2. Bank size is measured by quarter-end assets. The sample period is 2018Q3 to 2019Q4 for the post-relief period, and from 2010Q3 to 2018Q2 for the pre relief period.

Table 2.1: Summary Statistics

	N	mean	sd	p5	p25	p50	p75	p95
Assets	39228	27.905	163.363	1.056	1.376	2.232	5.845	81.776
Assets growth rate	33616	7.655	11.504	-7.081	1.484	5.809	11.632	29.310
Assets growth rate (sd)	38578	8.668	5.071	2.385	4.736	7.663	11.719	18.100
Profits	39228	1.624	1.328	0.223	1.074	1.503	1.948	3.063
Total admin expenses	39228	0.201	0.184	0.000	0.083	0.173	0.278	0.517
Legal	39228	0.026	0.051	0.000	0.000	0.000	0.037	0.129
Data processing	39228	0.101	0.110	0.000	0.000	0.083	0.149	0.295
Advisory	39228	0.028	0.055	0.000	0.000	0.000	0.039	0.132
Printing and supplies	39228	0.015	0.027	0.000	0.000	0.000	0.029	0.065
Auditing	39228	0.012	0.024	0.000	0.000	0.000	0.022	0.059
Communications	39228	0.013	0.025	0.000	0.000	0.000	0.025	0.059
# employees	39228	0.217	0.106	0.070	0.148	0.206	0.271	0.398
Salaries	39228	1.618	0.812	0.736	1.235	1.514	1.804	2.705

Note: This table reports the summary statistics of all the U.S. banks from 2001 to 2019. We use the highest-level ownership as the observation unit because the Dodd–Frank Act applies to both the bank level and bank holding company (BHC) level. The data is quarterly. Assets are reported in billions. Profits and expenses are annualized and reported as a percentage of the assets. The number of employees is reported as per million of assets.

Table 2.2: Maximum Likelihood Estimation of Regulatory Costs

Panel A: $10 billion threshold

		Estimated value	S.E.
β	Exponent of the power law distribution	1.112	[0.001]
σ	Measurement error volatility (in %)	4.258	[0.386]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	10.973	[0.086]
τ	Cost of regulation (% of profit)	0.405	[0.066]

Panel B: $50 billion threshold

		Estimated value	S.E.
β	Exponent of the power law distribution	1.083	[0.002]
σ	Measurement error volatility (in %)	2.290	[0.498]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	52.393	[0.517]
τ	Cost of regulation (% of profit)	0.106	[0.046]

Note: This table presents the results of the maximum likelihood estimation of the regulatory costs using the power law distribution. The sample period is from 2010Q3 to 2018Q2. The estimation sample of Panel A includes banks with assets between $3 billion and $40 billion. The estimation sample of Panel B includes banks with assets above $40 billion. Standard errors are obtained via the inverse of the Hessian matrix.

Table 2.3: Calibrated Parameters for Policy Experiment

Parameter	Value	Target	Data	Model
μ_z: productivity growth	7.700	Asset growth	7.700	7.700
σ_z: productivity diffusion	8.700	Asset growth std	8.700	8.700
λ: exit rate	4.400	Exit rate	4.400	4.400
θ: semi-elasticity of funding supply	61.728	Profit margin	1.620	1.620
ρ: discount rate	7.000	Market-to-book	1.100	1.228
z_n: productivity of new entrants	-3.470	Size of entrants	0.250	0.239
A: total factor productivity	8.000	Lending rate	5.000	4.924
c_e: entry costs	0.120	Average bank size	19.067	21.781
$\overline{m}$: mass of entry in the long-run	1.000	Normalize to one	-	-
α: curvature of production function	0.300	Cooper et al. (2007)	-	-
η: elasticit of entry	100.000	Moll (2018)	-	-
τ_{10}: regulatory cost at \$10B	0.406	Table 2	-	-
τ_{50}: regulatory cost at \$50B	0.106	Table 2	-	-

Note: Columns 1-2 of this table report the calibrated parameters for the counterfactual simulations. Columns 3-5 report the targeted moments in the data and their model counterpart.

Table 2.4: Model Fit: Untargeted Moments

	Pre-Dodd–Frank		Post-Dodd–Frank	
Moments	Data	Model	Data	Model
Share of banks: small banks	89.609	87.787	89.318	87.777
Share of banks: medium banks	6.569	7.223	7.065	7.230
Share of banks: big banks	3.822	4.990	3.617	4.993
Share of assets: small banks	10.385	5.714	9.557	5.711
Share of assets: medium banks	10.472	7.327	7.556	7.311
Share of assets: big banks	79.142	86.958	82.887	86.978
Imbalance ratio: $10 billion	1.091	0.881	0.421	0.558
Imbalance ratio: $20 billion	1.182	0.897	0.895	0.909
Imbalance ratio: $30 billion	1.444	1.032	1.240	1.001
Imbalance ratio: $40 billion	0.833	1.012	1.045	1.069
Imbalance ratio: $50 billion	0.710	0.925	0.458	0.560

Note: This table reports the untargeted moments in the data and their model counterpart. Banks are categorized as small (<$10 billion), medium ($10 billion to $50 billion), and big (>$50 billion). The imbalance ratio is defined as the number of banks in a narrow interval above the threshold divided by the number of banks in a narrow interval below the threshold (e.g. $10 billion-$11 billion vs. $9 billion-$10 billion). We use a $1 billion window for $10, $20, and $30 billion thresholds, and $2 billion window for $40 and $50 billion thresholds. The choice of the window size is consistent with the estimated marginal bank in Table 2.2.

Table 2.5: Counterfactual Simulation

Counterfactual	Baseline	Dodd-Frank	Regulatory relief
		Panel (a): all banks	
Mass of banks	11.836	-0.184 %	-0.094 %
Market-to-book	1.228	-0.221 %	-0.104 %
Lending quantity	257.805	-0.065 %	-0.032 %
Lending rate	0.049	0.046 %	0.023 %
Output	42.313	-0.020 %	-0.010 %
		Panel (b): annual profits by size group	
Small banks	0.023	0.068 %	0.033 %
Medium banks	0.358	-0.399 %	-0.221 %
Big banks	6.150	-1.268 %	-0.593 %
		Panel (c): asset shares by size group	
Small banks	0.057	-0.061 %	-0.034 %
Medium banks	0.073	-0.216 %	-0.123 %
Big banks	0.870	0.022 %	0.013 %
		Panel (d): shares of banks by size group	
Small banks	0.878	-0.012 %	-0.006 %
Medium banks	0.072	0.089 %	0.042 %
Big banks	0.050	0.075 %	0.042 %

Note: This table reports the results of counterfactual simulations. Column (1) reports the equilibrium quantities and prices in the baseline economy without the Dodd–Frank Act. Columns (2) and (3) report the simulated economies with the Dodd–Frank Act and the 2018 regulatory relief, respectively. The values are reported as percentage changes with respect to the baseline economy.

Table 2.6: Estimated Regulatory Costs of Dodd–Frank on Banks Based on Surveys

Source	Sample	Estimate	Date
Bank Director Magazine	Survey of 10 banks	9.9	1/1/2017
American Action Forum	Estimation from Federal Register	1.8	7/1/2016
JPMorgan and Citigroup	Survey of 2 banks	0.9	2012, 2014
Federal Reserve Bank of Minneapolis	Estimation of cost of new hires	1.1	3/1/2013
Bank Director and Grant Thornton LLP Survey	Survey of 130 senior executives	Qualitative	6/1/2013
KPMG 2013 Community Banking Survey	Survey of 100 senior executives	Qualitative	10/1/2013
Florida Chamber Fundation	Survey of 75 banks	Qualitative	7/1/2012
Mercatus Center's Small Bank Survey	Survey of 200 banks	Qualitative	2/1/2014
Risk Management Association Survey	Survey of 230 senior executives	Qualitative	3/1/2013

Note: This table presents a list of surveys on the regulatory costs of the Dodd–Frank Act on banks. For sources with quantitative estimates, we translate the estimates into percentages of net income.

Table 2.7: Regulatory Costs of the Dodd–Frank Act Estimated by Difference-in-Differences

	(1) # employees	(2) Salaries	(3) Total admin expenses	(4) Legal	(5) Data processing	(6) Advisory	(7) Printing	(8) Auditing	(9) Com-munications
Treat * Post	-0.012	-0.043	-0.012	0.008	0.001	-0.003	0.001	0.000	-0.009**
	[0.012]	[0.081]	[0.018]	[0.008]	[0.014]	[0.008]	[0.004]	[0.005]	[0.004]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,816	6,816	5,931	3,135	5,067	2,429	2,476	1,323	2,643
Adj. R-squared	0.515	0.039	0.115	0.063	0.006	0.037	0.197	0.043	0.302

Note: This table presents the estimated regulatory costs from the difference-in-difference regression shown in equation (2.18). The sample includes U.S. banks with assets between 3 billion and 40 billion at the time of the introduction of Dodd-Frank and covers the period of 2003Q1–2018Q2. The dependent variables are normalized by assets. *Post* is a dummy equal to 1 for all years after the Dodd–Frank Act (after 2010Q3). *Treat* is a dummy equal to 1 if the bank's total consolidated assets are above $10 billion as of 2010Q2, and 0 otherwise. Robust standard errors clustered at the bank level are showed in brackets. All regressions include bank and time fixed effects, and control for the log of number of bank branches.

Table 2.8: Regulatory Costs of the Dodd–Frank Act Estimated by Regression Discontinuity

	# employees	Salaries	Total admin expenses	Legal	Data processing	Advisory	Printing	Auditing	Com-munications
Treatment Effect	0.010	0.222	0.025	0.011	0.007	-0.003	-0.005	0.021^{**}	-0.010^{*}
	[0.024]	[0.182]	[0.030]	[0.021]	[0.023]	[0.015]	[0.005]	[0.010]	[0.005]
Observations	559	559	505	256	400	296	203	148	256
Adj. R-squared	0.011	0.004	-0.004	-0.005	0.006	0.008	-0.003	0.028	0.003
Bandwidth	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38	1.38

Note: This table presents the estimated regulatory costs from the regression discontinuity design shown in equation (2.19). The sample includes U.S. banks with assets between 8.62 and 11.38 billion from 2010Q3 to 2018Q2. The dependent variables are normalized by assets. The treatment effect is the increase in the regulation-related expenses when a bank's assets exceed the regulatory threshold. Robust standard errors clustered at the bank-level are showed in brackets.

Table 2.9: Maximum Likelihood Estimation of Regulatory Costs: Post 2018 Regulatory Relief

$10 billion threshold

		Estimated value	S.E.
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	10.701	[0.159]
τ	Cost of regulation (% of profit)	0.219	[0.095]

$50 billion threshold

		Estimated value	S.E.
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	50.138	[3.707]
τ	Cost of regulation (% of profit)	0.003	[0.012]

Note: This table presents the results of the maximum likelihood estimation of the regulatory costs using the power law distribution. The sample period is from 2018Q3 to 2019Q4. The estimation sample of Panel A includes banks with assets between $3 billion to $40 billion. The estimation sample of Panel B includes banks with assets above $40 billion. Standard errors are obtained via the inverse of the Hessian matrix.

Table 2.10: Maximum Likelihood Estimation of Regulatory Cost: Log Normal Distribution

Panel A: $10 billion threshold

		Estimated value	S.E.
μ_q	Average undistorted log(asset)	1.999	[0.007]
σ_q	Undistorted log(asset) std.	0.674	[0.005]
σ	Measurement error volatility (in %)	4.027	[0.378]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	10.888	[0.086]
τ	Cost of regulation (% of profit)	0.342	[0.062]

Panel B: $50 billion threshold

		Estimated value	S.E.
μ_q	Average undistorted log(asset)	5.040	[0.022]
σ_q	Undistorted log(asset) std.	1.062	[0.016]
σ	Measurement error volatility (in %)	2.355	[0.533]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	52.552	[0.533]
τ	Cost of regulation (% of profit)	0.120	[0.048]

Note: This table presents the results of the maximum likelihood estimation of the regulatory costs using the log normal distribution. The sample period is from 2010Q3 to 2018Q2. The estimation sample of Panel A includes banks with assets between $3 billion to $40 billion. The estimation sample of Panel B includes banks with assets above $40 billion. Standard errors are obtained via the inverse of the Hessian matrix.

Table 2.11: Regulatory Cost Estimates: Placebo Tests

Panel A: $10 billion threshold, pre Dodd–Frank

		Estimated value	S.E.
β	Exponent of the power law distribution	1.112	[0.002]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	10.002	[0.239]
τ	Cost of regulation (% of profit)	0.005	[0.003]

Panel B: $50 billion threshold, pre Dodd–Frank

		Estimated value	S.E.
β	Exponent of the power law distribution	1.085	[0.003]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	51.221	[0.717]
τ	Cost of regulation (% of profit)	0.029	[0.034]

Panel C: $20 billion threshold, post Dodd–Frank

		Estimated value	S.E.
β	Exponent of the power law distribution	1.112	[0.001]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	20.016	[0.618]
τ	Cost of regulation (% of profit)	0.004	[0.003]

Panel D: $40 billion threshold, post Dodd–Frank

		Estimated value	S.E.
β	Exponent of the power law distribution	1.085	[0.002]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	40.509	[1.546]
τ	Cost of regulation (% of profit)	0.008	[0.071]

Note: This table presents the results of the maximum likelihood estimation of the regulatory costs using the power law distribution. The sample periods of Panels A and B are from 2001Q1 to 2007Q4. The sample periods of Panels C and D are from 2010Q3 to 2017Q4. The estimation sample of Panels A and C includes banks with assets between $3 billion and $40 billion. The estimation sample of Panel B includes banks with assets above $40 billion. The estimation sample of Panel D includes banks with assets above $30 billion. Standard errors are obtained via the inverse of the Hessian matrix.

Table 2.12: Maximum Likelihood Estimation of Regulatory Costs at $10 billion: Transition Dynamics

Panel A: Drop first year

		Estimated value	S.E.
β	Exponent of the power law distribution	1.112	[0.001]
σ	Measurement error volatility (in %)	4.236	[0.411]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	11.000	[0.090]
τ	Cost of regulation (% of profit)	0.426	[0.071]

Panel B: Drop first two years

		Estimated value	S.E.
β	Exponent of the power law distribution	1.112	[0.002]
σ	Measurement error volatility (in %)	4.255	[0.453]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	11.040	[0.097]
τ	Cost of regulation (% of profit)	0.458	[0.079]

Panel C: Drop first three years

		Estimated value	S.E.
β	Exponent of the power law distribution	1.112	[0.002]
σ	Measurement error volatility (in %)	4.027	[0.543]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	10.998	[0.110]
τ	Cost of regulation (% of profit)	0.425	[0.087]

Note: This table presents the results of the maximum likelihood estimation of the regulatory costs using the power law distribution. The sample period is 2011Q3-2018Q2, 2012Q3-2018Q2, and 2013Q3-2018Q2 for panels A, B, and C, respectively. The estimation sample includes banks with assets between $3 billion and $40 billion. Standard errors are obtained via the inverse of the Hessian matrix.

Chapter 3

Fuzzy Bunching[1]

3.1 Introduction

Bunching is ubiquitous in finance. For instance, DeFusco and Paciorek (2017) and Buchak, Matvos, Piskorski, and Seru (2018a) document that mortgage borrowers bunch at the conforming loan threshold to take advantage of lower interest rates. Bachas, Kim, and Yannelis (2020) document that small business lenders bunch at Small Business Administration (SBA) regulatory threshold to qualify for government guarantees. Iliev (2010) document that public firms bunch below regulatory thresholds of public float to avoid disclosure and internal governance regulations required by the SEC. Finally, Degeorge, Patel, and Zeckhauser (1999) document that managers systematically manage earnings to meet thresholds such as past earnings or industry benchmarks.

In recent years, there has been a growing interest in applying the bunching estimation technique developed in the public finance and labor literature (e.g., Saez (2010); Chetty, Friedman, Olsen, and Pistaferri (2011); Kleven and Waseem (2013)) to study finance topics. Bunching estimation complements traditional empirical methodologies such as regression discontinuity and difference-in-differences because it does not require random assignment around the threshold (Kleven, 2016). Instead, bunching estimation embraces endogenous selection and uses the extent of bunching around the threshold to infer underlying economic parameters.

Many finance applications, however, present new challenges for this methodology. The existing bunching approach exploits a sharp spike at the regulatory threshold in the bunching variable's

[1]This chapter is based on Alvero and Xiao (2020). We thank Samuel Antill, Tobias Berg, Michael Best, Matias Cattaneo, Ramona Dagostino, Olivier Darmouni, Anthony DeFusco, William Diamond, John Friedman, Leonard Goff, Wei Jiang, Ernest Liu, Yiming Ma, Claudia Robles-Garcia, Philipp Schnabl, Philip Strahan, Stefanie Stantcheva, Adi Sunderam, Paul Tetlock, and participants at the WFA, the SFS Cavalcade, the WFA-CFAR Conference at WUSTL, the Swiss National Bank Workshop, and the Columbia Finance Lunch Seminar for helpful comments and discussions.

probability density function, as shown in the top panel of Figure 3.1. In data with little noise, researchers can easily determine, typically through visual inspection, a narrow interval that contains the density spike. The abnormal bunching mass within the spike interval is then used as the key moment to identify the underlying economic parameters. In many finance settings, however, bunching is quite fuzzy. For instance, evidence has steadily accumulated that U.S. banks bunch below the $10 billion regulatory threshold to avoid regulations imposed by the Dodd–Frank Act of 2010.[2] However, unlike the classical bunching setting in which taxpayers can bunch their income at the tax threshold relatively easily, it is more difficult for banks to precisely bunch their size at the $10 billion threshold due to random inflows and outflows of deposits. In addition, bank size is also observed with large measurement errors because of the complex accounting rules in determining asset values. Due to these noises in the data, the bunching mass is diffused around the threshold, as shown in the top panel of Figure 3.2. Estimates of the sharp bunching approach are quite sensitive to the interval over which the bunching mass is calculated. Such fuzzy bunching patterns also arise in many other finance settings,[3] heightening the need for a formal approach to address this issue.

Another challenge for the sharp bunching approach in finance settings is the small sample size. When applying the sharp bunching approach, researchers need to estimate the empirical probability density function by discretizing the bunching variable into small bins and counting the fraction of observations falling into each bin. This step can be easily done in the classic bunching studies because they use administrative tax return data with millions of observations. However, in many finance settings, the sample size is quite small, making density estimation difficult. For instance, the U.S. bank data only have around 400 bank-quarter observations in the size interval of $9 billion to $11 billion between 2010 to 2017. The empirical probability density function constructed from this small sample is extremely noisy and sensitive to the bin size, as shown in the middle and bottom panels of Figure 3.2.

To address these challenges, this paper introduces a new fuzzy bunching approach. The key

[2] See Morgan and Yang (2016), Hou and Warusawitharana (2018), and Alvero, Ando, and Xiao (2022).

[3] See also Degeorge et al. (1999), Iliev (2010), Gourio and Roys (2014), Garicano, Lelarge, and Van Reenen (2016), and Ewens, Xiao, and Xu (2020) for other examples of fuzzy bunching patterns.

insight is that, instead of identifying bunching from the bunching mass in the probability density function, one can identify bunching from the area of a bulge in the cumulative distribution function, as shown in the lower panel of Figure 3.1. Intuitively, as agents bunch, some probability mass is shifted from above to below the threshold, creating a bulge around the threshold in the cumulative distribution function. This bunching area is strictly increasing to the bunching incentives but is invariant to the magnitude of the noise. Therefore, this moment can identify bunching regardless of the magnitude of noise. In contrast, the bunching mass used by the sharp bunching approach is a function of both the bunching incentive and the magnitude of the noise. Thus, one can only identify the extent of bunching unless there is little noise in the bunching variable using the sharp bunching approach.

The fuzzy bunching approach also offers a better way to visualize bunching in many finance settings because it works with the cumulative distribution rather than the probability density function. In the sharp bunching approach, the empirical probability density function has to be estimated using a density estimation, which requires a large sample size. Thus, in many finance settings with sparse firm-level data, the density estimation could be quite noisy or infeasible. In comparison, the fuzzy bunching approach works with the empirical cumulative distribution function, which is directly observable in the data. Indeed, as shown in Figure 3.3, the empirical cumulative distribution of the U.S. banks display a clear bulge around the $10 billion threshold, even if the empirical probability density function is quite fuzzy. Furthermore, using the empirical cumulative distribution function also eliminates the ambiguity in choosing tuning parameters such as the bin size for density estimation.

We illustrate the advantages of the fuzzy bunching approach in both simulated and real data. We first consider a simulated example in which banks face a size-based regulation. Some banks above the threshold bunch to avoid regulatory costs and the extent of bunching reveals the magnitude of regulatory costs. We apply both the sharp and fuzzy bunching approaches to the simulated bank size data to estimate the amount of assets that banks are willing to give up in order to avoid regulatory costs. When there is little noise in bank assets, both sharp and fuzzy bunching approaches can accurately identify the extent of bunching. However, when there is substantial noise, the sharp bunching approach tends to underestimate the extent of the bunching, while the fuzzy bunching approach still generates robust estimates. Moreover, the sharp bunching estimates are also sensitive to the tuning parameters, such as the width of the spike interval and the bin size. In comparison, these tuning parameters are not needed for the fuzzy bunching approach because it works with the cumulative distribution function.

We then apply the fuzzy bunching approach to real data. We first use the fuzzy bunching approach to replicate two classical bunching studies that use the sharp bunching approach: Saez (2010) and Kleven and Waseem (2013). These two papers use administrative tax return data with large numbers of observations, so both the sharp and fuzzy bunching approaches should be able to identify the bunching range in these settings. Indeed, we find that the fuzzy bunching approach generates similar results to the original estimates, which reassures the fuzzy bunching approach's validity.

Next, we apply the fuzzy bunching approach to the U.S. bank data studied by Alvero, Ando, and Xiao (2022). This dataset is much noisier than the administrative taxpayer data in Saez (2010) and Kleven and Waseem (2013). As a result, the fuzzy bunching approach should have a comparative advantage in this setting. Indeed, we find the fuzzy bunching estimate is consistent with the maximum likelihood estimate in Alvero et al. (2022). In comparison, the sharp bunching estimates are much smaller than the maximum likelihood estimate and are sensitive to the specifications of spike intervals and bin sizes.

Finally, we provide step-by-step guidance for applied researchers on how to implement the fuzzy bunching approach in finance research. We also discuss advanced issues such as how to handle non-optimizing agents and round-number bunching. We hope that this guide will be useful to academics in future research.

This paper first contributes to the bunching estimation literature. The bunching literature has long recognized the challenges introduced by data noise. For instance, in the seminal work that started the bunching literature, Saez (2010) noted that the bunching estimates could be sensitive to the choice of the spike interval around the Earned Income Tax Credit thresholds (Saez, 2010, Table 2, Panel B). Subsequent bunching works such as Chetty et al. (2011) address the problem by using a wider spike interval to calculate the diffused bunching mass. We show that this approach does not fully resolve the issue because a wider spike interval contains regions in which the actual probability density function falls below the counterfactual density function. Another strand of literature such as Garicano et al. (2016), Gourio and Roys (2014), and Alvero et al. (2022) address the data noise problem by estimating a full structural model where the magnitude of the noise is explicitly estimated. However, this approach requires assumptions on the data-generating process and is quite involved.[4] This paper addresses this issue following the sufficient statistic approach of the bunching literature, which is considerably simpler and involves fewer assumptions.[5]

This paper also contributes to a growing body of research that uses the bunching approach to finance topics.[6] The applications of the bunching approach to finance have stimulated several methodological innovations. First, the classical bunching studies in public finance usually estimate bunching using a single cross-section of data. DeFusco, Johnson, and Mondragon (2020) develop a new approach by leveraging the time-series dimension of the mortgage data to address the issue of extensive responses. Second, the classical bunching studies assume that agents can only manipulate a single dimension, taxable income. Bachas, Liu, and Morrison (2020) extend the classic one-dimensional bunching to a two-dimensional one to study the Small Business Administration (SBA) loan guarantees. Finally, the classic bunching approach usually uses a known discontinuity in policy cost to estimate elasticity related to preferences. In contrast, Alvero et al. (2022) shows that the bunching approach can be reversed to estimate the change in policy cost if the preference is given. Following the prior works in the bunching literature, this paper focuses on the identification

[4]The full structural approach used by Garicano et al. (2016) and Gourio and Roys (2014) allows rich counterfactual analysis, which cannot be done with our approach.

[5]The sufficient statistic approach is a mix of reduced-form and structural methods. It derives a small set of relationships from a model (or a class of models) that is robust to the data generating process and can be measured empirically (Chetty, 2009).

[6]Notable examples include DeFusco and Paciorek (2017); DeFusco, Johnson, and Mondragon (2020); Dagostino (2018); Bachas, Kim, and Yannelis (2020); Bachas, Liu, and Morrison (2020); Antill (2020). These studies use administrative loan-level data, which have little data noise.

of the bunching approach while sets aside the issue of statistical inference.[7] This allows us to clarify our key innovation is to propose a new moment to identify bunching in the presence of data noise. The standard errors of the fuzzy bunching estimates can be computed using the bootstrap method following the existing bunching literature.

The paper proceeds as follows. Section 3.2 introduces the fuzzy bunching approach. Section 3.3 illustrates this approach in simulated data. Section 3.4 applies this approach to three distinct real data sets. Section 3.5 considers extensions of the fuzzy bunching approach. Section 3.6 concludes.

3.2 Bunching estimation

We first present the conceptual framework of the bunching approach. Then, we discuss the sharp bunching approach and its challenges when the data are noisy. Finally, we introduce the new fuzzy bunching approach.

3.2.1 Payoff function and policy discontinuity

Bunching estimation considers a group of heterogeneous agents who choose a bunching variable q to maximize their payoff:

$$\max_{q} \pi(q; z, \theta) - \kappa(q), \tag{3.1}$$

where $\pi(q; z, \theta)$ is the policy-independent component of the payoff. z captures the heterogeneity across agents. Agents with a larger z choose larger q in the absence of regulatory distortion, as shown in the upper panel of Figure 3.4. θ captures the rate of payoff loss under choice distortion. Larger θ makes it more costly to deviate from the optimum, as shown in the lower panel of Figure 3.4. $\kappa(q)$ is the policy cost function, which features either a discrete change in either the level or the slope at a threshold, $\underline{q}$. If the policy cost function features a discrete change in the level of the incentive, so the threshold $\underline{q}$ is referred to as a notch point. If the policy cost function features

[7] See also Blomquist, Newey, Kumar, and Liang (2021), who also focus on identification of the bunching approach.

a discrete change in the slope of the incentive, then the threshold $\underline{q}$ is referred to as a kink point. We provide three examples of the payoff function in the prior literature.

Example 1: taxpayer bunching at kink points of tax schedule (Saez, 2010)

$$\max_q \underbrace{q - \frac{z}{1+\theta}\left(\frac{q}{z}\right)^{1+\theta} - \tau q}_{\text{policy-independent}} - \underbrace{\mathbf{1}_{q \geq \underline{q}} k [q - \underline{q}]^+}_{\text{policy-dependent}}, \tag{3.2}$$

where q is pre-tax income; z is the ability of the agent; θ is the inverse labor supply elasticity, $\frac{z}{1+\theta}\left(\frac{q}{z}\right)^{1+\theta}$ is the disutility of working, τ is the tax rate for income below $\underline{q}$, k is the change in the *marginal* tax rate at the threshold $\underline{q}$, which applies to the income exceeding the threshold, $[q - \underline{q}]^+$.

Example 2: taxpayer bunching at notch points of tax schedule (Kleven and Waseem, 2013)

$$\max_q \underbrace{q - \frac{z}{1+\theta}\left(\frac{q}{z}\right)^{1+\theta} - \tau q}_{\text{policy-independent}} - \underbrace{\mathbf{1}_{q \geq \underline{q}} k q}_{\text{policy-dependent}}, \tag{3.3}$$

where q is pre-tax income; z is the ability of the agent; θ is the inverse labor supply elasticity, $\frac{z}{1+\theta}\left(\frac{q}{z}\right)^{1+\theta}$ is the disutility of working, τ is the tax rate when income is below $\underline{q}$, and k is the change in the *average* tax rate at the threshold $\underline{q}$, which applies to all the income, q.

Example 3: banks bunching at notch points of tiered regulation (Alvero et al., 2022)

$$\max_q \underbrace{\theta(z - q) \exp q}_{\text{policy-independent}} - \underbrace{\mathbf{1}_{q \geq \underline{q}} k}_{\text{policy-dependent}}, \tag{3.4}$$

where q is the log bank assets; z is the level of demand; θ is the inverse semi-elasticity of demand, $\theta(z - q)$ is the profit margin, and k is the change in the regulatory costs when bank assets exceed the regulatory threshold.

3.2.2 Indifference condition of the marginal buncher

From the agents' payoff function, we can derive the optimal bunching choice for each agent. Define $q_0(z) = \arg\max \pi(q; z, \theta)$ as the undistorted choice of the agents. Agents with a large z will optimally choose a larger q_0 so that agents can be indexed by either the underlying heterogeneity z or the undistorted choice q_0.

When $\kappa(q)$ features a kink point, agents with $q_0 \in [\underline{q}, \overline{q}]$ bunch at $\underline{q}$, where $\overline{q}$ is given by the following indifference condition of the marginal agent:

$$\pi'(\overline{q}; z(\overline{q}), \theta) = \kappa'(\underline{q}). \tag{3.5}$$

When $\kappa(q)$ features a notch point, agents with $q_0 \in [\underline{q}, \overline{q}]$ bunch at $\underline{q}$, where $\overline{q}$ is given by the following indifference condition of the marginal agent:

$$\pi(\overline{q}; z(\overline{q}), \theta) - \kappa(\overline{q}) = \pi(\underline{q}; z(\overline{q}), \theta) - \kappa(\underline{q}). \tag{3.6}$$

We illustrate the intuition of optimal bunching choice using the simplest policy cost function, $\kappa = \mathbf{1}_{q \geq \underline{q}} k$, where k can be interpreted as a fixed regulatory cost when the bunching variable q crosses the regulatory threshold $\underline{q}$. As shown in the upper panel of Figure 3.5, the policy costs create a discrete drop in the payoff function at the regulatory threshold, $\underline{q}$. If an agent's optimal undistorted choice q_0 is just above the regulatory threshold, it is more profitable to bunch at $\underline{q}$ to avoid the discrete increase in policy costs. However, if an agent's optimal undistorted choice q_0 far exceeds the regulatory threshold, bunching is too costly. This agent will instead incur the regulatory costs k as shown in the middle panel of Figure 3.5. A marginal agent $\overline{q}$ is indifferent between bunching and not bunching, as shown in the bottom panel of Figure 3.5.

The indifference condition of the marginal agent establishes a one-to-one correspondence between the policy cost k and the payoff loss rate θ: $\pi(\overline{q}; z(\overline{q}), \theta) - k = \pi(\underline{q}; z(\overline{q}), \theta)$. Given the marginal agent $\overline{q}$, we can use the above indifference condition to calculate the underlying economic parameters. For example, in the public finance and labor literature, researchers know the policy cost k (e.g., tax schedule). The goal is to identify the rate of payoff loss θ (e.g., labor supply elasticity). Conversely, if the rate of payoff loss is known, we can also estimate the regulatory costs faced by the agents.

The goal of bunching estimation is to identify the marginal agent, $\overline{q}$, or equivalently, the bunching range, $\Delta q \equiv \overline{q} - \underline{q}$, from the bunching pattern of the distribution of the bunching variable.

3.2.3 Estimate the marginal agent: sharp bunching

Saez (2010) proposes the sharp bunching approach, which identifies the bunching range Δq from the density spike at the regulatory threshold as illustrated in the top panel of Figure 3.1. Formally, define $g(.)$ as the probability density functions of the bunching variable, q, and $g_0(.)$ as the undistributed counterpart, q_0. $g_0(.)$ is often referred to as the "counterfactual distribution". The total abnormal densities, B, in a narrow interval around the threshold, $[q^-, q^+]$, is given by:

$$B \equiv \int_{q^-}^{q^+} \left(g\left(q\right) - g_0\left(q\right)\right) dq, \tag{3.7}$$

where $[q^-, q^+]$ is refereed to as the spike interval. The bunching range, $\Delta q \equiv \overline{q} - \underline{q}$, can be calculated using the following formula:

$$\Delta q_{SB} \equiv \frac{B}{g_0(\underline{q})}, \tag{3.8}$$

where $g_0(\underline{q})$ is the counterfactual probability density at the regulatory threshold. The above formula can be derived from equation (3.7) by assuming that the counterfactual density $g_0(\underline{q})$ is locally constant in the bunching range $[\underline{q}, \overline{q}]$. If the counterfactual density changes significantly in the bunching range, then the average densities in the bunching range, $\overline{g}_0 \equiv \int_{\underline{q}}^{\overline{q}} g_0\left(q\right) dq/(\overline{q} - \underline{q})$, can be used to replace $g_0(\underline{q})$. Note that the average density depends on the upper bound $\overline{q}$, so Δq needs to solved implicitly from equation (3.8).

3.2.4 Implementation procedure of the sharp bunching approach

Supposing the researchers observe a sample of bunching variables, $\{q_i\}_{i=1,2,\ldots,N}$, then the sharp bunching estimators can be implemented using the following procedure:

1. Group observation into J bins, $\left\{[b_{j-1}, b_j)\right\}_{j=1,2,\ldots,J}$. Estimate the empirical probability density in each bin, $\hat{g}(j) = \frac{n_j}{\delta}$, where $n_j \equiv \frac{1}{N}\sum_{i=1,2,\ldots,N} \mathbb{1}_{b_{j-1}<x_i\leq b_j}$ is the fraction of observations in a bin and δ is the bin size.

2. Estimate the counterfactual probability densities $\hat{g_0}$ by fitting a smooth polynomial function to the remaining observations after dropping the observations in the excluded range $[q_l, q_u]$.[8] Note that the excluded range should contain all the observations potentially affected by bunching, including the sharp spike at the regulatory threshold (and the hole above the regulatory threshold for bunching at notches).

3. Calculate the bunching mass B by integrating the actual and counterfactual density functions over the spike interval, $[q^-, q^+]$.[9] The bunching range, Δq, is calculated by using equation (3.8).

3.2.5 Challenges of the data noise for sharp bunching approach

The sharp bunching approach assumes that the bunching variable, q, can be (1) precisely controlled by the agents and (2) perfectly observed by researchers. This assumption is reasonable in the administrative data on taxable income typically used in the prior studies. However, in certain finance settings, the observed bunching variable may contain a noise term u with a probability density function $h(u)$:

$$x = q + u. \tag{3.9}$$

[8]An iterative process could be used to determine $[q_l, q_u]$: (1) start with a relatively large excluded range; (2) estimate the marginal agent; (3) update the excluded range such that it just contains the marginal agent; (4) repeat the process until it converges.

[9]The spike interval can be determined visually or by using an automatic procedure that sets the boundary of the interval as the points where the actual density function first crosses the counterfactual density function on each side of the threshold.

This noise term could arise due to many reasons. First, models are only an approximation of reality. In a theoretical bunching model, agents can control the bunching variable perfectly. However, there is large randomness in the real world so that agents could not set the bunching variable precisely at the regulatory threshold. For instance, Alvero et al. (2022) study a setting in which banks' assets (the bunching variable) fluctuate with the deposit inflows and outflows. Ewens et al. (2020) study a setting in which firms' market capitalization (the bunching variable) fluctuates with share prices. Depending on the magnitude of the noise, the density spike may display certain degrees of diffusion. Second, even if a theoretical model truthfully reflects the reality, no dataset can perfectly measure the underlying theoretical concept, as exemplified by Erickson and Whited (2000). Measure errors are a fact of life in empirical research. In the bunching context, Garicano et al. (2016) study a setting in which regulators use slightly different notions of employment size (the bunching variable) from the data that researchers have access to, which leads to measurement errors.

Data noise poses a challenge for the sharp bunching approach. Researchers only observe the distribution of the bunching variable with the noise, defined as $f(x)$, rather than the distribution of the bunching variable alone $g(q)$. Using $f(.)$ in the sharp bunching approach in equation (3.7) may lead to biased estimates. Consider the following example. The bunching variable follows a uniform distribution $g_0(q) = U[0,1]$ without regulation. After the regulation is introduced, agents in $[\underline{q}, \overline{q}]$ choose to bunch at $\underline{q}$. The bunching mass $g_0 \Delta q$ is diffused by a noise term, which also follows a uniform distribution $u \sim U[-\sqrt{3}\sigma, \sqrt{3}\sigma]$, where σ is the standard deviation of the noise term. The actual probability density function of the observed bunching variable is given by

$$f(x) = \begin{cases} 1, & 0 \leq x \leq \underline{q} - \sqrt{3}\sigma, \\ 1 + \frac{\Delta q}{2\sqrt{3}\sigma}, & \underline{q} - \sqrt{3}\sigma < x \leq \underline{q}, \\ \frac{\Delta q}{2\sqrt{3}\sigma}, & \underline{q} < x \leq \underline{q} + \sqrt{3}\sigma, \\ 0, & \underline{q} + \sqrt{3}\sigma < x \leq \overline{q}, \\ 1, & \overline{q} \leq x \leq 1. \end{cases} \tag{3.10}$$

We apply the sharp bunching approach to this example. Given a spike interval $[q^-, q^+] = [\underline{q} - \epsilon, \underline{q} + \epsilon]$, the estimated bunching range using the sharp bunching approach is[10]

$$\Delta q_{SB} = \frac{B(\epsilon)}{g_0(\underline{q})} = \Delta q - \Delta q \max\left\{1 - \frac{\epsilon}{\sqrt{3}\sigma}, 0\right\} - \epsilon. \tag{3.11}$$

Equation (3.11) shows the sharp bunching estimate is a function of both the true bunching range Δq and the magnitude of the noise σ. The estimated bunching range converges to the true bunching range when the noise converges to zero and an infinitely tight spike interval is chosen: $\Delta q_{SB} \to 0\Delta q$ if $\sigma \to 0$ and $\epsilon \to 0$. However, if the noise is non-zero, the true bunching range is unidentified because there are two structural parameters, Δq and σ, but only one equation.[11] In other words, the true bunching range can only be identified from the bunching mass B if the magnitude of the noise σ is known (or is zero).

The bunching literature has recognized the issue of data noise since the seminal paper by Saez (2010). Table 2 Panel B of Saez (2010) shows that the sharp bunching estimate could be sensitive to the spike interval chosen by researchers. Chetty et al. (2011) propose a solution by using a wider spike interval to calculate the bunching mass. However, this approach only partially addresses the problem because there is a trade-off in the spike interval: a narrow spike interval may miss some

[10]See Appendix C.2 for derivation.

[11]The identification issue created by data noise is not the same as Blomquist et al. (2021), who argue that the bunching range Δq cannot be identified when the counterfactual density $g_0(\underline{q})$ is unrestricted.

abnormal densities, leading to a bias term, $-\Delta q \max\left\{1-\frac{\epsilon}{\sqrt{3}\sigma},0\right\}$; while a wide spike interval may contain the region where the actual probability density function falls below the counterfactual density function, leading to another bias term, $-\epsilon$. No spike interval can simultaneously make these two bias terms zero when the noise is present.

In many existing bunching studies, noise in the bunching variable is a secondary issue because the noise in the administrative data used by these studies is usually quite small. Chetty et al. (2011) find their estimates are indeed not sensitive to changes in spike interval in their setting. However, the noise can become more severe in many finance settings where administrative data are unavailable.

3.2.6 Fuzzy bunching

We now introduce the fuzzy bunching approach. As shown in the lower panel of Figure 3.1, in the presence of bunching, the cumulative distribution function displays a bulge around the regulatory threshold. The area of this bulge, A, is given by:

$$A \equiv \int \left(F\left(x\right) - F_0\left(x\right)\right) dx, \tag{3.12}$$

where F is the actual cumulative distribution function of the bunching variable; F_0 is the counterfactual cumulative distribution function of the bunching variable. Note that the integration can be calculated over the entire support of the data or a smaller interval containing the whole bulge. Theoretically, they should give rise to the same result because $F(.)$ overlaps with $F_0(.)$ outside of the bulge.

We can identify the bunching range using the bunching area, A:

$$\Delta q_{FB} \equiv \sqrt{\frac{2A}{f_0\left(\underline{q}\right)}}, \tag{3.13}$$

where A is the bunching area defined in equation (3.12); $f_0(\underline{q})$ is the counterfactual probability density function of the bunching variable at the regulatory threshold. If the counterfactual density changes significantly in the bunching range, then the average densities in the bunching range, $\overline{f_0} \equiv (F_0(\overline{q}) - F_0(\underline{q}))/(\overline{q} - \underline{q})$, can be used to replace $f_0(\underline{q})$. Note that the average density depends on the upper bound $\overline{q}$, so Δq needs to be solved implicitly from equation (3.13). Formally, we can establish the following results:

Proposition 1. *Let $G_0(q) \in C^2$ be the cumulative distribution function of the undistorted bunching variable, $h(u) \in C^2$ be the probability density function of the data noise. When Δq is sufficiently small and the noise converges to zero, we have*

$$\Delta q_{FB} = \Delta q. \tag{3.14}$$

Proof: See Appendix C.1.

The requirement that $G_0(q)$ is twice continuously differentiable corresponds to the "smoothness" identification assumption of the sharp bunching approach (Kleven, 2016), which requires that the counterfactual probability density of the bunching variable should be continuously differentiable, $g_0(q) \in C^1$. Intuitively, this assumption means that the probability density of the bunching variable should be smooth in the absence of the threshold-based policy. So non-smoothness implies that there is bunching. A potential violation of this assumption would be that other confounding policies change discretely at the same threshold, creating bunching in the absence of the specific policy being analyzed. In such cases, the estimate captures the combined effects of all the policies. Another potential violation is that the threshold happens to be located at a salient round number and, therefore, is subject to round-number bunching. The smoothness assumption can potentially be verified if the pre-regulation distribution is observable. Note that the condition that Δq is small can be relaxed if the average density in the bunching range, rather than the density at the regulatory threshold, is used as the denominator in equation (3.13).

To illustrate the robustness of the fuzzy bunching approach in small and noisy data, we apply the fuzzy bunching approach to the theoretical example in section 3.2.3. The fuzzy bunching approach can recover the true parameter even in the presence of the noise:[12]

$$\Delta q_{FB} = \sqrt{\frac{2A}{f_0\left(\underline{q}\right)}} = \Delta q \ \forall \sigma. \tag{3.15}$$

3.2.7 Implementation procedure of the fuzzy bunching approach

This section discusses how to implement the conceptual approach laid out in the previous section in the data. Supposing the researchers observe a sample of bunching variables, $\{x_i\}_{i=1,2,...,N}$. The fuzzy bunching approach can be implemented using the following procedure:[13]

1. Construct the actual cumulative distribution function as the following

$$\hat{F}(x) = \frac{1}{N} \sum_{i=1,2,...,N} \mathbb{1}_{x_i \leq x}. \tag{3.16}$$

2. Estimate the counterfactual cumulative distribution function $\hat{F}_0$ by fitting a smooth polynomial function to the remaining sample after dropping the observations in the excluded range, $[x_l, x_u]$:

$$\hat{F}(x_i) = \sum_{p=1}^{P} \beta_p (x_i)^p + v_i \ , \forall x_i \notin [x_l, x_u], \tag{3.17}$$

where P is the degree of the polynomial, v_i is the regression residual. The excluded range, $[x_l, x_u]$, should span the entire range affected by the bunching response, i.e., the area featuring the bunching bulge. Similar to the sharp bunching approach, the excluded range can be determined visually.

The counterfactual cumulative distribution function is given by

$$\hat{F}_0(x) = \sum_{p=1}^{P} \hat{\beta}_p x^p, \tag{3.18}$$

[12] See Appendix C.2 for derivation.

[13] The fuzzy bunching estimation code can be found in `https://sites.google.com/site/kairongxiao/data-and-code`.

where $\hat{\beta}_p$'s are estimated from regression (3.17).

3. Calculate the bunching area A by integrating the actual and counterfactual cumulative distribution functions using equation (3.12). Theoretically, the integration interval can be the whole support of data or any smaller interval that contains the bunching bulge. However, a smaller interval is usually preferable in practice because it reduces the fitting errors when estimating the counterfactual distribution.

4. Calculate the counterfactual density around the threshold as the following:

$$\hat{f}_0(\underline{q}) = \hat{F}_0'(x) = \sum_{p=1}^{P} \hat{\beta}_p p \underline{q}^{p-1}, \tag{3.19}$$

   where $\hat{\beta}_p$'s are estimated from regression (3.17). If the bunching range is sufficiently wide and the counterfactual density changes significantly in the bunching range, then the average densities in the bunching range, $(\hat{F}_0(\overline{q}) - \hat{F}_0(\underline{q}))/(\overline{q} - \underline{q})$, can be used to replace $\hat{f}_0(\underline{q})$.

5. Calculate the bunching range, Δq, using equation (3.13). The standard error of the bunching estimate can be constructed using a bootstrapping procedure in which bootstrap samples are created by drawing with replacement from the observed sample of bunching variable, $\{x_i\}_{i=1,2,...,N}$. The bootstrap standard error is then the standard deviation of the bunching estimates obtained from the bootstrap samples.

In applications in which the pre-regulation sample is available (e.g. DeFusco et al. (2020) and Ewens et al. (2020)), researchers can construct the counterfactual distribution from the pre-regulation sample, $\{x_i(0)\}_{i=1,2,...,N}$:[14]

$$\hat{F}_0(x) = \frac{1}{N} \sum_{i=1,2,...,N} \mathbb{1}_{x_i(0) \leq x}. \tag{3.20}$$

[14]Using the pre-regulation sample to form counterfactual distribution addresses the issue raised by Blomquist et al. (2021) that restricting the counterfactual distribution to a known functional form may appear unusual in some cases.

3.2.8 Implementation of fuzzy and sharp bunching: a comparison

In this section, we compare the implementation procedure of the fuzzy bunching approach with that of the sharp bunching approach. This comparison illustrates the practical advantages of the fuzzy bunching approach in small and noisy data.

The first advantage of the fuzzy bunching approach is that it avoids specification of the spike interval, $[q^-, q^+]$, which is required to calculate the bunching mass in equation (3.7) for the sharp bunching approach. The spike interval is typically determined visually in the sharp bunching approach. When the bunching variable contains substantial noise or the sample size is not big enough to precisely estimate density, determining the spike interval could be quite challenging. Furthermore, as shown in the theoretical sample in section 3.2.5, when the data noise diffuses the bunching mass around the threshold, there may exist no spike interval that can identify the true parameter. In comparison, the fuzzy bunching approach does not require researchers to choose the spike interval, which removes the ambiguity associated with this choice.

The second advantage of the fuzzy bunching approach is that it avoids density estimation. When using the sharp bunching approach, researchers need to estimate the empirical probability density function typically with the histogram approach.[15] When the data are sparse, estimating the probability density inside each bin is challenging because there could be no observations falling into a narrow bin. As a result, the estimated density function may contain spurious holes and spikes, as shown in Figure 3.2, making it difficult to determine the spike interval. Furthermore, the estimated probability density function is potentially sensitive to bin size, δ. As a result, many robustness checks regarding this choice are often performed.[16] In comparison, the fuzzy bunching approach uses the cumulative distribution function, which can be constructed without specifying a bin size, as shown in equation (3.16). This approach removes the ambiguity associated with the bin size choice in density estimation.

[15] There is a large literature in statistics and econometrics on alternative approaches of density estimation. See McCrary (2008) and Cattaneo, Jansson, and Ma (2020) for references.

[16] For example, see Bachas, Kim, and Yannelis (2020).

3.3 Simulation evidence

Before applying our fuzzy bunching approach to real data, we validate the fuzzy bunching approach with simulated data. We first describe the simulated data. Then we compare the fuzzy bunching approach with the sharp bunching approach.

3.3.1 Simulated data

We consider a setting in which banks face a size-based regulation: banks incur a regulatory cost k if their assets exceed $\exp \underline{q}$ = \$10 billion. Banks choose log asset q to maximize their profits net of regulatory costs: $\pi(q; z, \theta) - \kappa(q) = \theta(z - q) \exp q - k\mathbf{1}_{q \geq \underline{q}}$. z is the level of demand, which determines the undistorted size of a bank. θ is the inverse semi-elasticity of demand, which takes a value of 1%. We set the true bunching range as $\Delta q = 0.5$, corresponding to roughly \$15 million regulatory costs.[17] To check the bunching approach's robustness to alternative data generating processes, we simulate the undistorted size using (1) a normal distributed with $q_0 \sim N(2, 0.7)$ and (2) an exponential distribution with the rate parameter $\lambda = 0.1$. Bank size contains a noise, which follows a normal distribution: $u \sim N(0, \sigma)$.

We report the results in Table 3.1. Each row reports a different scenario. We run 500 simulations for each scenario. We report the estimated bunching range normalized by the true value, $\widehat{\Delta q}/\Delta q$.[18] We report the bootstrapped standard errors in the brackets below. Column 1 shows the case in which the undistorted log assets follow a normal distribution. Column 2 shows the case in which the undistorted log assets follow an exponential distribution.

3.3.2 Results with simulated data

The first panel of Table 3.1 shows the results for the fuzzy bunching approach (FB). The first row reports the scenario with no noise ($\sigma = 0\%$). In this case, the fuzzy bunching approach precisely uncovers the true parameter value. We then increase the standard deviation of the noise

[17]The regulatory cost implied by a bunching range can be calculated using equation (3.6).
[18]Equivalently, we can report the estimated regulatory cost.

to 2% and 4% in the second and third rows, respectively. Consistent with Proposition 1, the fuzzy bunching estimates are invariant to the magnitude of the noise. We obtain similar results when using an alternative distribution for the bunching variable, as shown in column 2.

We compare the fuzzy bunching approach with the sharp bunching approach, which is reported in panel 2 of Table 3.1. Note that the sharp bunching approach requires researchers to specify the spike interval, $[q^-, q^+]$. Following the prior sharp bunching literature, we use an automatic procedure that sets the boundary of the spike interval as the points where the actual density function first crosses the counterfactual density function on each side of the threshold. Next, we set bin size 0.005 to make sure that there are enough observations to estimate density. We again start with the ideal scenario in which there is no noise ($\sigma = 0\%$). The sharp bunching approach precisely uncovers the true parameter value. We then examine the robustness of the sharp bunching approach to the noise by comparing the estimates across the rows. We find that the estimates become downward biased as large noise diffuses the bunching mass around the threshold. The results are similar if we use the exponential distribution instead of the normal distribution.

The sharp bunching approach requires researchers to choose the bin size to estimate the probability density function. In panel 3 of Table 3.1, we reestimate the bunching range under an alternative specification of the bin size. We find the estimates are sensitive to bin size choice, especially when the noise is present. In comparison, the fuzzy bunching approach does not need this tuning parameter because it uses the cumulative distribution function rather than the probability density function.

One may worry that our results could be driven by the automatic procedure used to determine the spike interval. To address this concern, we manually choose a spike interval around the threshold $[q^-, q^+] = [\underline{q} - \epsilon, \overline{q} + \epsilon]$, and set ϵ to 0.01 in panel 4 of Table 3.1. We find that the sharp bunching approach again underestimates the bunching range in the presence of the noise. Furthermore, the sharp bunching estimates are sensitive to the spike interval choice, as shown in panel 5.

The intuition behind the above results can be illustrated in Figure 3.6. Row 1 plots the theoretical cumulative distribution function and the probability density function in the no-noise scenario. The bunching incentive creates a triangle in the cumulative distribution function (column 1) a sharp spike in the probability density function (column 2). In this case, both sharp and fuzzy bunching approaches can identify the true bunching range. Row 2 plots the scenario with noise. Introducing the noise (the middle left panel) does not have a first-order effect on the size of the bunching area: the height of the "triangle" becomes smaller, but the base becomes larger, and the area stays approximately constant. Therefore, the fuzzy bunching approach can identify the bunching range regardless of the magnitude of the noise. In comparison, the noise diffuses the density spike used by the sharp bunching approach. A narrow spike interval misses some of the mass, while a wide spike interval includes regions where the actual density function falls below the counterfactual one.

Finally, the fuzzy bunching approach also has a practical advantage in the small finite sample, as shown in row 3, which plots the empirical distribution functions with 500 observations. The bunching bulge in the cumulative distribution function can still be seen clearly, but the bunching spike in the empirical probability density function becomes quite blurry.

3.4 Application to real data

So far, we have shown the performance of the fuzzy bunching approach in simulated data. Now we proceed to apply this approach to real data. We consider three applications. First, we consider U.S. income tax return data studied by Saez (2010), which feature bunching created by changes in marginal tax rates. Second, we consider the Pakistan income tax return data studied by Kleven and Waseem (2013), which feature bunching created by changes in average tax rates. Finally, we consider the bank asset data studied by Alvero et al. (2022), which feature bunching created by changes in regulatory costs. Although these three applications feature different economic contexts, they all entail the estimation of the bunching range, Δq.[19] The U.S. and Pakistan income tax return data have many observations and little noise, so both the sharp and fuzzy bunching approaches should provide similar estimates. However, the U.S. bank data have a small sample size and

[19]Translating the bunching range to structural parameters, however, requires context-specific indifference conditions.

substantial noise in the bunching variable, which is more suitable for the fuzzy bunching approach.

3.4.1 Application 1: Taxpayers bunching at tax kinks

Saez (2010) documents significant bunching at the U.S. federal income tax schedule's kink points using the Individual Public Use Tax Files constructed by the IRS. The data are available from 1960 to 2004. The number of observations per year is between 80,000 and 200,000.

We apply our fuzzy bunching approach to the data in Saez (2010). In particular, we consider the sample of tax filers with one dependent child around the \$8,580 threshold, where the marginal tax rate increases from -34% to 0%. This discrete change creates a significant bunching mass at this threshold as shown by the solid red line in the lower panel of Figure 3.7 (Panel 4 of Figure 4 in Saez, 2010).[20] To make the estimates more comparable, we follow the same assumptions as in Saez (2010) to form the counterfactual distribution, which is the dashed blue line in the lower panel of Figure 3.7. Note that the counterfactual distribution of the taxable income has higher densities on the right-hand side of the threshold. This is because the earning potential (or ability) is smooth, so a higher tax rate makes the taxable income more concentrated on the right-hand side of the threshold.

Applying the fuzzy bunching approach to the cumulative distribution function, we find the marginal taxpayer has an income of $\overline{q} = \$11,921$. Using equation (3) in Saez (2010), our estimate of the marginal taxpayer implies an income elasticity of 1.124. This value is close to the estimate of 1.101 using the sharp bunching approach as shown in row 1 and column 3 of Table 2 in Saez (2010).[21]

3.4.2 Application 2: Taxpayers bunching at tax notches

Our second application uses the administrative data of taxable income in Pakistan studied in Kleven and Waseem (2013). The data include the universe of personal income tax returns filed for

[20] Note that there is no hole in the probability densities above the threshold because the bunching in the context of Saez (2010) is created by a discrete change in the slope of incentives rather than the level.

[21] Because the individual-level data are confidential, we use the kernel density function of the earnings estimated by Saez (2010) directly. As a result, we can only compare the point estimate but not the bootstrap standard errors.

the tax years 2006–2009, which have about 4 million observations. The large sample size allows an accurate estimate of the probability density function around the threshold of tax rate changes.

We apply our fuzzy bunching approach to the data of Kleven and Waseem (2013). We consider the PKR 500K threshold, where the average tax rate changes from 10% to 12.5%. The change in tax rates creates a sharp spike in the probability density function of the taxable income as shown in Figure 3.8 (Figure VI, Panel E in Kleven and Waseem, 2013). Note that bunching in the Pakistan data is created by changes in the average tax rate (notches) rather than the marginal tax rate (kinks) as in the U.S. data studied by Saez (2010). Specifically, there is a dominated region just above the threshold because the level of after-tax income will experience a discrete drop if the taxable income exceeds the notch point.

We follow Kleven and Waseem (2013) to fit a five-degree polynomial to the density function after excluding the potential region that is affected by the discontinuity, $[q_l, q_u] = [480, 550]$, to estimate the counterfactual distribution in the absence of regulation.[22] The red and blue lines in the lower panel of Figure 3.8 show the empirical and counterfactual cumulative distribution functions, respectively. The fuzzy bunching approach generates an estimated bunching range of 43. This estimate is close to the result from the sharp bunching approach in Kleven and Waseem (2013), which is 40 (Table II column 9 in Kleven and Waseem, 2013).[23] The successful replication of existing sharp bunching results reassures the fuzzy bunching approach's validity in large samples.

3.4.3 Application 3: Banks bunching at regulatory size thresholds

So far, we have shown that both the sharp and fuzzy bunching approaches work well in settings with large sample sizes and little noise. We now consider the U.S. bank data studied by Alvero et al. (2022), a setting with small sample size and substantial noise. Unlike the individual-level tax data, which have millions of observations, there are only around 400 observations between \$9 billion and \$11 billion in the post-Dodd–Frank period. As a result, the empirical probability density function is sensitive to the choice of bin size, as shown in Figure 3.2, which can potentially

[22] We again use the density function of the taxable income estimated by Kleven and Waseem (2013) directly.

[23] Note that we follow Kleven and Waseem (2013) to estimate the fraction of non-optimizing agents.

affect the results of the sharp bunching approach. Furthermore, it is difficult to determine a narrow interval containing the sharp spike because the data noise diffuses it. Alvero et al. (2022) estimate a full structural model using maximum likelihood. We apply our fuzzy bunching approach to this data and compare our estimates with those from maximum likelihood estimation and those from sharp bunching estimation.

Column 1 of Table 3.2 presents the estimate of regulatory costs using the fuzzy bunching approach. We fit a smooth polynomial with degree six to the cumulative distribution after excluding the region that is potentially affected by the discontinuity, $[q_l, q_u] = [\ln 9.0, \ln 11.5]$, to estimate the counterfactual distribution. Alternatively, we can use the pre-regulation data to form the counterfactual distribution, and the results are quite similar. The estimated bunching range Δq is 0.10, which means that the marginal bank is willing to give up 10% of the assets to avoid the Dodd–Frank regulations. The fuzzy bunching estimate is consistent with the results from the maximum likelihood estimation in Alvero et al. (2022), which is also 0.10.

We now compare the estimates from the fuzzy bunching approach with the sharp bunching approach, which are presented in columns 2-4 in Table 3.2. Similar to the fuzzy bunching approach, we need to estimate the counterfactual distribution function. To this end, we fit a smooth polynomial with degree five to the probability density function after excluding the region potentially affected by the discontinuity, $[q_l, q_u] = [\ln 9.0, \ln 11.5]$. The counterfactual distribution can also be formed using the pre-regulation data. Overall, we find the sharp bunching estimates range from 0.01 to 0.03, which are smaller than the estimate from the maximum likelihood estimation in Alvero et al. (2022). This result is consistent with our simulation evidence in Section 3.3 that the sharp bunching approach turns to underestimate the extent of bunching when the bunching mass is diffused.

As we discussed in Section 3.3, the sharp bunching approach may be sensitive to the bin size and the spike interval in noisy data. Therefore, we compare the estimates under different combinations of bin size and spike intervals. We first fix the bin size to 0.01 but vary the spike interval in columns 2 and 3. We find that the estimated bunching range is indeed quite sensitive to the spike interval. Next, we fix the spike interval and vary the bin size in column 4. Again, the estimated bunching range is sensitive to this tuning parameter. Finally, we use the automatic approach described in Section 3.2.4 to determine the spike interval in column 5. The estimate is still much smaller than that from the maximum likelihood estimation in Alvero et al. (2022). Overall, this exercise highlights the difficulty of applying the existing sharp bunching approach to the small and noisy data and the comparative advantage of the fuzzy bunching approach in such situations.

3.5 Extensions

This section discusses two extensions of the baseline framework presented above: non-optimizing agents and round-number bunching.

3.5.1 Non-optimizing agents

So far, we assume all agents can optimize in response to regulation. However, some agents may not respond to the bunching incentive in certain situations because of inattention or some other optimization frictions. Kleven and Waseem (2013) shows how to adjust the baseline sharp bunching approach with non-optimizing agents. In this section, we discuss how to adjust the baseline fuzzy bunching approach with non-optimizing agents.

Suppose a fraction α of agents is non-optimizing. The bunching range based on the fuzzy bunching approach is given by the following formula:

$$\Delta q_{FB\ friction} \equiv \sqrt{\frac{2A}{(1-\alpha) f_0\left(\underline{q}\right)}}, \tag{3.21}$$

where A is the bunching area defined in equation (3.12), f_0 is the counterfactual probability density function of the bunching variable, $\underline{q}$ is the regulatory threshold, and α is the fraction of non-optimizing agents.[24]

The above formula requires an estimate of the fraction of non-optimizing agents, α. Kleven

[24] See Appendix C.3 for derivation.

and Waseem (2013) show that this fraction can be estimated from the probability mass in the "dominated region," defined as the region just above the regulatory threshold. The intuition is the following. As shown by the top right panel of Figure 3.6, if all agents can optimize, then the dominated region should be empty because all of the agents in this region should strictly prefer to bunch. However, if we observe some agents inside this "dominated region" as shown in the top right panel of Figure C.1, that implies that these agents are likely facing optimization frictions.

We extend the fuzzy bunching approach to estimate the fraction of non-optimizing agents in the presence of noise. The idea is that the noise should move the same amount of probability mass to each side of the threshold. Thus, if we subtract half of the bunching mass from the dominated region, we can recover the true mass in the dominated region. Formally, we can derive the following formula of the fraction of non-optimizing agents:

$$\alpha_{FB} = \frac{2\left(F\left(\overline{q}\right) - F\left(\underline{q}\right)\right)}{f_0\left(\underline{q}\right)\Delta q} - 1, \tag{3.22}$$

where F is the actual cumulative distribution function, and $f_0\left(\underline{q}\right)$ is the counterfactual probability density function evaluated at $\underline{q}$.[25] Note that the fraction of non-optimizing agents, α, and the bunching range, Δq, are mutually dependent. We need to jointly solve equation (3.21) and equation (3.22) to get the fraction of non-optimizing agents, α, and bunching range, Δq.[26]

Note that both optimization frictions and noise lead to non-zero probability density above the regulatory threshold. The difference is that optimization frictions reduce the bunching area, but the noise does not. This asymmetry allows us to identify the fraction of non-operating agents when data noise is present.

[25] See Appendix C.4 for derivation.

[26] In Online Appendix Table C.1, we show that the fuzzy bunching approach is robust in the presence of optimization friction.

3.5.2 Round-number bunching

In some situations (e.g. Kleven and Waseem (2013), Bachas et al. (2020), Antill (2020)), agents may bunch at round numbers because they constitute a natural reference points. When the policy threshold is located at a round number, researchers need to net out the confounding effect by using similar round numbers that are not policy thresholds (Kleven, 2016). Following a similar logic, the bunching range based on the fuzzy bunching approach adjusted by round number bunching is given by:

$$\Delta q_{FB\, round-number} \equiv \sqrt{\frac{2A}{f_0\left(\underline{q}\right)}} - \frac{1}{R}\sum_{r=1}^{R}\sqrt{\frac{2A_r}{f_0\left(\underline{q_r}\right)}}, \tag{3.23}$$

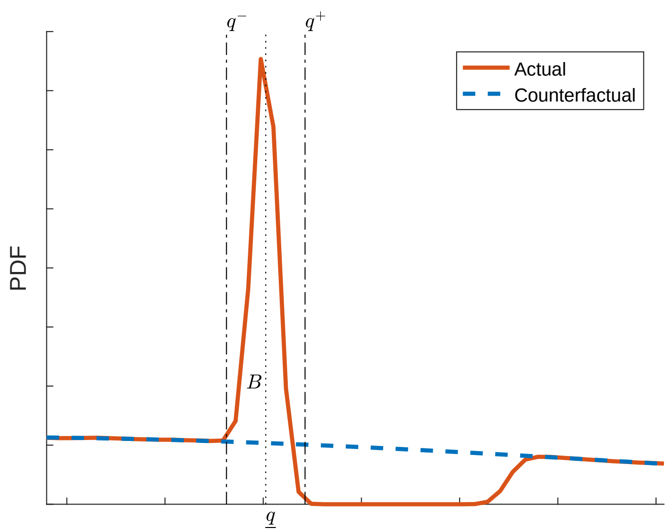
The second term of the above formula is the average round-number bunching for round numbers similar to the policy threshold. A_r is the bunching area generated by round number r, $f_0\left(\underline{q_r}\right)$ is the counterfactual density around round number r, and R is the total number of round numbers that are similar to the policy threshold.

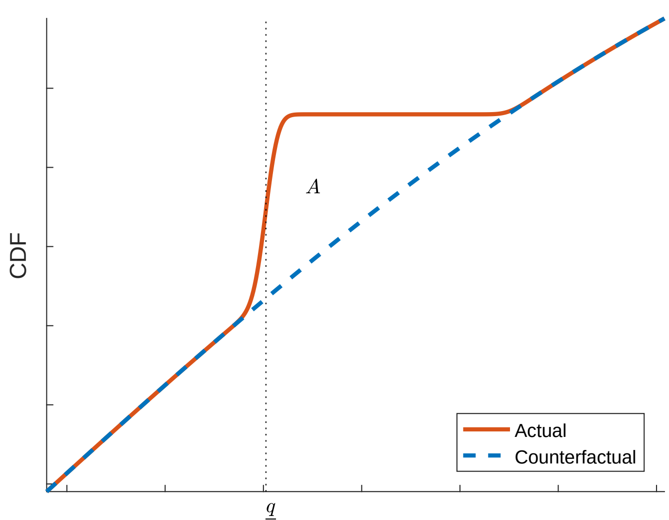
3.6 Conclusion

This paper introduces the fuzzy bunching approach that identifies bunching from the bunching area in the cumulative distribution function rather than the bunching mass in the probability density function. We show that the fuzzy bunching approach can identify bunching more robustly when the bunching variable is noisy. The fuzzy bunching approach can also provide a better way to visualize and measure bunching when the sample size is small. We hope the fuzzy bunching approach can broaden the scope of the bunching methodology and allow researchers to explore empirical settings in which traditional identification strategies such as difference-in-differences and regression discontinuity do not apply.

3.7 Figures and Tables

Figure 3.1: Bunching in Probability Density Function v.s. Cumulative Distribution Function

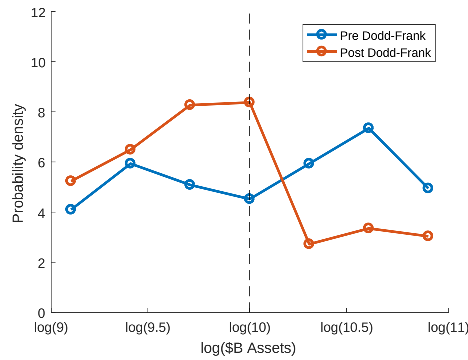




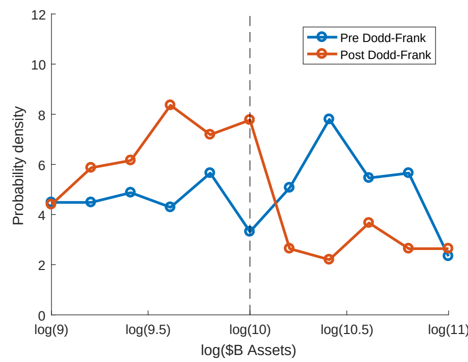
Note: The figure displays the bunching patterns in the probability density function (upper panel) and the cumulative distribution function (lower panel). $\underline{q}$ is the threshold that creates bunching. B is the excess bunching mass in the actual probability density function used for the sharp bunching approach. $[q^-, q^+]$ is the spike interval used for the sharp bunching approach. A is the bunching area between the actual and counterfactual cumulative distribution functions used for the fuzzy bunching approach.

Figure 3.2: Probability Density Function of Bank Assets Before and After the Dodd–Frank Act

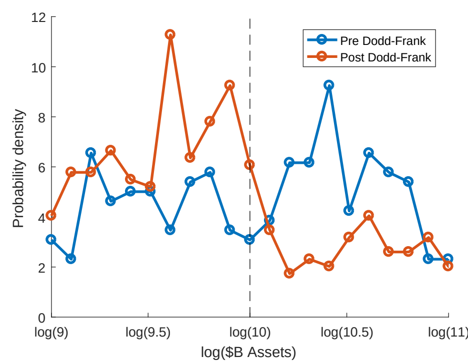
Bin size = 0.03



Bin size = 0.02

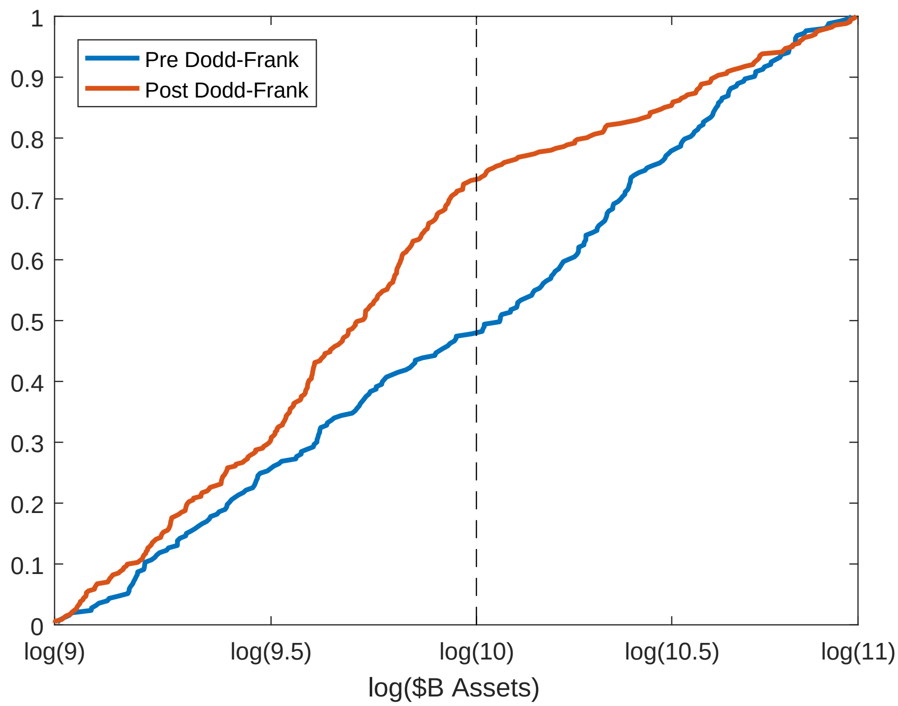


Bin size = 0.01



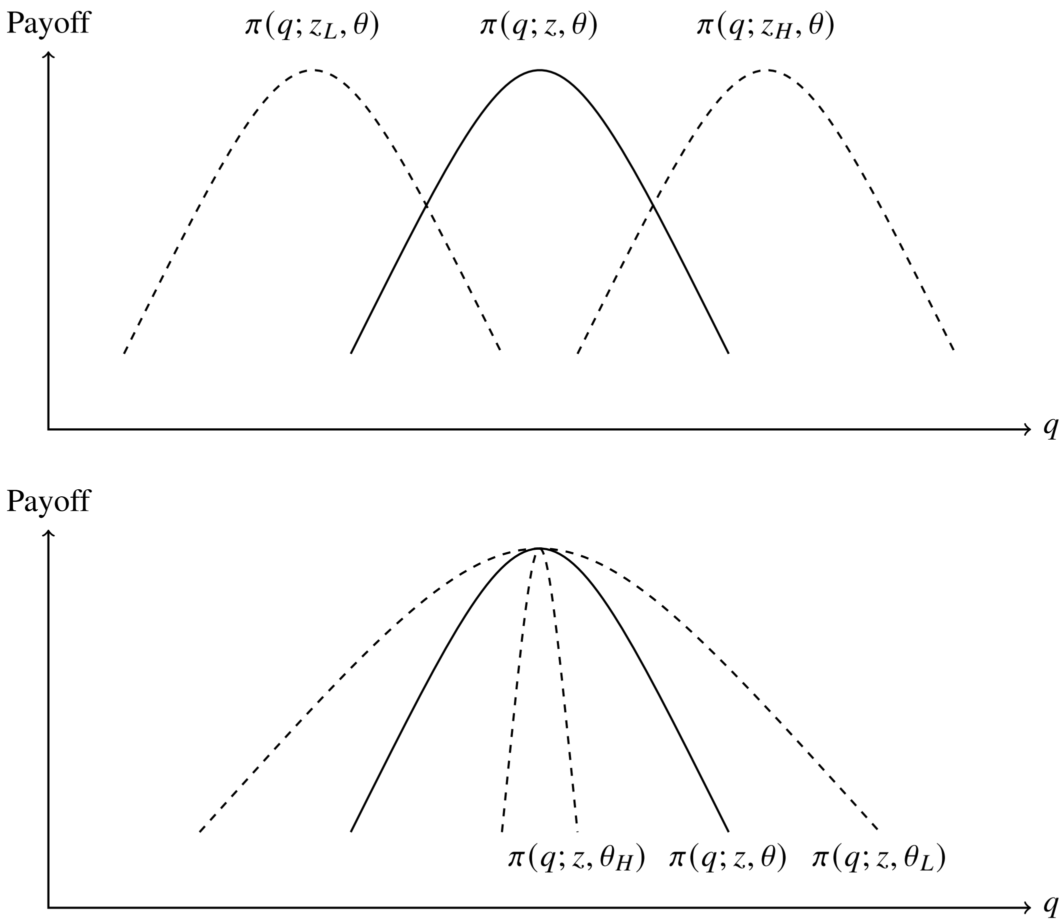
Note: This figure shows the probability density function of bank assets between $9 billion and $11 billion before and after the Dodd–Frank Act. The bin size of the probability density function is set to 0.03, 0.02, and 0.01 respectively. The pre- and post-Dodd–Frank periods are defined as 2000Q3–2007Q3 and 2010Q3–2018Q3, respectively.

Figure 3.3: Cumulative Distribution of Bank Assets Before and After the Dodd–Frank Act



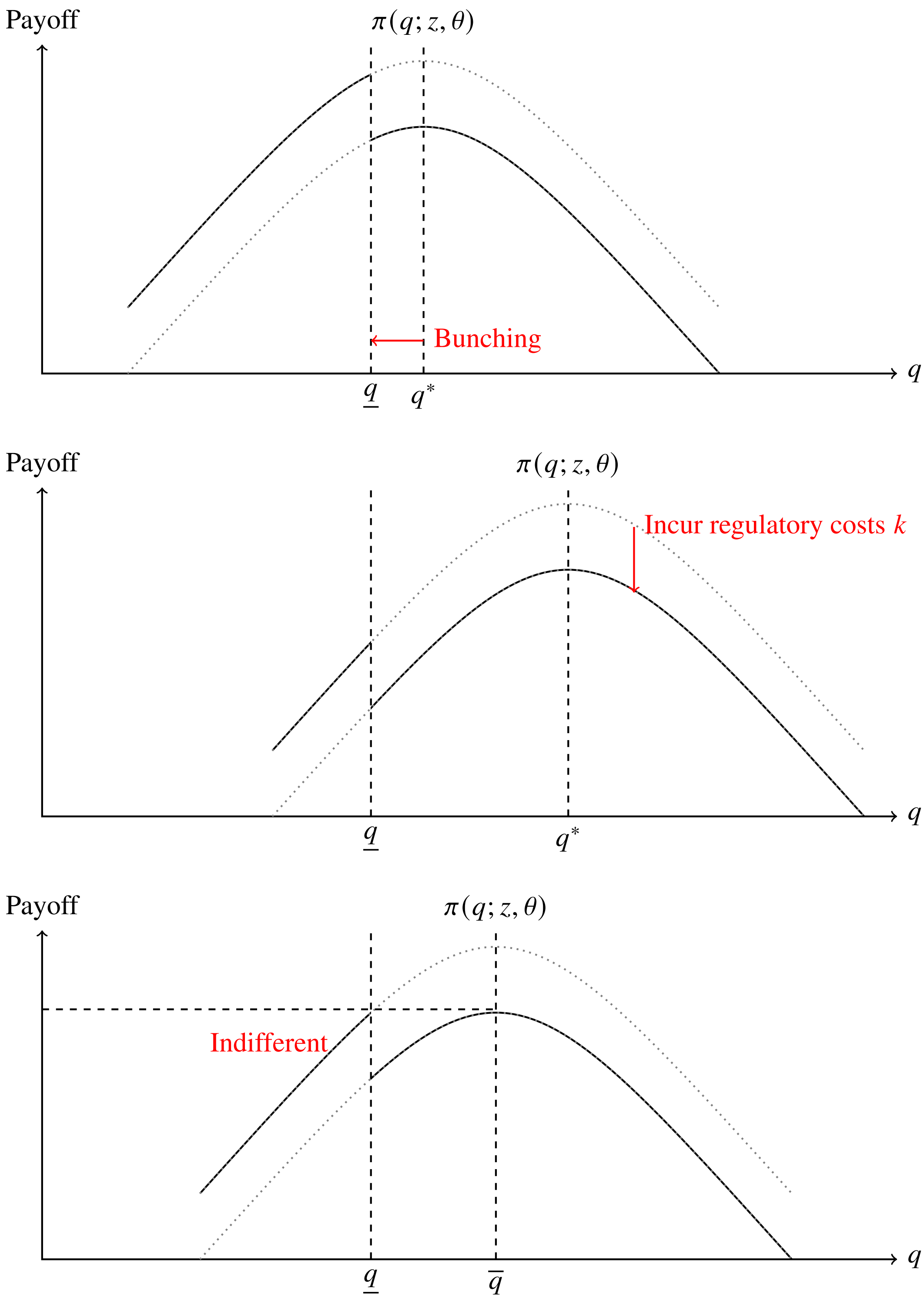
Note: This figure shows the cumulative distribution function of bank assets between $9 billion and $11 billion before and after the Dodd–Frank Act. The pre- and post-Dodd–Frank periods are defined as 2000Q3–2007Q3 and 2010Q3–2017Q3, respectively.

Figure 3.4: Bunching Model: The Policy-independent Component of Payoff Function



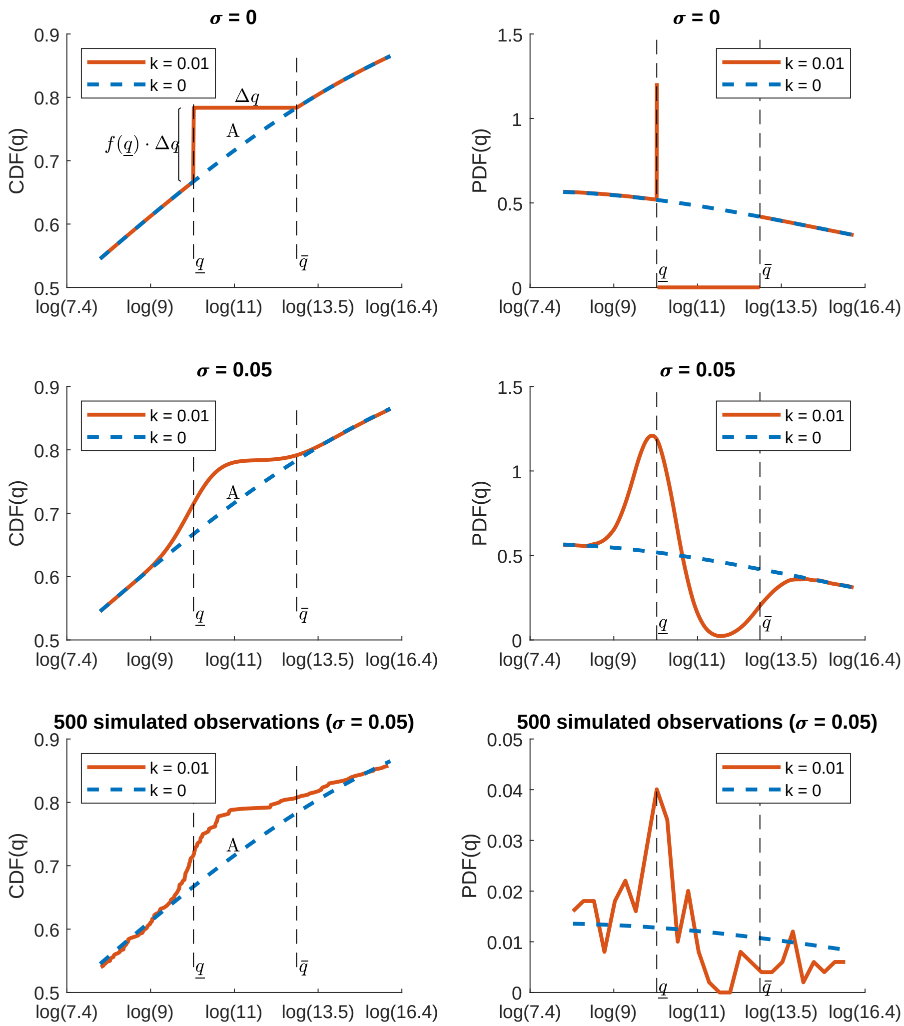
Note: This figure shows the policy-independent component of the payoff function. The upper panel shows the payoff functions of different agents, indexed by z. The lower panel shows the payoff functions with different payoff loss rates, θ.

Figure 3.5: Bunching Model: The Payoff Function with Policy Costs



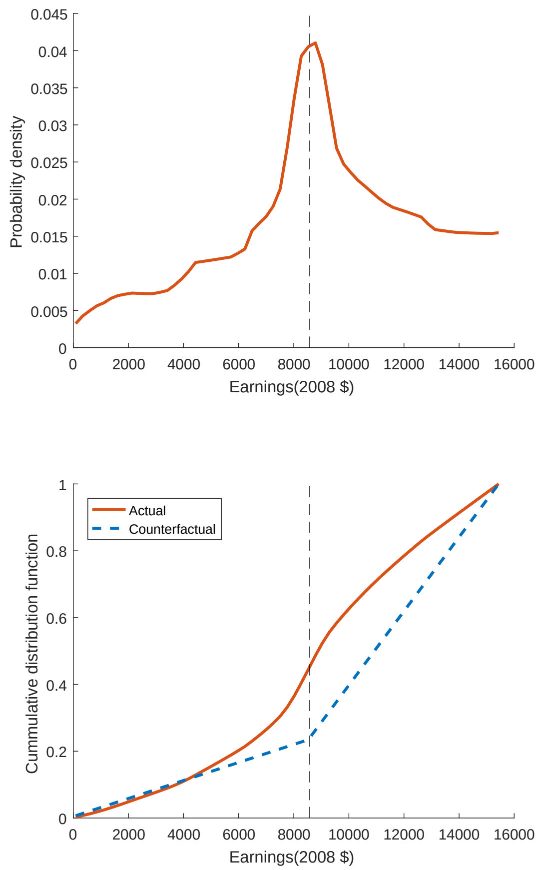
Note: This figure shows the policy-induced discontinuity in the payoff functions for different agents. The upper panel represents an agent who is just above the regulatory threshold and chooses to bunch. The middle panel represents an agent that chooses not to bunch because it is far above the regulatory threshold. The bottom panel represents an agent that is indifferent between bunching and not bunching.

Figure 3.6: Hypothetical Example: Banks Bunch at Regulatory Threshold



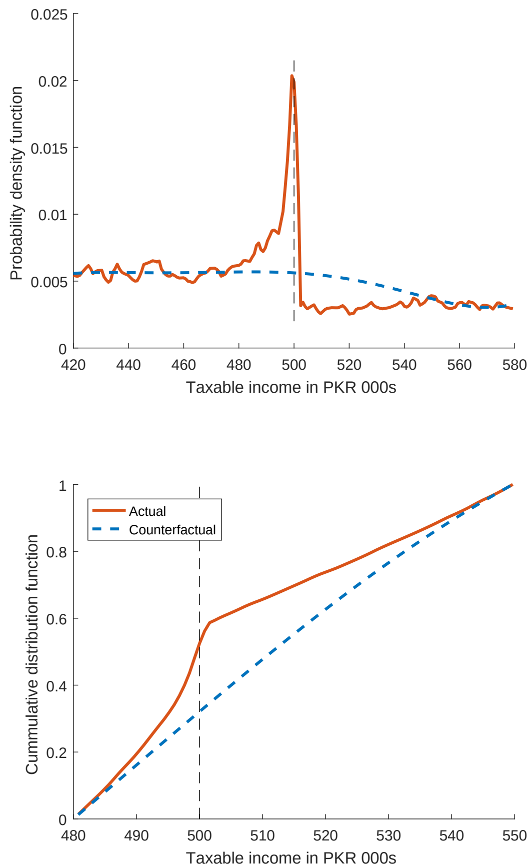
Note: This figure shows a simulated example of asset size distribution when banks bunch at the $10 billion regulatory threshold. The data-generating process is described in Section 3.3. The left and right columns show the cumulative distribution function and the probability density function, respectively. The horizontal axis is the log bank assets. The vertical axis shows the probability densities and the cumulative probability densities on the left and right panels, respectively. The top panel shows the theoretical distributions with no noise. The middle panel shows the theoretical distributions with noise. The bottom panel shows the empirical distributions with 500 observations and noise. The probability density function on the bottom right panel is based on bin sizes of 0.025.

Figure 3.7: Distributions of Taxable Income around the $8580 Kink



Note: The figure displays the probability density for tax filers with one dependent child. The charts include all years 1995–2004. The bandwidth is $400. The vertical dashed line marks the $8580 kink point of the tax rate schedule at which the marginal tax rate increases from −34% to 0%. The data source is Figure 4, Panel A of Saez (2010). The solid line is the actual distribution and the dash line is the counterfactual distribution. We follow Saez (2010) to allow the counterfactual densities to be different across the two sides of the threshold.

Figure 3.8: Distributions of Taxable Income around the 500,000 Rupee Notch



Note: The figure shows the empirical distribution of taxable income in Pakistan from 2006 to 2009. The upper panel is the empirical histogram of the number of taxpayers with a bin size of 500 rupees. The lower panel is the empirical cumulative distribution function. The vertical dashed line marks the 500,000 rupee notch point of the tax rate schedule at which the average tax rate increases from 10% to 12.5%. The blue dashed line is the counterfactual cumulative distribution function. The data source is Figure IV, Panel E of Kleven and Waseem (2013). The solid line is the actual distribution and the dash line is the counterfactual distribution. We follow Kleven and Waseem (2013) to fit a polynomial of five degrees to estimate the counterfactual distribution.

Table 3.1: Fuzzy Bunching versus Sharp Bunching in Simulated Data

	σ (%)	**q ~ N(2,0.7)**	**q ~ exp(λ=0.1)**
FB	0	1.035	1.041
		[0.049]	[0.052]
	2	1.009	1.034
		[0.056]	[0.056]
	4	0.984	1.026
		[0.062]	[0.061]
SB bin size: 0.005 ε: Auto	0	1.000	1.013
		[0.074]	[0.069]
	2	0.861	0.906
		[0.075]	[0.069]
	4	0.625	0.759
		[0.149]	[0.117]
SB bin size: 0.010 ε: Auto	0	0.989	1.000
		[0.075]	[0.069]
	2	0.889	0.907
		[0.064]	[0.062]
	4	0.774	0.810
		[0.083]	[0.072]
SB bin size: 0.005 ε: 0.010	0	0.974	0.989
		[0.064]	[0.067]
	2	0.318	0.358
		[0.049]	[0.057]
	4	0.156	0.176
		[0.036]	[0.042]
SB bin size: 0.005 ε: 0.020	0	0.954	0.967
		[0.071]	[0.065]
	2	0.593	0.631
		[0.063]	[0.066]
	4	0.304	0.340
		[0.050]	[0.059]

Note: This table presents the estimation results using simulated data. We run 500 simulations for each set of parameters. We set the bunching range of log-assets to 0.5. We report the average estimates over true value ($\widehat{\Delta q}/\Delta q$) and their standard deviations in parenthesis. We use a sample of 500 observations of bank assets. The standard deviation of noise σ is 0%, 2%, and 4%, respectively. The left and right columns draw the undistorted assets q from the normal distribution and the exponential distribution, respectively. For the sharp bunching approach, we use different combinations of spike intervals and bin sizes. In panels 2 and 3, the spike interval is determined based on the automatic procedure that sets the boundary of the spike interval as the points where the actual density function first crosses the counterfactual density function on each side of the threshold. In panels 4 and 5, the spike interval is fixed ex ante as $[\underline{q} - \epsilon, \underline{q} + \epsilon]$.

Table 3.2: Regulatory Costs of the Dodd–Frank Act at $10 Billion Threshold

Approach	(1) FB	(2) SB	(3) SB	(4) SB	(5) SB
Bunching range (Δq)	0.10 [0.01]	0.01 [0.00]	0.03 [0.01]	0.02 [0.00]	0.03 [0.02]
Tuning parameters					
Spike interval $[q^-, q^+]$	-	ln(9.7)-ln(10.0)	ln(9.5)-ln(10.0)	ln(9.5)-ln(10.0)	Auto
Bin size (δ)	-	0.010	0.010	0.020	0.010

Note: This table presents the estimated regulatory costs k using the fuzzy bunching approach (column 1) and sharp bunching approach (columns 2-4) in the sample of banks around the $10 billion threshold. Bootstrapped standard errors are reported in brackets. Columns 2-3 represent a different combination of spike interval and bin size for the sharp bunching approach. In column 4, the spike interval is determined based on the automatic procedure that sets the boundary of the spike interval as the points where the actual density function first crosses the counterfactual density function on each side of the threshold.

References

Acharya, V. V., A. N. Berger, and R. A. Roman (2018). Lending implications of us bank stress tests: Costs or benefits? *Journal of Financial Intermediation 34*, 58–90.

Al-Ubaydli, O. and P. A. McLaughlin (2017). Regdata: A numerical database on industry-specific regulations for all united states industries and federal regulations, 1997–2012. *Regulation & Governance 11*(1), 109–123.

Alvero, A., S. Ando, and K. Xiao (2022). Watch what they do, not what they say: Estimating regulatory costs from revealed preferences.

Alvero, A. and W. Renxuan (2022). Esg investing in emerging markets: Betting on firm fundamentals or riding investor preferences.

Alvero, A. and K. Xiao (2020). Fuzzy bunching. *Available at SSRN 3611447*.

AMAC (2017). ESG responsible investment research report.

AMAC (2021). Fund manager's green investment self-assessment report (2021).

Amiraslani, H., K. Lins, H. Servaes, and A. Tamayo (2021). Trust, social capital, and the bond market benefits of ESG performance. European Corporate Governance Institute (ECGI) - Finance Working Paper No. 535/2017.

Amstad, M. and Z. He (2019, February). Chinese bond market and interbank market. Working Paper 25549, National Bureau of Economic Research.

Ando, S. (2021). Size-dependent policies and risky firm creation. *Journal of Public Economics 197*, 104404.

Ang, A., J. Bai, and H. Zhou (2018). The Great Wall of debt: Real estate, political risk, and Chinese local government financing cost. Georgetown McDonough School of Business Research Paper (2603022), 15–02.

Antill, S. (2020). Are bankruptcy professional fees excessively high? *Available at SSRN 3554835*.

Atkeson, A. G., A. d'Avernas, A. L. Eisfeldt, and P.-O. Weill (2019). Government guarantees and the valuation of american banks. *NBER Macroeconomics Annual 33*(1), 81–145.

Bachas, N., O. S. Kim, and C. Yannelis (2020). Loan guarantees and credit supply. *Journal of Financial Economics*.

Bachas, N., E. Liu, and D. Morrison (2020). Market power in small business lending: A two dimensional bunching approach. *Working paper*.

Bachas, N., E. Liu, and C. Yannelis (2019). Market power in small business lending. Technical report.

Baker, M., D. Bergstresser, G. Serafeim, and J. Wurgler (2018, October). Financing the response to climate change: The pricing and ownership of U.S. green bonds. Working Paper 25194, National Bureau of Economic Research.

Bauer, R. and D. Hann (2010). Corporate environmental management and credit risk. *Available at SSRN 1660470*.

Bauer, R., T. Ruof, and P. Smeets (2021). Get real! Individuals prefer more sustainable investments. *The Review of Financial Studies 34*(8), 3976–4043.

Begenau, J. and T. Landvoigt (2021). Financial regulation in a quantitative model of the modern banking system. Technical report, National Bureau of Economic Research.

Ben Slimane, M., T. Le Guenedal, T. Roncalli, and T. Sekine (2019). ESG investing in corporate bonds: Mind the gap. *Available at SSRN 3683472*.

Benetton, M. (2021). Leverage regulation and market structure: A structural model of the uk mortgage market. *The Journal of Finance 76*(6), 2997–3053.

Berg, F., K. Fabisik, and Z. Sautner (2020). Is history repeating itself? The (un) predictable past of ESG ratings. European Corporate Governance Institute–Finance Working Paper 708.

Berk, J. and J. H. van Binsbergen (2021). The impact of impact investing. *Available at SSRN 3909166*.

Berndt, A., D. Duffie, and Y. Zhu (2020). The decline of too big to fail. *Available at SSRN 3497897*.

Blomquist, S., W. K. Newey, A. Kumar, and C.-Y. Liang (2021). On bunching and identification of the taxable income elasticity. *Journal of Political Economy 129*(8), 000–000.

Bolton, P. and M. Kacperczyk (2021). Do investors care about carbon risk? *Journal of Financial Economics 142*(2), 517–549.

Bondioli, M., M. Goldberg, N. Hu, C. Li, O. Maalaoui, and H. J. Stein (2021). The bloomberg corporate default risk model (drsk) for private firms. *Quantitative Risk Analytics, Bloomberg L.P.*

Bouwman, C. H., S. Hu, and S. A. Johnson (2018). Differential bank behaviors around the dodd–frank act size thresholds. *Journal of Financial Intermediation 34*, 47–57.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru (2018a). Beyond the balance sheet model of banking: Implications for bank regulation and monetary policy. *Working paper*.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru (2018b). Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics 130*(3), 453–483.

Calomiris, C. W., H. Mamaysky, and R. Yang (2020). Measuring the cost of regulation: A text-based approach. Technical report, National Bureau of Economic Research.

Calomiris, C. W. and D. Nissim (2014). Crisis-related shifts in the market valuation of banking activities. *Journal of Financial Intermediation 23*(3), 400–435.

Campello, M., L. Chen, and L. Zhang (2008). Expected returns, yield spreads, and asset pricing tests. *The Review of Financial Studies 21*(3), 1297–1338.

Carey, M. P. (2014). *Cost-benefit and other analysis requirements in the rulemaking process*. Congressional Research Service.

Carletti, E., I. Goldstein, and A. Leonello (2020). The interdependence of bank capital and liquidity. Technical report.

Cattaneo, M. D., M. Jansson, and X. Ma (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association 115*(531), 1449–1455.

CFA (2019). ESG integration in China: Guidance and case studies.

Chava, S. (2014). Environmental externalities and cost of capital. *Management Science 60*(9), 2223–2247.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annual Review of Economics 1*(1), 451–488.

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. *The Quarterly Journal of Economics 126*(2), 749–804.

Choi, J., S. Hoseinzade, S. S. Shin, and H. Tehranian (2020). Corporate bond mutual funds and asset fire sales. *Journal of Financial Economics 138*(2), 432–457.

Chousakos, K. T. and G. B. Gorton (2017). Bank health post-crisis. Technical report, National Bureau of Economic Research.

Coates, J. C. (2014). Cost-benefit analysis of financial regulation: Case studies and implications. *Yale Law Journal 124*, 882.

Cochrane, J. H. (2014). Challenges for cost-benefit analysis of financial regulation. *The Journal of Legal Studies 43*(S2), S63–S105.

Cooper, R., J. Haltiwanger, and J. L. Willis (2007). Search frictions: Matching aggregate and establishment observations. *Journal of Monetary Economics 54*, 56–78.

Corbae, D. and P. D'Erasmo (2020a). Capital buffers in a quantitative model of banking industry dynamics. Technical report, Technical report.

Corbae, D. and P. D'Erasmo (2020b). Rising bank concentration. *Journal of Economic Dynamics and Control 115*, 103877.

Cortés, K. R., Y. Demyanyk, L. Li, E. Loutskina, and P. E. Strahan (2020). Stress tests and small business lending. *Journal of Financial Economics 136*(1), 260–279.

Crifo, P., M.-A. Diaye, and R. Oueghlissi (2017). The effect of countries' ESG ratings on their sovereign borrowing costs. *The Quarterly Review of Economics and Finance 66*, 13–20.

Crippa, M., D. Guizzardi, M. Muntean, E. Schaaf, E. Solazzo, F. Monforti, J. Olivier, and E. Vignati (2020). Fossil CO2 emissions of all world countries.

Dagostino, R. (2018). The impact of bank financing on municipalities' bond issuance and the real economy. Technical report.

Darmouni, O. (2020). Informational frictions and the credit crunch. *The Journal of Finance 75*(4), 2055–2094.

DeFusco, A. A., S. Johnson, and J. Mondragon (2020). Regulating household leverage. *The Review of Economic Studies 87*(2), 914–958.

DeFusco, A. A. and A. Paciorek (2017). The interest rate elasticity of mortgage demand: Evidence from bunching at the conforming loan limit. *American Economic Journal: Economic Policy 9*(1), 210–40.

Degeorge, F., J. Patel, and R. Zeckhauser (1999). Earnings management to exceed thresholds. *The Journal of Business*.

Derwall, J. and K. Koedijk (2009). Socially responsible fixed-income funds. *Journal of Business Finance & Accounting 36*(1-2), 210–229.

Diab, A. and G. M. Adams (2021). ESG assets may hit $53 trillion by 2025, a third of global AUM. Bloomberg Intelligence.

Ding, Y., W. Xiong, and J. Zhang (2022). Issuance overpricing of China's corporate debt securities. *Journal of Financial Economics 144*(1), 328–346.

Edmans, A. (2011). Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics 101*(3), 621–640.

Egan, M., A. Hortaçsu, and G. Matvos (2017). Deposit competition and financial fragility: Evidence from the us banking sector. *American Economic Review 107*(1), 169–216.

Ehlers, T. and F. Packer (2017). Green bond finance and certification. *BIS Quarterly Review September 2017*.

Eisenbach, T. M., D. O. Lucca, and R. M. Townsend (2021). Resource allocation in bank supervision: Trade-offs and outcomes. Technical report.

Ellul, A., C. Jotikasthira, and C. T. Lundblad (2011). Regulatory pressure and fire sales in the corporate bond market. *Journal of Financial Economics 101*(3), 596–620.

Erickson, T. and T. M. Whited (2000). Measurement error and the relationship between investment and q. *Journal of Political Economy 108*(5), 1027–1057.

Ewens, M., K. Xiao, and T. Xu (2020). Regulatory costs of being public: Evidence from bunching estimation. *Working paper*.

Feldman, R., K. Heinecke, and J. Schmidt (2013). Quantifying the costs of additional regulation on community banks. *Economic Policy Paper 13*(3).

Ferriani, F. (2022). Issuing bonds during the COVID-19 pandemic: Is there an ESG premium? *Available at SSRN 4042802*.

Fuster, A., M. C. Plosser, and J. I. Vickery (2018). Does cfpb oversight crimp credit? *FRB of New York Staff Report*.

Gabaix, X. (2016). Power laws in economics: An introduction. *Journal of Economic Perspectives 30*(1), 185–206.

Garicano, L., C. Lelarge, and J. Van Reenen (2016). Firm size distortions and the productivity distribution: Evidence from france. *American Economic Review 106*(11), 3439–79.

Ge, W. and M. Liu (2015). Corporate social responsibility and the cost of corporate bonds. *Journal of Accounting and Public Policy 34*(6), 597–624.

Ghouma, H., H. Ben-Nasr, and R. Yan (2018). Corporate governance and cost of debt financing: Empirical evidence from Canada. *The Quarterly Review of Economics and Finance 67*, 138–148.

Goldstein, I., A. Kopytov, L. Shen, and H. Xiang (2022, April). On ESG investing: Heterogeneous preferences, information, and asset prices. Working Paper 29839, National Bureau of Economic Research.

Gompers, P., J. Ishii, and A. Metrick (2003). Corporate governance and equity prices. *The Quarterly Journal of Economics 118*(1), 107–156.

Goss, A. and G. S. Roberts (2011). The impact of corporate social responsibility on the cost of bank loans. *Journal of Banking & Finance 35*(7), 1794–1810.

Gourio, F. and N. Roys (2014). Size-dependent regulations, firm size distribution, and reallocation. *Quantitative Economics 5*(2), 377–416.

Graham, A. and J. J. Maher. Environmental liabilities, bond ratings, and bond yields. *Environmental Accounting*.

Hachenberg, B. and D. Schiereck (2018). Are green bonds priced differently from conventional bonds? *Journal of Asset Management 19*(6), 371–383.

Halling, M., J. Yu, and J. Zechner (2021). Primary corporate bond markets and social responsibility. Swedish House of Finance Research Paper (20-13).

Harberger, A. C. (1964). The measurement of waste. *American Economic Review 54*(3), 58–76.

Hartzmark, S. M. and A. B. Sussman (2019). Do investors value sustainability? A natural experiment examining ranking and fund flows. *The Journal of Finance 74*(6), 2789–2837.

Hasan, I., C. K. Hoi, Q. Wu, and H. Zhang (2017). Social capital and debt contracting: Evidence from bank loans and public bonds. *Journal of Financial and Quantitative Analysis 52*(3), 1017–1047.

Heinkel, R., A. Kraus, and J. Zechner (2001, December). The effect of green investment on corporate behavior. *Journal of Financial and Quantitative Analysis 36*(4), 431–449.

Hill, A. and Y. Jia (2021). The internationalization of the China corporate bond market. International Capital Market Association (ICMA), Zurich.

Hinkes-Jones, L. (2017). How much did dodd-frank cost? don't ask banks. *The Bureau of National Affairs*.

Hoepner, A., I. Oikonomou, B. Scholtens, and M. Schröder (2016). The effects of corporate and country sustainability characteristics on the cost of debt: An international investigation. *Journal of Business Finance & Accounting 43*(1-2), 158–190.

Hong, H. and M. Kacperczyk (2009, July). The price of sin: The effects of social norms on markets. *Journal of Financial Economics 93*(1), 15–36.

Hou, D. and M. Warusawitharana (2018). Effects of fixed nominal thresholds for enhanced supervision. *FEDS Notes* (2018-07), 19.

Huntington, D. S. (2010). Summary of dodd-frank financial regulation legislation. In *Harvard Law School Forum on Corporate Governance and Financial Regulation*.

Ilhan, E., P. Krueger, Z. Sautner, and L. T. Starks (2021). Climate risk disclosure and institutional investors. *Swiss Finance Institute Research Paper* (19-66).

Iliev, P. (2010). The effect of sox section 404: Costs, earnings quality, and stock prices. *The Journal of Finance 65*(3), 1163–1196.

Izzo, M. F. and B. S. Magnanelli (2012). Does it pay or does firm pay? The relation between CSR performance and the cost of debt. *Available at SSRN 1986131*.

J.P.Morgan (2021). Overcoming the ESG data challenge in China.

Karpf, A. and A. Mandel (2018). The changing value of the 'green' label on the US municipal bond market. *Nature Climate Change 8*(2), 161–165.

Kashyap, A. K., D. P. Tsomocos, and A. P. Vardoulakis (2020). Optimal bank regulation in the presence of credit and run-risk. Technical report, National Bureau of Economic Research.

Kempf, A. and P. Osthoff (2007). The effect of socially responsible investing on portfolio performance. *European Financial Management 13*(5), 908–922.

Kim, M., J. Surroca, and J. A. Tribó (2014). Impact of ethical behavior on syndicated loan rates. *Journal of Banking & Finance 38*, 122–144.

Kisin, R. and A. Manela (2016). The shadow cost of bank capital requirements. *The Review of Financial Studies 29*(7), 1780–1820.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics 8*, 435–464.

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

Klock, M. S., S. A. Mansi, and W. F. Maxwell (2005). Does corporate governance matter to bondholders? *Journal of Financial and Quantitative Analysis 40*(4), 693–719.

Koijen, R. S. and M. Yogo (2015). The cost of financial frictions for life insurers. *American Economic Review 105*(1), 445–75.

Koijen, R. S. and M. Yogo (2019). A demand system approach to asset pricing. *Journal of Political Economy 127*(4), 1475–1515.

La Rosa, F., G. Liberatore, F. Mazzi, and S. Terzani (2018). The impact of corporate social performance on the cost of debt and access to debt financing for listed european non-financial firms. *European Management Journal 36*(4), 519–529.

Lau, P., A. Sze, W. Wan, and A. Wong (2022). The economics of the greenium: How much is the world willing to pay to save the earth? *Environmental and Resource Economics 81*, 379–408.

Luo, H. A. and R. J. Balvers (2017). Social screens and systematic investor boycott risk. *Journal of Financial and Quantitative Analysis 52*(1), 365–399.

Martin, P. R. and D. V. Moser (2016). Managers' green investment disclosures and investors' reaction. *Journal of Accounting and Economics 61*(1), 239–254.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714.

Menz, K.-M. (2010). Corporate social responsibility: Is it rewarded by the corporate bond market? a critical note. *Journal of Business Ethics 96*(1), 117–134.

Moll, B. (2018). Distributional macroeconomics.

Morgan, D. P. and B. Yang (2016, Oct). Fear of $10 billion. *Liberty Street Economics*.

Nandy, M. and S. Lodh (2012). Do banks value the eco-friendliness of firms in their corporate lending decision? Some empirical evidence. *International Review of Financial Analysis 25*, 83–93.

Nelson, C. R. and A. F. Siegel (1987). Parsimonious modeling of yield curves. *Journal of Business*, 473–489.

Nelson, S. (2018). Private information and price regulation in the us credit card market. Technical report.

Nozawa, Y. (2017). What drives the cross-section of credit spreads?: A variance decomposition approach. *The Journal of Finance 72*(5), 2045–2072.

Oikonomou, I., C. Brooks, and S. Pavelin (2014). The effects of corporate social performance on the cost of corporate debt and credit ratings. *Financial Review 49*(1), 49–75.

Parker, R. (2018a). Hyping the cost of regulation. *The Regulatory Review*.

Parker, R. W. (2018b). The faux scholarship foundation of the regulatory rollback movement. *Ecology LQ 45*, 845.

Pastor, L., R. F. Stambaugh, and L. A. Taylor (2021a, June). Dissecting green returns. Working Paper 28940, National Bureau of Economic Research.

Pastor, L., R. F. Stambaugh, and L. A. Taylor (2021b). Sustainable investing in equilibrium. *Journal of Financial Economics 142*, 550–571.

Pedersen, L. H., S. Fitzgibbons, and L. Pomorski (2021). Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics 142(2)*, 572–597.

PingAn (2020). ESG investment in China.

Posner, E. and E. G. Weyl (2013). Benefit-cost analysis for financial regulation. *American Economic Review 103*(3), 393–97.

Posner, R. A. (1971). Taxation by regulation. *The Bell Journal of Economics and Management Science*, 22–50.

PRI (2020). ESG and alpha in China.

Riedl, A. and P. Smeets (2017). Why do investors hold socially responsible mutual funds? *The Journal of Finance 72*(6), 2505–2549.

Robles-Garcia, C. (2019). Competition and incentives in mortgage markets: The role of brokers. Technical report.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy 2*(3), 180–212.

Sarin, N. and L. H. Summers (2016). Have big banks gotten safer? *Brookings Papers on Economic Activity*.

Seltzer, L., L. T. Starks, and Q. Zhu (2020). Climate regulatory risks and corporate bonds. Nanyang Business School Research Paper (20-05).

Sharfman, M. P. and C. S. Fernando (2008). Environmental risk management and the cost of capital. *Strategic Management Journal 29*(6), 569–592.

Simkovic, M. and M. B. Zhang (2020). Regulation and technology-driven entry: Measurement and micro-evidence. Technical report.

Stellner, C., C. Klein, and B. Zwergel (2015). Corporate social responsibility and eurozone corporate bonds: The moderating role of country sustainability. *Journal of Banking & Finance 59*, 538–549.

Viscusi, W. K. and J. E. Aldy (2003). The value of a statistical life: a critical review of market estimates throughout the world. *Journal of Risk and Uncertainty 27*(1), 5–76.

Whited, T. M., Y. Wu, and K. Xiao (2021). Risk incentives for banks with market power when interest rates are low. *Journal of Monetary Economics*.

Xiao, K. (2020). Monetary transmission through shadow banks. *The Review of Financial Studies 33*(6), 2379–2420.

Ye, K. and R. Zhang (2011). Do lenders value corporate social responsibility? Evidence from China. *Journal of Business Ethics 104*(2), 197–206.

Zerbib, O. D. (2019). The effect of pro-environmental preferences on bond prices: Evidence from green bonds. *Journal of Banking & Finance 98*, 39–60.

Zhang, R. L. (2021). *ESG and Cost of Debt*. Stanford University.

Appendix A: Chapter 1

A.1 ESG scores

A.1.1 Sustainalytics

Sustainalytics, which has recently been bought by Morningstar, is a leading ESG research provider for emerging markets, covering over 11,000 companies globally. The company's ESG Risk Ratings aims to identify financially material ESG risks at the firm level by measuring the magnitude of a company's unmanaged ESG risks. More specifically, a company's ESG Risk Rating is comprised of a numeric score and risk category. A higher score indicates more unmanaged risks. The firm scores are updated annually.

Sustainalytics follows a three-step process to find the ESG Risk Ratings. The first step is to find the company's exposure, which measures the company's vulnerability to ESG risks. The second step is to assess the unmanageable risks and management gap. The unmanageable risks are industry specific, while the management gap is company specific. The last step is to simply sum the unmanageable risks and management gap to obtain the unmanaged risks.

A company's exposure is computed as the industry exposure multiplied by the beta of the company. Industry exposure is determined by a set of ESG issues relevant to the industry in which the company operates. The estimate of the industry exposure considers external data (e.g., CO_2 emissions), company reporting, and third-party research. The beta of the company reflects the degree of the deviation of the company's exposure to ESG issues from the industry average. The betas enable the ESG ratings to be company specific. The computation of the betas has three stages: modeling of industry-specific beta indicators, analysis of firm-specific qualitative overlays, and application of technical correction factors.

To quantify the unmanageable risks, Sustainalytics find the industry-specific manageable risk factor (MRF). The determination of the MRF is based on four primary industry factors, hence ensuring the comparability of the ratings across companies and industries. Based on the MRF, Sustainalytics can calculate the unmanageable risks by multiplying the company exposure and (1 - MRF).

The other part of the unmanaged risks is the management gap, which is the failure of the company to sufficiently cope with manageable risks. This is computed as the product of the manageable risks and management score (in %). The management score stems from a set of management indicators and outcome-focused indicators. Management indicators assess a company's policies, management systems, and certifications, while outcome-focused indicators measure management performance, either directly in quantitative terms (e.g., CO_2 emissions) or by assessing a company's involvement in ESG-concerned controversial events.

A.1.2 RepRisk

The main score that RepRisk provides J.P. Morgan with is called the "RepRisk Index" (RRI). It is a score based on an algorithm developed by RepRisk, which dynamically captures a company's reputation risk exposure to ESG issues. The RRI depends on related risk incidents and reflects a company's actual risk management performance, as opposed to its ESG goals and policies. As a result, it is not influenced by potential greenwashing in companies' sustainable reports. It scrapes millions of sources, including, but not limited to, local regulators' reports, regional median outlets, press releases, NGO reports, think tank articles, law journals, blogs, and thought pieces. The RRI ranges from 0 to 100 and is updated daily. A higher value means higher risk exposure. There are three specified RRI indexes that provide dynamic measurements: (1) current RRI reflects the current level of attention that media and stakeholders pay toward ESG issues for a certain company, (2) peak RRI is equal to the highest level of the RRI over the last two years, acting as a proxy for the overall reputation risk exposure related to the ESG issues of a company, and (3) the RRI trend, which shows the increase or decrease of the RRI within the past 30 days. The RRI calculation combines two components: incident intensity and incident value. Incident intensity captures the frequency of risk incidents. It increases when the number of incidents associated with the company has risen in the past two months. Incident values are calculated as a time-weighted mean of the company's incidents from the last two years. RepRisk uses two years because it is their estimated time for a company to be able to mitigate any ESG conduct violation. The incident values are rated based on severity, reach of the information source, and novelty. Severity reflects the harshness of the risk incident or criticism. The reach of the information source is based on readership and circulation, as well as by the importance of the information in a specific country. Novelty reflects the newness of the issues addressed for the company; whether it is the first time a company has been exposed to a specific ESG issue in a certain location. The RepRisk dataset is incident based, which means each risk incident is automatically tagged in the different metrics mentioned above to all the related companies that have been identified in the risk incident.

A.2 Chinese onshore bond market

A special feature of the Chinese bond market is the distinction between enterprise bonds and corporate bonds. Both of these types of bonds are issued by corporations, so they would be interpreted as "corporate bonds" in the U.S. However, they differ in China because of the ownership of their issuers and regulatory supervisor of their issuing activity. Specifically, enterprise bonds are issued by state-owned enterprises (SOEs) that are not listed on stock markets. On average, enterprise bonds have longer maturities than corporate bonds. Moreover, most enterprise bonds are debenture bonds. Since August 2021, enterprise bonds and financial bonds have been exempted from mandatory credit ratings, which means corporate bonds are the only bonds that require credit ratings.[1]

Enterprise and corporate bonds exclude bonds issued by financial institutions. Financial bond issuers include commercial banks, investment banks (securities companies), and insurance companies. Medium-term notes (MTN) and commercial papers (CP) trade on IB only. One interesting fact is that although most MTNs have maturities between three to five years, as in the U.S. market, there are 183 bonds with maturities longer than 10 years. This could be because of the issuance of listed SOEs, who cannot be listed on IB because of the regulation. To access the IB, these publicly listed SOEs use MTNs as long-term instruments to access the IB.

[1] Enterprise bonds and corporate bonds have equal requirements on minimum assets. However, these constraints are never binding for giant SOEs.

A.3 EM corporate client survey methodology

Institutional investors positioning can be approximated using the J.P. Morgan Emerging Markets Client Survey, which is a monthly survey of institutional investors that captures their positioning across 200+ EM fixed-income currency assets. The survey distinguishes between three types of investors:

1. Dedicated investors: investors with dedicated or pure emerging market portfolios.
2. Crossover investors: investors with a benchmark containing small or no allocations to emerging markets.
3. Trading investors: investors with a total return benchmark, which are usually hedge funds or proprietary traders.

The participants are asked to rank their positioning in each assets as "Very Overweight," "Overweight," "Neutral," "Underweight," "Very Underweight." From these answers, J.P. Morgan creates an overall $ExposureScore$ at the asset class level by taking the weighted average of the numerically encoded answers of all N participants. The weights correspond to the share of assets under management of the respondent compared with the total assets under management of the participants. Formally:

$$ExposureScore_t = \sum_{i=1}^{N} w_{i,t} \times score_{i,t},$$

where

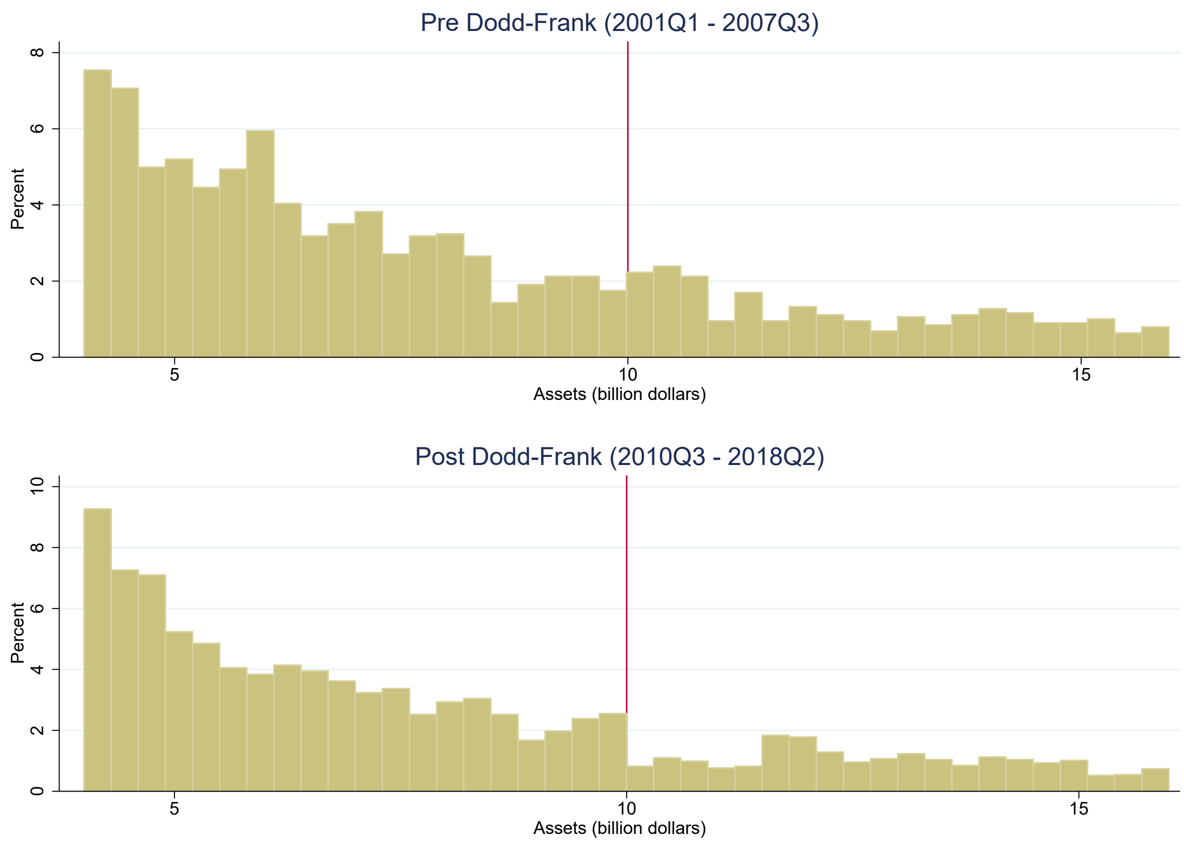
$$w_{i,t} = \frac{AUM_{i,t}}{\sum_{i=1}^{N} AUM_{i,t}}$$

$$score_{i,t} = \begin{cases} +10 & \text{Very Overweight} \\ +5 & \text{Overweight} \\ 0 & \text{Neutral} \\ -5 & \text{Underweight} \\ -10 & \text{Very Underweight} \end{cases}.$$

This is done for several asset classes that have different levels of aggregation.

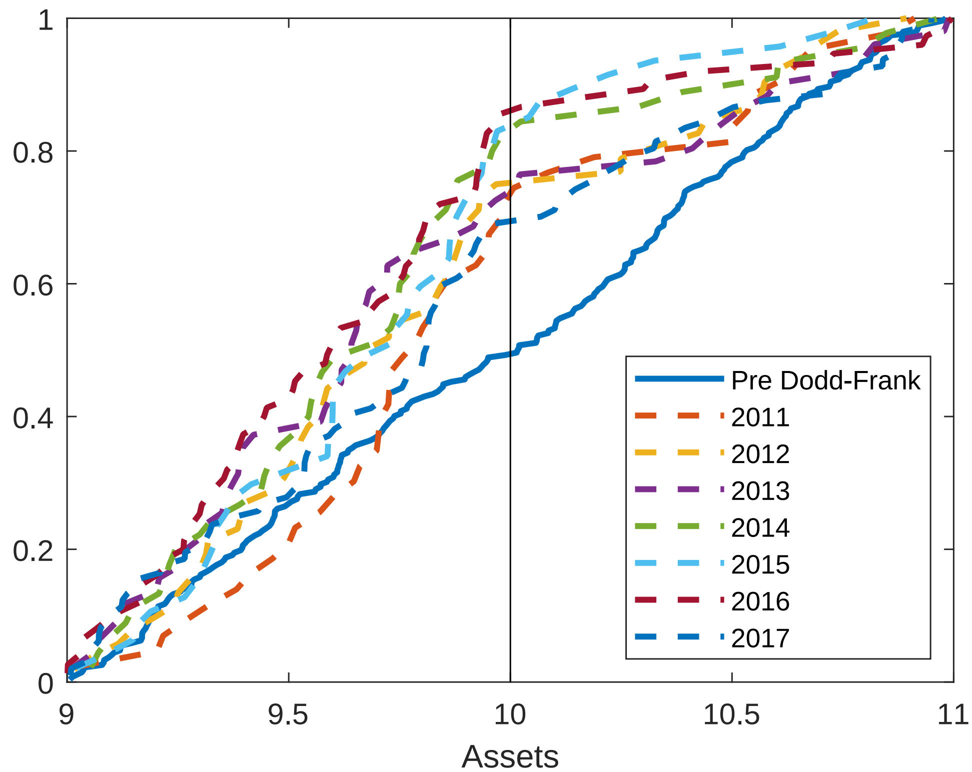
Appendix B: Chapter 2

Figure B.1: Histogram of Bank Size Around Regulatory Threshold



Note: This figure shows the histogram of bank size around the $10 billion threshold before and after the Dodd–Frank Act, respectively.

Figure B.2: Bank Profit Margin Across Size Distribution



Note: This figure shows the empirical cumulative distribution of bank size conditional on the narrow intervals around the $10 billion regulatory threshold year by year. Bank size is measured by quarter-end assets.

Table B.1: Chi-Square Goodness of Fit Test of the Distribution of Bank Assets

Distribution	P-val	Chi-Square Stat.
Power Law	0.36	3.90
Uniform	<0.001	5531.49
Weibull	<0.001	40.31
Normal	<0.001	170.62
Lognormal	<0.001	40.00
Exponential	<0.001	104.17
Extreme Value	<0.001	161.22
Gamma	<0.001	65.72

Note: This table reports the p-values and Chi-square statistics from a Chi-square goodness of fit test of the quarterly distributions of bank assets between $1B and $250B. The sample period is 2001Q1 to 2007Q3. The Chi-square goodness of fit test tests whether the null hypothesis of the data following a particular distribution can be rejected. One can reject the null hypothesis with 5% statistical significance if the p-value is below 0.05.

Table B.2: Maximum Likelihood Estimation of Regulatory Costs with Different Assets Interval

Panel A: $10 billion threshold

		Estimated value	S.E.
β	Exponent of the power law distribution	1.113	[0.001]
σ	Measurement error volatility (in %)	4.258	[0.386]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	10.973	[0.086]
τ	Cost of regulation (% of profit)	0.405	[0.066]

Panel B: $50 billion threshold

		Estimated value	S.E.
β	Exponent of the power law distribution	1.085	[0.002]
σ	Measurement error volatility (in %)	2.290	[0.498]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	52.393	[0.537]
τ	Cost of regulation (% of profit)	0.106	[0.046]

Note: This table presents the results of the maximum likelihood estimation of the regulatory costs using the power law distribution. The sample period is from 2010Q3 to 2018Q2. The estimation sample of Panel A includes banks with assets between $3 billion and $30 billion. The estimation sample of Panel B includes banks with assets above $30 billion. Standard errors are obtained via the inverse of the Hessian matrix.

Table B.3: Maximum Likelihood Estimation of Regulatory Costs using Both Pre and Post Dodd–Frank Data

Panel A: $10 billion threshold

		Estimated value	S.E.
β	Exponent of the power law distribution	1.112	[0.001]
σ	Measurement error volatility (in %)	4.261	[0.386]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	10.974	[0.086]
τ	Cost of regulation (% of profit)	0.406	[0.066]

Panel B: $50 billion threshold

		Estimated value	S.E.
β	Exponent of the power law distribution	1.084	[0.002]
σ	Measurement error volatility (in %)	2.291	[0.498]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	52.393	[0.528]
τ	Cost of regulation (% of profit)	0.106	[0.046]

Note: This table presents the results of the maximum likelihood estimation of the regulatory costs using the power law distribution. The sample period is from 2010Q3 to 2018Q2 for the post Dodd–Frank period, and from 2001Q1 to 2007Q3 for the pre Dodd–Frank period (which excludes the financial crisis). The estimation sample of Panel A includes banks with assets between $3 billion and $40 billion. The estimation sample of Panel B includes banks with assets above $40 billion. Standard errors are obtained via the inverse of the Hessian matrix.

Table B.4: Counterfactual Simulation: Robustness

Counterfactual	Lower discount rate	Higher entry cost	Lower exit rate
		Panel (a): all banks	
Mass of banks	-0.236%	-0.193 %	-0.512 %
Market-to-book	-0.234%	-0.223 %	-0.167 %
Lending quantity	-0.085%	-0.067 %	-0.255 %
Lending rate	0.060%	0.047 %	0.179 %
Output	-0.026%	-0.020 %	-0.077 %
		Panel (b): annual profits by size group	
Small banks	0.096%	0.076 %	0.062 %
Medium banks	-0.395%	-0.399 %	-0.405 %
Big banks	-1.246%	-1.272 %	-0.682 %
		Panel (c): asset shares by size group	
Small banks	-0.067%	-0.063 %	-0.231 %
Medium banks	-0.247%	-0.230 %	-0.354 %
Big banks	0.028%	0.022 %	0.012 %
		Panel (d): shares of banks by size group	
Small banks	-0.012%	-0.012 %	-0.035 %
Medium banks	0.086%	0.083 %	0.035 %
Big banks	0.100%	0.083 %	0.035 %

Note: This table reports the results of the robustness of counterfactual simulations with respect to alternative parameter values. The values are expressed as the percentage changes from the baseline economy without the Dodd–Frank Act to the the economy with the Dodd–Frank Act. Column (1) reports the results when the discount rate is decreased by 10% relative to the baseline value. Columns (2) reports the results when the entry cost is increased by 10% relative to the baseline value. Column (3) reports the result when the exit rate is calibrated to the value used by Corbae and D'Erasmo (2020b) (1%), which only includes bank failures recorded by FDIC.

Table B.5: Maximum Likelihood Estimation of Regulatory Costs using Alternative Regulatory Assets

Panel A: Average of past four quarters

		Estimated value	S.E.
β	Exponent of the power law distribution	1.112	[0.001]
σ	Measurement error volatility (in %)	5.182	[0.582]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	11.003	[0.110]
τ	Cost of regulation (% of profit)	0.429	[0.087]

Panel B: End of last calendar year

		Estimated value	S.E.
β	Exponent of the power law distribution	1.113	[0.001]
σ	Measurement error volatility (in %)	4.488	[0.343]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	11.069	[0.085]
τ	Cost of regulation (% of profit)	0.483	[0.071]

Panel C: Minimum of the most recent four quarters

		Estimated value	S.E.
β	Exponent of the power law distribution	1.112	[0.001]
σ	Measurement error volatility (in %)	4.339	[0.385]
$\exp(\overline{q})$	Assets of marginal bank ($ Billion)	10.966	[0.082]
τ	Cost of regulation (% of profit)	0.400	[0.063]

Note: This table presents the results of the maximum likelihood estimation of the regulatory costs at the $10 billion threshold using the power law distribution. The sample period is from 2010Q3 to 2018Q2 for the post Dodd–Frank period. The estimation sample of includes banks with assets between $3 billion and $40 billion. Total consolidated assets are measured as the average of the most recent four quarters, most recent end-of-calendar-year value, and the minimum value in the most recent four quarters for panels A, B, and C, respectively. Standard errors are obtained via the inverse of the Hessian matrix.

B.1 Institutional background of the CBA of financial regulation

The major statutes that apply to financial regulators include (1) the Paperwork Reduction Act (PRA), which requires regulators to justify the collection of information from the public to minimize its burden and maximize the utility of information collected;[1] (2) the Regulatory Flexibility Act (RFA), which requires regulators to assess and consider alternatives to the burden of regula-

[1] Pub. L. No. 104–13, 109 Stat. 163 (1995) (codified at 44 U.S.C. Â§Â§ 3501-3520).

tion on small entities;[2] and (3) the Congressional Review Act (CRA), which requires regulators to submit proposed rules—along with any cost-benefit analysis the agencies have conducted—to Congress and the Government Accountability Office (GAO).[3]

The cost-benefit analysis often forms the basis for judicial review and Congressional oversight of regulatory actions. For instance, in the 2002 SEC mutual fund governance reform, the regulator proposed to increase the required share of independent directors from 50% to 75% for mutual funds. Following the SEC's proposal, the Chamber of Commerce, a business-oriented interest group, sued the SEC for failure to quantify the costs of the rule. In its defense, the SEC argued that staffing would be discretionary, and the SEC had no basis for knowing how many chairs would hire staff or how many staff each chair would hire. However, the court rejected this argument and maintained that "estimating the costs for an individual fund is pertinent to an assessment of the requirement."[4] In the end, the rule was struck down by the court.

[2]Pub. L. No. 96-354, 94 Stat. 1164 (1980) (codified at 5 U.S.C. Â§Â§ 601–612).
[3]Pub. L. No. 104–121, 110 Stat. 847 (1996) (codified as amended in 5 U.S.C. Â§ 801 et seq.).
[4]Chamber of Commerce v. SEC, 412 F.3d 133 (D.C. Cir. 2005).

B.2 Derivation of Regulatory Costs

Provided the different regulatory thresholds are far enough from each other, we can derive the regulatory cost τ_i in the same manner for all i. Therefore, we drop the subscript i in the following derivation. We know from the profit indifference condition of the marginal bank in equation (2.7) that τ can be written as follows:

$$1 - \tau = \frac{(R - r(\underline{q}|\overline{z}))\exp\left(\underline{q}\right)}{(R - r\left(\overline{q}|\overline{z}\right))\exp\left(\overline{q}\right)}. \tag{B.1}$$

We conduct a Taylor expansion for $r(\underline{q}|\overline{z})$ at the optimal size, $\overline{q}$

$$r\left(\underline{q}|\overline{z}\right) = r\left(\overline{q}|\overline{z}\right) + r_q\left(\overline{q}|\overline{z}\right)\left(\underline{q} - \overline{q}\right) + O\left((\underline{q} - \overline{q})^2\right). \tag{B.2}$$

The numerator of equation (B.1) becomes

$$\begin{aligned}\left(R - r\left(\underline{q}|\overline{z}\right)\right)\exp\left(\underline{q}\right) &= \left\{R - r\left(\overline{q}|\overline{z}\right) - r_q\left(\overline{q}|\overline{z}\right)\left(\underline{q} - \overline{q}\right) - O\left((\underline{q} - \overline{q})^2\right)\right\}\exp\left(\underline{q}\right) \\ &= \left\{(R - r\left(\overline{q}|\overline{z}\right))\left(1 - \underline{q} + \overline{q}\right) - O\left((\underline{q} - \overline{q})^2\right)\right\}\exp\left(\underline{q}\right),\end{aligned} \tag{B.3}$$

where the second equality uses the result of equation (2.2) that the profit margin at the optimal size equals the inverse semi-elasticity $R - r\left(\overline{q}|\overline{z}\right) = r_q\left(\overline{q}|\overline{z}\right)$. Plugging in equation (B.3) into equation (B.1), we get

$$1 - \tau = \left(1 - \underline{q} + \overline{q}\right)\exp\left(\underline{q} - \overline{q}\right) - O\left(\left(\underline{q} - \overline{q}\right)^2\exp\left(\underline{q} - \overline{q}\right)\right). \tag{B.4}$$

Thus, when $\underline{q}$ is close to $\overline{q}$, we get

$$1 - \tau \simeq \left(\overline{q} - \underline{q} + 1\right)\exp\left(\underline{q} - \overline{q}\right). \tag{B.5}$$

When the inverse semi-elasticity of funding supply does not depend on the size, i.e., $r_{qq} = 0$, the higher-order term $O\left((\underline{q} - \overline{q})^2 \exp\left(\underline{q} - \overline{q}\right)\right) = \frac{1}{2(R - r(\overline{q}|\overline{z}))} r_{qq} \left(\underline{q} - \overline{q}\right)^2 \exp\left(\underline{q} - \overline{q}\right)$ in the Taylor expansion disappears and the approximation becomes exact.

B.3 Heterogeneity in bank profit margin

This appendix considers the effect of heterogeneous demand semi-elasticity on the sufficient statistic formula of regulatory cost, equation (2.8). Suppose the funding supply function is $r(q|z,\theta) = \frac{1}{\theta}(q-z)$ and the semi-elasticity, θ, follows a distribution $x(\theta)$ so that it is heterogeneous across banks. Consider the bunching decision for a sub-sample of banks which has semi-elasticity θ:

$$\max_q \pi(q|z,\theta) = \max_q (R - \frac{1}{\theta}(q-z)) \exp(q) \cdot (1 - \tau \mathbb{1}_{q \geq \underline{q}}). \tag{B.6}$$

The indifference condition between bunching and not bunching of the marginal bank within this subset is given by:

$$1 - \tau = \frac{\left(R - \frac{1}{\theta}(\underline{q} - \overline{z}_\theta)\right) \exp\left(\underline{q}\right)}{\left(R - \frac{1}{\theta}(\overline{q}_\theta - \overline{z}_\theta)\right) \exp\left(\overline{q}_\theta\right)}, \tag{B.7}$$

where the optimal size when the bank does not bunch is given by

$$\overline{q}_\theta = \theta R + \overline{z}_\theta - 1. \tag{B.8}$$

The subscript θ indicates that the log asset size and the productivity may potentially depend on θ. Using the above optimal size equation to replace $\overline{z}_\theta$ with $\overline{q}_\theta$ in the first term in the numerator of equation (B.7), we get

$$\begin{aligned} &R - \frac{1}{\theta}(\underline{q} - \overline{z}_\theta) \\ =&R - \frac{1}{\theta}(\underline{q} - \overline{q}_\theta + \theta R - 1) \\ =&\frac{1}{\theta}(\overline{q}_\theta - \underline{q} + 1) \end{aligned} \tag{B.9}$$

The first term in the denominator of equation (B.7) is the undistorted profit margin, which equals the reciprocal of the semi-elasticity:

$$R - \frac{1}{\theta}(\overline{q}_\theta - \overline{z}_\theta) = \frac{1}{\theta}. \tag{B.10}$$

We can plug the above two equations into equation (B.7), we get

$$1-\tau=\frac{\frac{1}{\theta}(\overline{q}_{\theta}-\underline{q}+1)\exp\left(\underline{q}\right)}{\frac{1}{\theta}\exp\left(\overline{q}_{\theta}\right)}, \tag{B.11}$$

The semi-elasticity cancels out and we get:

$$\tau=1-\left(\overline{q}_{\theta}-\underline{q}+1\right)\exp\left(\underline{q}-\overline{q}_{\theta}\right). \tag{B.12}$$

Note that the size of the marginal bank only depends on τ, but not θ. Therefore, we can drop the subscript θ and we get the same sufficient statistic formula as equation (2.8):

$$\tau=1-\left(\overline{q}-\underline{q}+1\right)\exp\left(\underline{q}-\overline{q}\right). \tag{B.13}$$

In other words, the size of the marginal bank in the sub-sample of banks with semi-elasticity θ is the same as that in a different sub-sample of banks with a different semi-elasticity θ'. The heterogeneity in θ is absorbed by z. The above derivation shows that the mapping between the proportional regulatory tax, τ, and the marginal bunching banks' size, $\overline{q}$, does not depend on the value of semi-elasticity θ. As a result, the heterogeneous semi-elasticity does not affect the estimate of the proportional regulatory tax. However, the heterogeneous semi-elasticity does affect the dollar value of the regulatory costs, $\tau\frac{1}{\theta}\exp(q)$. Intuitively, for a given percentage of assets that banks are willing to give up, a lower semi-elasticity implies that banks forego more profits. Thus, the implied dollar value of regulatory cost is higher. This heterogeneity can be accounted for by multiplying the proportional tax rate with bank-specific expected profits.

B.4 Derivation of the stationary distribution

Guess $g(z) = c_z \exp(z)^{-\beta}$ and plug it to the Kolmogorov forward equation (2.14):

$$0 = -\frac{\partial}{\partial z}\left[\mu_z c_z \exp(z)^{-\beta}\right] + \frac{1}{2}\frac{\partial^2}{\partial z^2}\left[\sigma_z^2 c_z \exp(z)^{-\beta}\right] - \lambda c_z \exp(z)^{-\beta}, \tag{B.14}$$

for $z \neq z_n$. This leads to a quadratic equation

$$0 = -\mu_z \beta + \frac{1}{2}\sigma_z^2 \beta(\beta + 1) - \lambda. \tag{B.15}$$

This quadratic equation always has two real roots $\beta_+ > 0$ and $\beta_- < 0$ as long as $\lambda > 0$. The general solution to the Kolmogorov forward equation is

$$g(z) = c_- \exp(z)^{-\beta_-} + c_+ \exp(z)^{-\beta_+}. \tag{B.16}$$

$g(z)$ is integrable, which implies that $c_- = 0$ for $z > z_n$ and $c_+ = 0$ for $z < z_n$ (Moll, 2018). Therefore, the stationary distribution of the productivity is a double power law distribution.

$$g(z) = \begin{cases} c_z \exp(z - z_n)^{-\beta_-} & z < z_n \\ c_z \exp(z - z_n)^{-\beta_+} & z > z_n, \end{cases} \tag{B.17}$$

where z_n is the productivity of the new entrants.

Assuming the semi-elasticity of funding supply is constant, the undistorted log asset is a linear function of productivity according to equation (2.3). The stationary distribution of bank size is also a double power law distribution in the absence of regulatory distortion.

B.5 Derivation of maximum likelihood estimator: Power law

Define h and H as the probability density function and cumulative distribution function of the undistorted bank log-assets q. Using banks' optimal size choice described in equation (2.4), the theoretical density function of banks' log-assets with one regulatory threshold is given by the following distribution function

$$q \sim \begin{cases} h(q) & q \in (-\infty, \underline{q}) \\ H(\overline{q}) - H(\underline{q}) & q = \underline{q} \\ 0 & q \in (\underline{q}, \overline{q}] \\ h(q) & q \in (\overline{q}, \infty). \end{cases} \tag{B.18}$$

The empirical bank size a is observed with a structural error $u = \sigma v$ with $v \sim N(0, 1)$. Denote $\mathbb{P}(x \leq a|v)$ as the probability of observing log assets x below a given value a for a given realization of the structural error v. We have:

$$\mathbb{P}(x \leq a|v) = \mathbb{P}(q + \sigma v \leq a|v) = \mathbb{P}(q \leq a - \sigma v|v). \tag{B.19}$$

The above equation shows that the probability of observing size x below a given value a for a given realization of the structural error σv is the same as the probability that the theoretical size q is below $a - \sigma v$.

Using equations (B.18) and (B.19), the conditional cumulative distribution of the observed size a is given by:

$$\mathbb{P}(x \leq a|v) = \begin{cases} H(a - \sigma v) & a - \sigma v \in (-\infty, \underline{q}) \\ H(\overline{q}) & a - \sigma v \in [\underline{q}, \overline{q}] \\ H(a - \sigma v) & a - \sigma v \in (\overline{q}, \infty). \end{cases} \tag{B.20}$$

We can change the conditions such that they are written with respect to the structural error, v:

$$\mathbb{P}\left(x \le a|v\right) = \begin{cases} H(a-\sigma v) & v \in (\frac{1}{\sigma}(a-\underline{q}), \infty) \\ H(\overline{q}) & v \in [\frac{1}{\sigma}(a-\overline{q}), \frac{1}{\sigma}(a-\underline{q})] \\ H(a-\sigma v) & v \in (-\infty, \frac{1}{\sigma}(a-\overline{q})). \end{cases} \tag{B.21}$$

The unconditional cumulative distribution function of observed size a can be written as the convolution of the conditional CDF of q and the distribution of the structural error v:

$$\begin{aligned} \mathbb{P}\left(x \le a\right) =& \mathbb{E}\left[\mathbb{P}\left(x \le a \mid v\right)\right] \\ =& \int_{\frac{1}{\sigma}\left(a-\underline{q}\right)}^{\infty} H(a-\sigma v)\phi\left(v\right) dv \\ &+ H(\overline{q})\left[\Phi\left(\frac{1}{\sigma}\left(a-\underline{q}\right)\right) - \Phi\left(\frac{1}{\sigma}\left(a-\overline{q}\right)\right)\right] \\ &+ \int_{-\infty}^{\frac{1}{\sigma}(a-\overline{q})} H(a-\sigma v)\phi\left(v\right) dv. \end{aligned} \tag{B.22}$$

We take the derivative of equation (B.22) to get the probability density function of the observed size a:

$$\begin{aligned} f\left(a\right) =& \frac{d}{da}\mathbb{E}\left[\mathbb{P}\left(x \le a \mid v\right)\right] \\ =& \frac{d}{da}\int_{\frac{1}{\sigma}\left(a-\underline{q}\right)}^{\infty} H(a-\sigma v)\phi\left(v\right) dv \\ &+ \frac{d}{da}H(\overline{q})\left[\Phi\left(\frac{1}{\sigma}\left(a-\underline{q}\right)\right) - \Phi\left(\frac{1}{\sigma}\left(a-\overline{q}\right)\right)\right] \\ &+ \frac{d}{da}\int_{-\infty}^{\frac{1}{\sigma}(a-\overline{q})} H(a-\sigma v)\phi\left(v\right) dv. \end{aligned} \tag{B.23}$$

The second term in equation (B.23) is straightforward to compute:

$$\begin{aligned} &\frac{d}{da}H(\overline{q})\left[\Phi\left(\frac{1}{\sigma}\left(a-\underline{q}\right)\right) - \Phi\left(\frac{1}{\sigma}\left(a-\overline{q}\right)\right)\right] \\ =& \frac{1}{\sigma}H(\overline{q})\left[\phi\left(\frac{1}{\sigma}\left(a-\underline{q}\right)\right) - \phi\left(\frac{1}{\sigma}\left(a-\overline{q}\right)\right)\right]. \end{aligned}$$

For the first and third terms in equation (B.23), we use the Leibniz formula:

$$\frac{d}{da}\int_{t_1(a)}^{t_2(a)} f(a,v)\,dv = f(a,t_2(a))\,t_2'(a) - f(a,t_1(a))\,t_1'(a) + \int_{t_1(a)}^{t_2(a)} f_a(a,v)\,dv.$$

Using the Leibniz formula, the first term in equation (B.23) becomes:

$$\frac{d}{da}\int_{\frac{1}{\sigma}\left(a-\underline{q}\right)}^{\infty} H(a-\sigma v)\,\phi(v)\,dv = -\frac{1}{\sigma}H\left(\underline{q}\right)\phi\left(\frac{1}{\sigma}\left(a-\underline{q}\right)\right) + \int_{\frac{1}{\sigma}\left(a-\underline{q}\right)}^{\infty} h(a-\sigma v)\,\phi(v)\,dv,$$

given that $\frac{1}{\sigma}H(-\infty)\cdot\phi(\infty)=0$.

The third term in equation (B.23) becomes:

$$\frac{d}{da}\int_{-\infty}^{\frac{1}{\sigma}(a-\overline{q})} H(a-\sigma v)\,\phi(v)\,dv = \frac{1}{\sigma}H(\overline{q})\,\phi\left(\frac{1}{\sigma}(a-\overline{q})\right) + \int_{-\infty}^{\frac{1}{\sigma}(a-\overline{q})} h(a-\sigma v)\,\phi(v)\,dv,$$

given that $\frac{1}{\sigma}H(\infty)\cdot\phi(\infty)=0$. In summary, the probability density of the observed bank size is given by:

$$\begin{aligned} f(a) = &-\frac{1}{\sigma}H\left(\underline{q}\right)\phi\left(\frac{1}{\sigma}\left(a-\underline{q}\right)\right) + \int_{\frac{1}{\sigma}\left(a-\underline{q}\right)}^{\infty} h(a-\sigma v)\,\phi(v)\,dv \\ &+\frac{1}{\sigma}H(\overline{q})\left[\phi\left(\frac{1}{\sigma}\left(a-\underline{q}\right)\right) - \phi\left(\frac{1}{\sigma}(a-\overline{q})\right)\right] \\ &+\frac{1}{\sigma}H(\overline{q})\,\phi\left(\frac{1}{\sigma}(a-\overline{q})\right) + \int_{-\infty}^{\frac{1}{\sigma}(a-\overline{q})} h(a-\sigma v)\,\phi(v)\,dv. \end{aligned}$$

By simplifying the terms, we have:

$$f\left(a\right) = \underbrace{\frac{1}{\sigma}\left[\phi\left(\frac{1}{\sigma}\left(a-\underline{q}\right)\right)\right]\left(H\left(\overline{q}\right)-H\left(\underline{q}\right)\right)}_{\text{Term 1}} + \underbrace{\int_{\frac{1}{\sigma}\left(a-\underline{q}\right)}^{\infty} h\left(a-\sigma v\right)\phi\left(v\right)dv}_{\text{Term 2}} + \underbrace{\int_{-\infty}^{\frac{1}{\sigma}\left(a-\overline{q}\right)} h\left(a-\sigma v\right)\phi\left(v\right)dv}_{\text{Term 3}}, \tag{B.24}$$

We now assume the undistorted bank assets follow a power-law distribution, $\exp(q) \sim c\exp(q)^{-\beta}$ over the strictly positive support $[q_{min}, \infty]$, where q_{min} corresponds to the minimum theoretical log-asset of a bank that could be observed. Given that we defined q as log-assets of a bank (see Section 3.2), the probability density function of q is:

$$\begin{aligned} h\left(q\right) &= \frac{d}{dq}\mathbb{P}\left(\exp(x) \leq \exp(q)\right) \\ &= \frac{d}{dq}\left(\frac{c}{1-\beta}e^{(1-\beta)q}\right) \\ &= c\exp\left((1-\beta)q\right), \end{aligned} \tag{B.25}$$

for $q \geq q_{min}$, and $h(q) = 0$ for $q < q_{min}$. The scaling parameter c is constrained to be equal to $(\beta-1)/\exp\left((1-\beta)q_{min}\right)$ such that the distribution integrates to one.

Plugging in equation (B.25) into equation (B.24), we can derive the first term of equation (B.24) as:

$$\text{Term 1} = \frac{1}{\sigma}\frac{c}{1-\beta}\left[\phi\left(\frac{1}{\sigma}\left(a-\underline{q}\right)\right)\right]\left(\exp\left((1-\beta)\,\overline{q}\right)-\exp\left((1-\beta)\,\underline{q}\right)\right).$$

The second term of equation (B.24) is:

$$
\begin{aligned}
\text{Term 2} = & \int_{\frac{1}{\sigma}\left(a-\underline{q}\right)}^{\infty} \mathbf{1}_{\{a-\sigma v \geq q_{min}\}} c \exp\left(\left(1-\beta\right)\left(a-\sigma v\right)\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}v^2\right) dv \\
=& c \exp\left(\left(1-\beta\right)a + \frac{1}{2}\left(\beta-1\right)^2\sigma^2\right) \int_{\frac{1}{\sigma}\left(a-\underline{q}\right)}^{\infty} \mathbf{1}_{\left\{v \leq \frac{1}{\sigma}\left(a-q_{min}\right)\right\}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(v-\left(\beta-1\right)\sigma\right)^2\right) dv \\
=& c \exp\left(\left(1-\beta\right)a + \frac{1}{2}\left(\beta-1\right)^2\sigma^2\right) \cdot \\
& \left(\Phi\left(\frac{1}{\sigma}\left(a-q_{min}\right) - \left(\beta-1\right)\sigma\right) - \Phi\left(\frac{1}{\sigma}\left(a-\underline{q}\right) - \left(\beta-1\right)\sigma\right)\right).
\end{aligned}
$$

The third term of equation (B.24) is:

$$
\begin{aligned}
\text{Term 3} = & \int_{-\infty}^{\frac{1}{\sigma}\left(a-\overline{q}\right)} \mathbf{1}_{\{a-\sigma v \geq q_{min}\}} c \exp\left(\left(1-\beta\right)\left(a-\sigma v\right)\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}v^2\right) dv \\
=& c \exp\left(\left(1-\beta\right)a + \frac{1}{2}\left(\beta-1\right)^2\sigma^2\right) \Phi\left(\frac{1}{\sigma}\left(a-\overline{q}\right) - \left(\beta-1\right)\sigma\right).
\end{aligned}
$$

To summarize, the probability density of the observed bank size is given by:

$$
\begin{aligned}
f\left(a\right) =& \frac{1}{\sigma}\frac{c}{1-\beta}\left[\phi\left(\frac{1}{\sigma}\left(a-\underline{q}\right)\right)\right]\left(\exp\left(\left(1-\beta\right)\overline{q}\right) - \exp\left(\left(1-\beta\right)\underline{q}\right)\right) \\
& + c \exp\left(\left(1-\beta\right)a + \frac{1}{2}\left(\beta-1\right)^2\sigma^2\right) \cdot \\
& \left(\Phi\left(\frac{1}{\sigma}\left(a-q_{min}\right) - \left(\beta-1\right)\sigma\right) - \Phi\left(\frac{1}{\sigma}\left(a-\underline{q}\right) - \left(\beta-1\right)\sigma\right) + \Phi\left(\frac{1}{\sigma}\left(a-\overline{q}\right) - \left(\beta-1\right)\sigma\right)\right).
\end{aligned}
\tag{B.26}
$$

Taking the log of equation (B.26), we can get the log-likelihood function for the MLE. Note that when $\tau = 0$, we have $\overline{q} = \underline{q}$. The first term of equation (B.26) equals 0 and this equation simplifies to:

$$
f\left(a\right) = c \exp\left(\left(1-\beta\right)a + \frac{1}{2}\left(\beta-1\right)^2\sigma^2\right) \Phi\left(\frac{1}{\sigma}\left(a-q_{min}\right) - \left(\beta-1\right)\sigma\right).
$$

B.6 Derivation of maximum likelihood estimator: Log-normal

The derivation of the maximum likelihood estimator under the log-normal distribution is the same as the power-law distribution until equation (B.25), where the undistorted asset size follows a log-normal distribution instead. As a result, the undistorted log-assets q follows a normal distribution with probability density function:

$$h\left(q\right) = \frac{1}{\sigma_q}\phi\left(\frac{q - \mu_q}{\sigma_q}\right), \tag{B.27}$$

where μ_q denotes the mean undistorted log-assets, and σ_q denotes the standard deviation. Plugging in equation (B.27) into equation (B.24), we can derive the first term of equation (B.24) as:

$$\text{Term 1} = \frac{1}{\sigma}\left[\phi\left(\frac{1}{\sigma}\left(a - \underline{q}\right)\right)\right]\left(\Phi\left(\frac{\overline{q} - \mu_q}{\sigma_q}\right) - \Phi\left(\frac{\underline{q} - \mu_q}{\sigma_q}\right)\right).$$

The second term of equation (B.24) is:

$$
\begin{aligned}
\text{Term 2} =& \int_{\frac{1}{\sigma}\left(a-\underline{q}\right)}^{\infty} \frac{1}{\sigma_q}\phi\left(\frac{a-\sigma v-\mu_q}{\sigma_q}\right)\phi\left(v\right)dv \\
=& \frac{1}{2\pi}\frac{1}{\sigma_q}\int_{\frac{1}{\sigma}\left(a-\underline{q}\right)}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{a-\sigma v-\mu_q}{\sigma_q}\right)^2-\frac{1}{2}v^2\right)dv \\
=& \frac{1}{2\pi}\frac{1}{\sigma_q}\int_{\frac{1}{\sigma}\left(a-\underline{q}\right)}^{\infty} \exp\left(-\frac{1}{2}\frac{\left(\sigma^2+\sigma_q^2\right)v^2-2\left(a-\mu_q\right)\sigma v+\left(a-\mu_q\right)^2}{\sigma_q^2}\right)dv \\
=& \frac{1}{2\pi}\frac{1}{\sigma_q}\int_{\frac{1}{\sigma}\left(a-\underline{q}\right)}^{\infty} \exp\left(-\frac{1}{2}\frac{\left(\sigma^2+\sigma_q^2\right)\left(v-\frac{a-\mu_q}{\sigma^2+\sigma_q^2}\sigma\right)^2-\frac{\left(a-\mu_q\right)^2}{\sigma^2+\sigma_q^2}\sigma^2+\left(a-\mu_q\right)^2}{\sigma_q^2}\right)dv \\
=& \frac{1}{\sigma_q}\exp\left(\frac{\frac{\left(a-\mu_q\right)^2}{\sigma^2+\sigma_q^2}\sigma^2-\left(a-\mu_q\right)^2}{2\sigma_q^2}\right)\frac{\sqrt{\frac{\sigma_q^2}{\sigma^2+\sigma_q^2}}}{\sqrt{2\pi}} \\
&\cdot\frac{1}{\sqrt{2\pi\frac{\sigma_q^2}{\left(\sigma^2+\sigma_q^2\right)}}}\int_{\frac{1}{\sigma}\left(a-\underline{q}\right)}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{v-\frac{a-\mu_q}{\sigma^2+\sigma_q^2}\sigma}{\sqrt{\frac{\sigma_q^2}{\sigma^2+\sigma_q^2}}}\right)^2\right)dv \\
=& \frac{1}{\sigma_q}\frac{\sqrt{\frac{\sigma_q^2}{\sigma^2+\sigma_q^2}}}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{\left(a-\mu_q\right)^2}{\sigma^2+\sigma_q^2}\right) \\
&\cdot\frac{1}{\sqrt{2\pi\frac{\sigma_q^2}{\sigma^2+\sigma_q^2}}}\int_{\frac{1}{\sigma}\left(a-\underline{q}\right)}^{\infty} \exp\left(-\frac{\sigma^2+\sigma_q^2}{2\sigma_q^2}\left(v-\sigma\frac{a-\mu_q}{\sigma^2+\sigma_q^2}\right)^2\right)dv \\
=& \frac{1}{\sqrt{2\pi\left(\sigma^2+\sigma_q^2\right)}}\exp\left(-\frac{1}{2}\frac{\left(a-\mu_q\right)^2}{\sigma^2+\sigma_q^2}\right)\left\{1-\Phi\left(\frac{\frac{1}{\sigma}\left(a-\underline{q}\right)-\sigma\frac{a-\mu_q}{\sigma^2+\sigma_q^2}}{\sqrt{\frac{\sigma_q^2}{\sigma^2+\sigma_q^2}}}\right)\right\}.
\end{aligned}
$$

The third term of equation (B.24) is computed similarly:

$$\text{Term } 3 = \int_{-\infty}^{\frac{1}{\sigma}(a-\overline{q})} \frac{1}{\sigma_q} \phi\left(\frac{a-\sigma v-\mu_q}{\sigma_q}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}v^2\right) dv$$

$$= \frac{1}{\sqrt{2\pi\left(\sigma^2+\sigma_q^2\right)}} \exp\left(-\frac{1}{2}\frac{\left(a-\mu_q\right)^2}{\sigma^2+\sigma_q^2}\right) \Phi\left(\frac{\frac{1}{\sigma}\left(a-\bar{q}\right)-\sigma\frac{a-\mu_q}{\sigma^2+\sigma_q^2}}{\sqrt{\frac{\sigma_q^2}{\sigma^2+\sigma_q^2}}}\right).$$

To summarize, the probability density of the observed bank log-assets is given by:

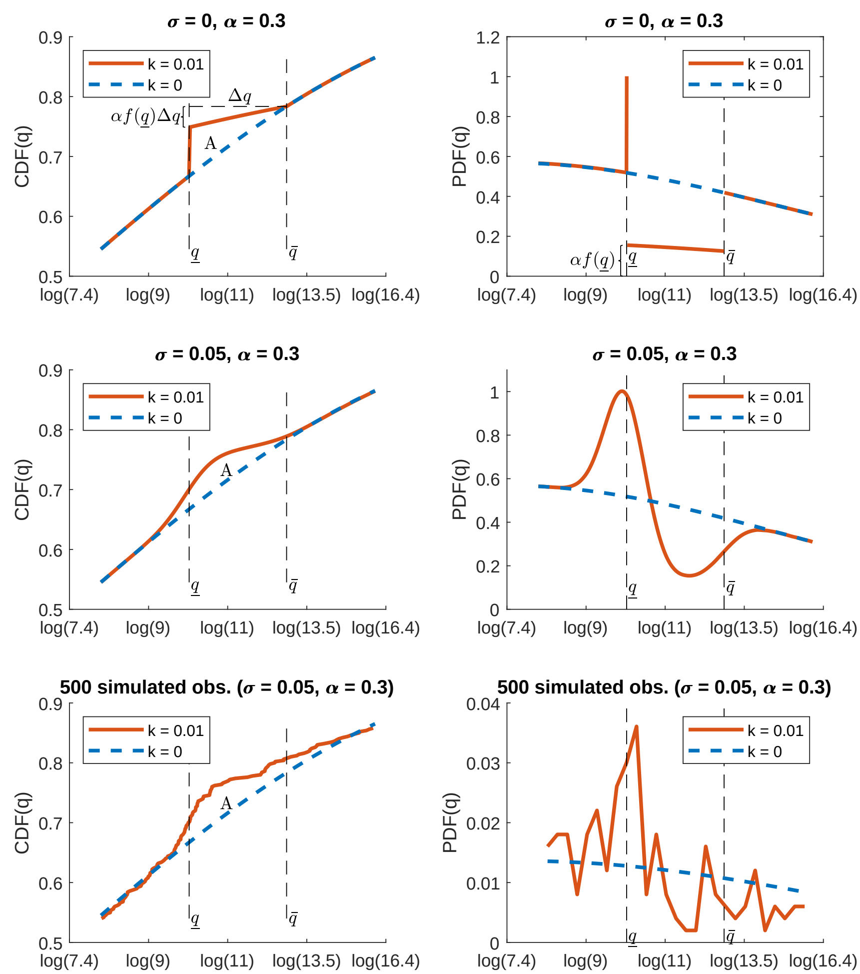
$$\begin{aligned} f(a) =& \frac{1}{\sigma}\phi\left(\frac{1}{\sigma}\left(a-\underline{q}\right)\right)\left(\Phi\left(\frac{\bar{q}-\mu_q}{\sigma_q}\right)-\Phi\left(\frac{\underline{q}-\mu_q}{\sigma_q}\right)\right) \\ &+ \frac{1}{\sqrt{2\pi\left(\sigma^2+\sigma_q^2\right)}} \exp\left(-\frac{1}{2}\frac{\left(a-\mu_q\right)^2}{\sigma^2+\sigma_q^2}\right)\left\{1-\Phi\left(\frac{\frac{1}{\sigma}\left(a-\underline{q}\right)-\sigma\frac{a-\mu_q}{\sigma^2+\sigma_q^2}}{\sqrt{\frac{\sigma_q^2}{\sigma^2+\sigma_q^2}}}\right)\right\} \\ &+ \frac{1}{\sqrt{2\pi\left(\sigma^2+\sigma_q^2\right)}} \exp\left(-\frac{1}{2}\frac{\left(a-\mu_q\right)^2}{\sigma^2+\sigma_q^2}\right) \Phi\left(\frac{\frac{1}{\sigma}\left(a-\bar{q}\right)-\sigma\frac{a-\mu_q}{\sigma^2+\sigma_q^2}}{\sqrt{\frac{\sigma_q^2}{\sigma^2+\sigma_q^2}}}\right). \end{aligned} \tag{B.28}$$

Taking log of equation (B.28), we can get the log-likelihood function for the MLE. Note that when $\tau = 0$, we have $\overline{q} = \underline{q}$. The first term of equation (B.28) equals 0 and this equation simplifies to:

$$f(a) = \frac{1}{\sqrt{2\pi\left(\sigma^2+\sigma_q^2\right)}} \exp\left(-\frac{1}{2}\frac{\left(a-\mu_q\right)^2}{\sigma^2+\sigma_q^2}\right).$$

Appendix C: Chapter 3

Figure C.1: Hypothetical Example: Banks Bunching with Non-optimizing Agents



Note: This figure shows a simulated example of asset size distribution when banks bunch at the \$10 billion regulatory threshold and a fraction $\alpha = 0.3$ of banks face optimization frictions. The data-generating process is described in Section 3.3. The left and right columns show the cumulative distribution function and the probability density function, respectively. The horizontal axis is the log bank assets. The vertical axis shows the probability densities and the cumulative probability densities on the left and right panels, respectively. The top panel shows the theoretical distributions with no noise. The middle panel shows the theoretical distributions with noise. The bottom panel shows the empirical distributions with 500 observations and noise. The probability density function on the bottom right panel is based on bin sizes of 0.025.

Table C.1: Fuzzy Bunching versus Sharp Bunching in Simulated Data with Non-optimizing Agents

	σ (%)	**q ~ N(2,0.7)**	**q ~ exp(λ=0.1)**
FB	0	1.010	1.102
		[0.117]	[0.131]
	2	1.053	1.155
		[0.178]	[0.230]
	4	1.093	1.183
		[0.192]	[0.190]
SB bin size: 0.005 ε: Auto	0	1.040	1.028
		[0.142]	[0.138]
	2	0.575	0.768
		[0.088]	[0.334]
	4	0.476	0.575
		[0.115]	[0.209]
SB bin size: 0.010 ε: Auto	0	1.022	1.031
		[0.136]	[0.132]
	2	0.602	0.803
		[0.083]	[0.338]
	4	0.529	0.660
		[0.079]	[0.276]
SB bin size: 0.005 ε: 0.010	0	1.015	1.002
		[0.140]	[0.134]
	2	0.421	0.455
		[0.073]	[0.080]
	4	0.460	0.496
		[0.077]	[0.083]
SB bin size: 0.005 ε: 0.020	0	0.987	1.001
		[0.139]	[0.137]
	2	0.437	0.478
		[0.057]	[0.066]
	4	0.444	0.476
		[0.070]	[0.080]

Note: This table presents the estimation results using simulated data. We run 500 simulations for each set of parameters. We set the bunching range of log-assets to 0.5. We set the fraction of non-optimizing agents to be 0.3. We report the average estimates over true value ($\widehat{\Delta q}/\Delta q$) and their standard deviations in parenthesis. We use a sample of 500 observations of bank assets. The standard deviation of noise σ is 0%, 2%, and 4%, respectively. The left and right columns draw the undistorted assets q from the normal distribution and the exponential distribution, respectively. For the sharp bunching approach, we use different combinations of spike intervals and bin sizes. In panels 2 and 3, the spike interval is determined based on the automatic procedure that sets the boundary of the spike interval as the points where the actual density function first crosses the counterfactual density function on each side of the threshold. In panels 4 and 5, the spike interval is fixed ex ante as $[\underline{q} - \epsilon, \underline{q} + \epsilon]$.

C.1 Proofs of Proposition 1

The bunching area A can be written as

$$A = \int (F(x) - F_0(x))\, dx = \int \int_{x-\overline{q}}^{x-\underline{q}} (G_0(\overline{q}) - G_0(x-u))\, h(u)\, du dx \tag{C.1}$$

We use the error bound theorem of the Trapezoid Rule, which states that for any function $y(x) \in C^2$, then

$$\int_a^b y(x)dx = \frac{y(a)+y(b)}{2}(b-a) + o((b-a)^2). \tag{C.2}$$

Since $G_0(.)$ and $h(.)$ are C^2, using the the error bound theorem, we have

$$\int_{x-\overline{q}}^{x-\underline{q}} (G_0(\overline{q}) - G_0(x-u))\, h(u) du \tag{C.3}$$

$$= \frac{1}{2}\left(G_0(\overline{q}) - G_0\left(\underline{q}\right)\right) h(x-\underline{q})\Delta q + o(\Delta q^2) \tag{C.4}$$

Hence:

$$A = \int \int_{x-\overline{q}}^{x-\underline{q}} (G_0(\overline{q}) - G_0(x-u))\, h(u)\, du dx \tag{C.5}$$

$$= \frac{1}{2}\left(G_0(\overline{q}) - G_0\left(\underline{q}\right)\right) \Delta q \int h\left(x-\underline{q}\right) dx + o(\Delta q^2) \tag{C.6}$$

$$= \frac{1}{2}\left(G_0(\overline{q}) - G_0\left(\underline{q}\right)\right) \Delta q + o(\Delta q^2). \tag{C.7}$$

Using the first-order Taylor expansion, $G_0(\overline{q}) = G_0\left(\underline{q}\right) + g_0 \Delta q + o(\Delta q)$, we have

$$A = \frac{1}{2}\left(G_0(\overline{q}) - G_0\left(\underline{q}\right)\right) \Delta q + o(\Delta q^2) \tag{C.8}$$

$$= \frac{1}{2} g_0(\underline{q})\Delta q^2 + o(\Delta q^2). \tag{C.9}$$

Note that when the noise converges to zero,

$$f_0\left(\underline{q}\right) = \int_{-\infty}^{+\infty} g_0\left(\underline{q} - u\right) h(u)du \rightarrow g_0\left(\underline{q}\right) \tag{C.10}$$

Therefore, we have

$$\Delta q_{FB} \equiv \sqrt{\frac{2A}{f_0\left(\underline{q}\right)}} = \sqrt{\frac{g_0(\underline{q})\Delta q^2 + o(\Delta q^2)}{g_0\left(\underline{q}\right)}} = \Delta q + o(\Delta q) \tag{C.11}$$

The above approximation becomes exact when the true bunching range Δq is sufficiently small:

$$\Delta q_{FB} \rightarrow \Delta q. \tag{C.12}$$

C.2 Derivation of the Sharp and Fuzzy Bunching Estimates in the Theoretical Example

For the sharp bunching approach, note that the total abnormal densities in the spike interval $[\underline{q} - \epsilon, \underline{q} + \epsilon]$ is given by

$$B(\epsilon) = \int_{\underline{q}-\epsilon}^{\underline{q}+\epsilon} \left(f\left(x\right) - f_0\left(x\right)\right) dx = f_0\Delta q - 2c(\sqrt{3}\sigma - \min\left\{\epsilon, \sqrt{3}\sigma\right\}) - \epsilon. \tag{C.13}$$

where $c = \frac{f_0\Delta q}{2\sqrt{3}\sigma}$ and $f_0 = 1$. Using the definition of the sharp bunching approach (bunching hole approach) in equation (3.8), we can get

$$\Delta q_{SB} = \Delta q - \Delta q \max\left\{1 - \frac{\epsilon}{\sqrt{3}\sigma}, 0\right\} - \epsilon. \tag{C.14}$$

For the fuzzy bunching approach, note that the cumulative distribution function is given by

$$F(x) = \begin{cases} F_0(x), & x \leq \underline{q} - \sqrt{3}\sigma, \\ F_0(x) + c(x - (\underline{q} - \sqrt{3}\sigma)), & \underline{q} - \sqrt{3}\sigma < x \leq \underline{q}, \\ F_0(\overline{q}) - c((\underline{q} + \sqrt{3}\sigma) - x), & \underline{q} < x \leq \underline{q} + \sqrt{3}\sigma, \\ F_0(\overline{q}), & \underline{q} + \sqrt{3}\sigma < x \leq \overline{q}, \\ F_0(x), & x \leq \overline{q}. \end{cases} \tag{C.15}$$

The bunching area is given by

$$\begin{aligned} A &= \int (F(x) - F_0(x))\, dx \\ &= \int_{\underline{q}-\sqrt{3}\sigma}^{\underline{q}} (x - (\underline{q} - \sqrt{3}\sigma))c dx + \int_{\underline{q}}^{\underline{q}+\sqrt{3}\sigma} (F_0(\overline{q}) - c\sqrt{3}\sigma + (x - \underline{q})c - F_0(x))dx \\ &+ \int_{\underline{q}+\sqrt{3}\sigma}^{\overline{q}} (F_0(\overline{q}) - F_0(x))dx \\ &= \int_{\underline{q}}^{\overline{q}} (F_0(\overline{q}) - F_0(x))dx + \int_{\underline{q}-\sqrt{3}\sigma}^{\underline{q}} (x - (\underline{q} - \sqrt{3}\sigma))c dx + \int_{\underline{q}}^{\underline{q}+\sqrt{3}\sigma} (-c\sqrt{3}\sigma + (a - \underline{q})c)dx \\ &= \frac{1}{2} f_0(\Delta q)^2 - \frac{1}{2}(\sqrt{3}\sigma - \sqrt{3}\sigma)^2 c \\ &= \frac{1}{2} f_0(\Delta q)^2. \end{aligned} \tag{C.16}$$

where the fourth equation uses $\int_{\underline{q}}^{\overline{q}} (F_0(\overline{q}) - F_0(x))dx = \frac{1}{2} f_0(\Delta q)^2$. Using the definition of the fuzzy bunching approach in equation (3.13), we can get

$$\Delta q_{FB} = \Delta q. \tag{C.17}$$

C.3 Derivation of the Fuzzy Bunching Approach with Optimization Frictions

The area between the two cumulative distribution function is given by:

$$\int \left(F\left(x\right) - F_0\left(x\right)\right) dx = \int \int_{x-\overline{q}}^{x-\underline{q}} (1-\alpha)\left(G\left(\overline{q}\right) - G\left(x-u\right)\right) h\left(u\right) du dx \tag{C.18}$$

Based on section C.1, we have

$$\Delta q = \sqrt{\frac{2\int \left(F\left(x\right) - F_0\left(x\right)\right) dx}{(1-\alpha) f_0\left(\underline{q}\right)}} \tag{C.19}$$

C.4 Derivation of the Fraction of Non-optimizing Agents

Because the noise has a mean of zero, half of the bunching mass is shifted into the dominated range. Therefore, we have the following equation:

$$\alpha = \frac{F\left(\overline{q}\right) - F\left(\underline{q}\right) - \frac{1}{2}(1-\alpha)\Delta q f_0(\underline{q})}{\Delta q f_0(\underline{q})} \tag{C.20}$$

where $F\left(\overline{q}\right) - F\left(\underline{q}\right)$ is the empirically observed mass in the dominated range, $\left[\underline{q}, \overline{q}\right]$, and $(1-\alpha)\Delta q f_0(\underline{q})$ is the bunching mass. We can solve α from this equation.