Watch What They Do, Not What They Say:
Estimating Regulatory Costs from Revealed Preferences*

October 1, 2019

Abstract

We propose a revealed preference approach to estimate the cost of banking regulation. Our methodology takes advantage of the fact that regulations usually do not apply until a regulated party reaches a certain size threshold. To avoid regulation, regulated parties around the threshold can downsize their assets to stay below the threshold if the regulatory cost outweigh the profits that they can earn from the additional assets. We can therefore infer the regulatory cost from the extent to which regulated parties downsize their assets to avoid regulation. Our estimator can be easily calculated using moments of the asset size distribution. Using simulation exercise, we show our estimator is robust to alternative data generating processes and exhibit superior performance compared to reduced-form estimators such as Difference-in-Differences and Regression Discontinuity design when endogeneous selection is severe and a valid instrument is absent. We apply our estimator to estimate the regulatory cost of the Dodd-Frank Act. Our estimates shows that the Dodd-Rank Act imposes a compliance cost of $1 million per year for banks above the $10 billion threshold. This estimate is significantly lower than self-reported estimates in surveys of banks.

JEL Classification Codes: G21, G28
Keywords: financial regulation, financial intermediation

*We thank seminar participants at the Swiss National Bank Workshop for helpful comments and discussions. This paper reflects the views of the authors and does not reflect the views of the IMF, its Executive Board, or IMF management.
1 Introduction

The aftermath of the 2008–09 financial crisis was one of the most active periods of financial regulation in U.S. history. Although the necessity of maintaining the integrity and safety of the financial system became a shared posterior after the near collapse of the U.S. financial system, considerable doubts have been subsequently raised about the costs imposed on financial institutions by these new regulations. Many argue that post-crisis financial regulation has created enormous compliance costs for financial institutions. For instance, the 2016 House Concurrent Budget Resolution estimates that “The estimated cost of Federal regulations are as high as $1.88 to $2.03 trillion per year” and “Dodd–Frank (Public Law 111–203) alone has resulted in more than $39.3 billion in regulatory compliance costs.” Such concern has prompted efforts to roll back some Obama-era rules. However, critics suggest that many studies which suggest high regulatory costs are conducted by interest groups which favor deregulation and many of their claims are unsubstantiated.

In the center of the current policy debate lies the challenging task of quantifying regulatory costs. Regulators generally do not have enough information to gauge the actual impact of regulation. Instead, they usually have to rely on self-reported estimates from regulated parties to guide their policies. However, regulated parties may have an incentive to inflate the compliance costs to justify seeking regulatory relief, which hinders the assessment of the regulatory outcomes (Crawford and Sobel (1982)).

In this paper, we use a revealed preference approach to circumvent the incentive problem inhibits the measurement of regulatory costs. Our methodology takes advantage of the fact

---

1In January 2017, the Trump administration issued an Executive Order that mandates “for every one new regulation issued, at least two prior regulations be identified for elimination.” In May 2018, Congress passed the Economic Growth, Regulatory Relief, and Consumer Protection Act, which scales back many banking regulations enacted after the crisis.

2For instance, Parker (2018) find that the $2 trillion regulatory costs cited in the 2016 House Concurrent Budget Resolution are derives from two studies funded by organizations having a strong financial or institutional stake in the outcome of their studies. The cost-estimation methodologies employed by these studies are also fundamentally flawed.
that regulations usually do not apply until banks reach a certain size threshold. To avoid regulation, banks around the threshold can downsize their assets to stay below the threshold if the compliance costs outweigh the profits that they can earn from the additional assets. We can therefore infer the private compliance cost from the extent to which banks downsize their assets to avoid regulation.

We start by constructing a simple model of banks’ choice of asset size in presence of size-based regulation. We derive a sufficient statistic of the regulatory cost using moments that can be easily calculated from the data. We provide simulation evidence that our sufficient statistic formula exhibits superior performance compared to reduced-form estimators such as Difference-in-Differences and Regression Discontinuity design when endogeneous selection is severe and a valid instrument is absent.

We then implement this methodology to estimate the cost of the Dodd-Frank Act for banks around the $10 billion threshold. The Dodd–Frank Act imposes tighter regulation on banks with assets greater than $10 billion. As shown in Figure 1, after the passage of the Dodd–Frank Act, numerous banks bunched below the $10 billion threshold to avoid regulation. Our formula shows that the magnitude of the bunching implies that the annual cost of regulation is around $1 million or 1.25% of the average net interest income. We compare our estimate with a full structural estimation and find the two estimates are quite similar. We also compare our estimate with surveys of banks. We find that our estimated regulatory costs are much lower than self-reported numbers by banks, even after considering indirect costs of regulation to consumers. This is consistent with the idea that survey-based estimates could be biased upward because of the incentive distortion problem.

Our approach has several advantages compared to the existing methodologies used to estimate regulatory costs. The first widely used methodology is conducting surveys of regulated parties. However, as argued above, regulated parties may have incentives to over estimate the regulatory costs to justify seeking regulatory relief. In contrast, our approach
infers the regulatory costs from the actions of regulated parties, thereby circumventing the incentive problem inherent in self-reported surveys. The second methodology is reduced-form regressions such as difference-in-differences (DID) and regression discontinuity design (RDD). The main challenge for these methods is the endogeneity problem when regulated parties can bunch below the regulatory threshold, which tends to bias estimates downward. Therefore, these methods are often limited to cases in which instruments are available to address the endogeneity problem. Our approach addresses the endogeneity problem by modeling the bunching behavior directly, which complements these reduced-form methodologies.

The third existing approach is to estimate full-fledged structural models. Compared to that approach, our methodology uses a sufficient statistics approach, which sidesteps the difficulty of specifying and estimating a full structural model.

This paper contributes to the literature that studies the costs and benefits of financial regulations. Although rigorous and data-driven regulatory intervention has enormous welfare implications for society, studies in this area remain considerably underdeveloped (Posner and Weyl (2013); Coates (2014); Cochrane (2014)). Coates (2014) reviews the rule-making process of six major financial regulations and argues that quantitative cost-benefit analyses by financial regulators are no more than “guesstimates.” Our paper offers an approach that can address the information asymmetry problem between regulators and regulated parties. Cochrane (2014) argues that the conventional methods of studying the costs and benefits of financial regulations are “not very good at and generally ignore the behavioral, market, general equilibrium, and political responses.” Our approach complements the conventional methods by explicitly modeling the endogenous responses of regulated parties. Our model is closely related to Garicano et al. (2016), who study the impact of size-contingent regulation on the labor market in France. Unlike Garicano et al. (2016), who estimate a fully specified structural model, we derive our formula under general assumptions of the distributions of underlying shocks. The main advantage of our approach is that it allows us to estimate the magnitude of regulatory costs without relying on a particular distribution. We see this result
in the spirit of the sufficient statistic approach introduced in the public finance literature by Chetty (2009). This sufficient statistic approach also bypasses the difficulty of a full structural estimation, making it appealing to a broader audience.

Second, this paper is related to the literature that uses bunching around points that feature discontinuities in incentives to elicit behavioral responses and estimate structural parameters. This approach has been widely used in public finance literature to estimate the labor supply elasticity to taxes (Saez (2010); Chetty et al. (2011); Kleven and Waseem (2013)). In the finance literature, DeFusco and Paciorek (2017) apply this methodology to study borrowers’ elasticity to interest rates in the mortgage market by exploiting the conforming loan threshold. Buchak et al. (2018) use the same bunching pattern to identify borrowers’ elasticity to mortgage size in a full demand system estimation. In the existing literature, the discontinuities in incentives is known (tax rates, subsidies for conforming loans) and the goal is to estimate the elasticity of response. In our case, the main goal of the estimation is to estimate the discontinuities in incentives (the regulatory costs). In addition, the existing literature usually estimates the bunching mass from the probability density function. In our setting, we derive our formula based on the cumulative distribution function because our data on banks are more sparse than the data on individual tax payers or mortgages.

2 Data and Institutional background

2.1 Data

We combine three main data sources: the Consolidated Reports of Condition and Income ("Call Reports", for commercial banks), the Consolidated Financial Statements for Holding Companies ("FRY-9C reports", for bank holding companies), and the Federal Deposit Insur-
ance Corporation Summary of Deposits ("FDIC SOD", for branches information). Since the Dodd–frank act applies to the highest holding bank\(^3\), we only consider the highest holding bank total consolidated asset for the model estimation of section 5.2 (e.g. if bank A is part of a highest holding bank B, we only keep bank B and its total consolidated asset). Moreover, we only keep banks between $3 and $50 billion in assets since our main focus is to study the cost of the $10 billion regulatory threshold. We provide in Table 1 the summary statistics for our sample. Over the two period, the sample covers 8,538 bank-quarters, with 157 banks per quarter on average, with a total of 408 unique banks. We notice that the average asset (conditional on being between 3 and 50 billion) is lower in the post Dodd–Frank period. The decrease can be attributed to the decrease in loans, and thus a general increase in capital ratio. Deposit and loan rates are lower in the post-DF period, with a small net increase in spreads from 4.5% to 4.6%.

### 2.2 The Dodd–Frank Act

The Dodd–Frank Wall Street Reform and Consumer Protection Act of 2010 (Dodd–Frank) is a centerpiece of the post-crisis financial reform. It introduces hundreds of new regulatory requirement, affecting a broad range of activities such as banking, derivatives trading, and housing finance to name a few. An exhaustive summary of the law can be found in Huntington (2010). The Dodd–Frank Act imposes more stringent regulatory scrutiny on banks based on asset size because larger banks are viewed to be more systematically important. For instance, banks whose assets exceed $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 of Dodd–Frank, and 4) create risk committees with independent directors. To avoid these regulations, banks around the

\(^3\)See article 12 CFR § 252.13
regulatory threshold start to bunch under $10 billion after the passage of Dodd–Frank Act as shown in Figure 1. This fact has also been documented in prior work (e.g., Morgan and Yang (2016) and Nicoletti et al. (2018)) and suggests that banks strategically avoid regulations.

2.3 Existing Methods Used to Estimate Regulatory Costs

In recent years, efforts to roll back the Dodd–Frank are getting up steam. Underlying this effort is the argument that overreaching regulations enacted by the Dodd–Frank Act impose high costs for the U.S. financial system. We provide in Table 2 a summary of studies that estimate the regulatory costs of the Dodd–Frank Act for banks. The existing quantitative studies employ two main methodologies. First, many studies use surveys of banks. For instance, McCormick (2017) estimates that the cost of crossing the $10 billion threshold is around $10 millions by surveying a group of bank CEOs. A report by the American Action Forum (Batkins and Goldbeck (2016)) estimates that implementing Dodd–Frank had cost $36 billion and 73 million paperwork hours as of July 2016. The Federal Register estimated the cost for all institutions at $10.4 billion, whereas the Government Accountability Office estimated the compliance cost for all banks to be $2.9 billion in the first 5 years. An important issue with surveys of banks is that banks may have incentives to over estimate the regulatory costs to justify seeking regulatory relief (Crawford and Sobel (1982)).

Another methodology used by existing studies is reduced-form regressions such as the Difference-in-Differences (DID) or Regression Discontinuity design (RD). For instance, Hinkes-Jones (2017) estimate regulatory costs by comparing legal and administrative expenses of affected banks with those of unaffected banks. More broadly, comparing banks across regulatory thresholds is a popular identification strategy. Balasubramanyan et al. (2019) study the impact of the Dodd-Frank Act on banks with over $10 billion of assets. This methodology faces a significant endogeneity challenge. As shown in Figure 1, banks endogenously choose...
To stay above the regulatory threshold or bunch below it. Therefore, unless there is an valid instrument, comparing banks above and below the regulatory threshold may lead to biased estimates of regulatory costs because of the endogeneity issue.

To illustrate the difficulty of using reduced-form regressions to identify the regulatory costs, we apply a Difference-in-Differences regression to estimate the regulatory costs imposed by the Dodd–Frank Act on banks with assets above $10 billion. The items we consider are: number of employees, total salaries, and expenses for legal, data processing, advisory, printing, stationery and supplies, auditing, and communications. We run a difference-in-difference regression covering the period 2005-2016 of the form:

\[
\text{Expenses}_{i,t} = \alpha + \beta_1 \text{Treat}_i + \beta_2 \text{Post}_t + \beta_3 \text{Treat}_i \times \text{Post}_t + \varepsilon_{i,t}
\]  

(1)

where \text{Treat} is a dummy equal to 1 if the bank total consolidated asset is between 10 and 14 billion as of 2010Q2, 0 otherwise, and \text{Post} is a dummy that takes 1 for all years after the Dodd–Frank act (after 2010). The regression coefficients of equation 1 are reported in Table 3. None of expense items increases significantly for the treated banks. This result is also confirmed by Figure 2 where we plot the average expenses for each bank group size, in which we find no significant increase in any expense measures for banks above $10 billion in assets.

This exercise highlights the challenges faced by regulators to obtain a rigorous quantitative assessment of regulatory costs. Regulators often have limited information on private compliance costs borne by regulated parties. Regulated parties have little incentive to reveal such costs because of a diverging interest. Finally, regulated parties may endogenously respond to regulation, which further complicates such analysis. To address these challenges, in the following section we present a structural approach. We show theoretically that banks’ endogenous responses reveal their private compliance cost. We derive a sufficient statistic
for regulatory costs based on moment that can be easily calculated from the data.

3 Model

We consider a very simple economy where banks provide a quantity of intermediation \( q \) at a price \( p \) to consumers. The quantity of intermediation \( q \) can be seen as the total assets of the bank, so we will refer as such in the rest of the paper; \( q \) is the log asset. Each bank faces a demand function \( q(p, z) \), where \( z \) is defined as the bank productivity. More productive banks (with high \( z \)) are able to provide a higher quantity of intermediation for the same price. Banks must pay a regulatory cost \( k \in \mathbb{R}_+ \) per year once its asset exceeds the threshold \( q \in \mathbb{R}_+ \). Denote \( \pi \) as the expected bank’s profit, the optimization problem can be written as follows:

\[
\pi = \max_p (p - c) \exp q(p, z) - k 1_{\{q(p, z) \geq q\}}
\]  

(2)

where \( c \) is the marginal cost of banks. This setup can also accommodate the case in which banks only pay the regulatory cost once by interpreting \( k \) as the amortized cost per period.

In the absence of regulation cost \((k = 0)\), the first order condition in (2) leads to bank’s undistorted price and quantity as

\[
p_0 = c + (-\partial_p q(p, z))^{-1}
\]

\[
q_0 = q(p_0, z)
\]
With regulation \((k > 0)\), banks whose log assets are above the regulatory threshold, \(q\), would incur a regulatory cost, \(k\), if they chose to maintain the original scale. Define \(z\) as the productivity that corresponds to the regulatory asset threshold.

\[ q = q(p_0, z) \]

Banks that are just above the regulatory threshold, \(q\), find it more profitable to downsize their assets to \(q\) and avoid the regulatory cost, \(k\). However, banks with much higher productivity will find such a strategy too costly and instead prefer to maintain their undistorted scale. Denote \(z\) as the productivity of the bank that is indifferent between downsizing and not downsizing. Denote \(q = q(p_0, z)\) as the undistorted log assets of the indifferent bank. We have

\[ k = (p_0 - c) \exp \bar{q} - (q^{-1}(q, z) - c) \exp q \tag{3} \]

where \(q^{-1}(q, z)\) is the inverse function of the demand. From this indifferent condition, one can express the regulatory cost as a function of the asset threshold \((q, \bar{q})\) from equation 3.

The top panel of Figure 3 plots the optimal log assets as a function of the undistorted assets before the regulation is in place. Banks whose undistorted log assets are below the regulatory threshold are not affected by the regulation \((q_0(z) \leq q)\). Banks whose undistorted log assets are just above the regulatory threshold choose to bunch at \(q\). Banks whose undistorted log assets are higher than the indifferent bank, \(\bar{q}\), choose to maintain their optimal scale.

To summarize, the optimal price and log asset with regulation becomes
\[ p^* = \begin{cases} 
  p_0 & \text{for } q_0 \in (-\infty, q] \\
  q^{-1}(q, z) & \text{for } q_0 \in [q, q] \\
  p_0 & \text{for } q_0 \in (q, \infty) 
\end{cases} \]

(4)

4 Empirical Strategy

We now explain how we apply our theoretical framework to the data. We first allow for some measurement error in the observed bank assets following Garicano et al. (2016). The observed log assets, \( a \), are measured with some error, \( u \).

\[ a = q^*(z) + u \]

This measurement error account for two important features in the data which is not captured by the theoretical model. First, in the data of Figure 1 the “bulge” in the distribution does not start precisely at the regulatory threshold, \( q \). Instead, it started slightly earlier. In contrast, the baseline model predicts that the “bulge” should start sharply at \( q \). Second, in the data of Figure 1, the region immediately to the right of the regulatory threshold is slightly upward slopping, indicating that there are still some observations just to the right of the regulatory threshold. However, the baseline model would rule out such possibility: it predicts a flat region to the right of the regulatory threshold. To capture these two features of the data, we allow for some measurement error in the observed assets.

Second, we parameterize the demand function faced by banks as a log linear form:

\[ q(p, z) = z - \beta p \]
where the parameter $\beta$ corresponds to the semi-elasticity of the demand function, pinning down the price sensitivity of consumer’s demand. With this simple demand function, the optimal price and log asset given regulation become

\[
p^* = \begin{cases} 
  c + \frac{1}{\beta} & , q^* = \begin{cases} 
    q_0 & q_0 \in (-\infty, q] \\
    q & \text{for } q_0 \in (q, \bar{q}] \\
    q_0 & q_0 \in (\bar{q}, \infty) 
  \end{cases} 
\end{cases}
\]

(5)

the regulatory cost can be expressed as

\[
k = \frac{1}{\beta} q [e^{\Delta q} - (1 + \Delta q)]
\]

(6)

where $\Delta q = \bar{q} - q$. This expression is directly obtained from 3 using the assumed log-linear demand. Note that we can express $\bar{z}$ as a function of $\bar{q}$ using the undistorted asset quantity for a bank with productivity $\bar{z}$. Indeed we have $q_0 \left( c + \frac{1}{\beta}, \bar{z} \right) = \bar{q} = \bar{z} - \beta c - 1$, which implies that $\bar{z} = 1 + \bar{q} + \beta c$. This is the reason why $\bar{z}$ does not appear in 6.

The demand elasticity parameter $\beta$ can be estimated by a standard logit regression in the IO literature as discussed in Appendix A. Given the demand elasticity parameter $\beta$ and the indifferent bank’s undistorted asset $q$, the regulatory cost $k$ is determined by the equation 3. In the following, we will discuss three different approaches to estimate the regulatory cost.

### 4.1 Estimating regulatory costs with sufficient statistic

The regulatory cost essentially depends on the undistorted log asset of the indifferent bank. If there is no measurement error, the marginal bank is easy to identify. As shown in the left
panel of 4, the indifferent bank is the end point of the flat region of the asset distribution. However, with measurement error, there is no clear flat region. In this section, we show how to estimate the undistorted log asset of the marginal bank, $q_i$, using several moments that can be easily constructed from the data.

Assuming that undistorted bank log asset $q_0$ and measurement error $u$ follow distributions with respective cumulative distribution functions $G(\cdot)$ and $H(\cdot)$, one can write the CDF of observed log asset $a$ as the convolution of $q_0$ and $u$:

$$F(a) = \mathbb{P}(x \leq a) = \int_u \mathbb{P}(x \leq a|u) \, dH(u)$$

where:

$$\mathbb{P}(x \leq a|u) = \mathbb{P}(q^* + u \leq a|u) = \mathbb{P}(q^* \leq a - u|u)$$

$$= \begin{cases} 
G(a - u) & a - u \in (-\infty, q] \\
G(q) & a - u \in (q, \bar{q}] \\
G(a - u) & a - u \in (\bar{q}, \infty) 
\end{cases}$$

Finally, we can rewrite the cumulative distribution function of $a$ as:

$$F(a) = \int_{-\infty}^{a-q} G(a - u) \, dH(u) + \int_{a-q}^{\bar{q}} G(q) \, dH(u) + \int_{\bar{q}}^{+\infty} G(a - u) \, dH(u)$$

Without regulation, the observed log asset is simply equal to the sum of the undistorted log asset and measurement error $q_0 + u$ for all values of $q$. The CDF is thus:

$$F_0(a) = \int_{-\infty}^{+\infty} G(a - u) \, dH(u)$$

\footnote{Recall that undistorted log asset is $q_0(z) = z - 1 - \beta c$, the distribution of productivity $z$ pins down the distribution of $q_0$.}
Having derived the CDF with and without regulatory cost, we can now make the following proposition:

**Proposition 1.** Assuming the cumulative distribution functions of $q_0$ and $u$ are continuously differentiable, $G(q_0), H(u) \in C^1$, one can approximate the bunching range $\Delta q$ as:

$$\Delta q \approx \sqrt{\frac{2 \cdot A}{f_0(q)}}$$  \hspace{1cm} (7)

where $A = \int_a (F(a) - F_0(a)) \, da$ is the area between the cumulative distribution function with and without regulation cost and $f_0(q)$ is the PDF of observed asset without regulation ($k = 0$), evaluated at $q$.

**Proof.** See Appendix B.1

The intuition behind this result is shown in Figure 4. When there is no measurement error (the left panel), $A$ is close to a triangle with an area of $\frac{1}{2} f(q) \Delta q^2$. Therefore, we can solve $\Delta q$ using the area $A$ with equation 7. It turns out that introducing measurement error does not have a first order effect on the area of $A$: the height of the “triangle” becomes smaller but the base becomes larger, and the area stays almost constant. Therefore, we can still use with equation 7 to calculate $\Delta q$ from $A$. Finally, plugging equation 7 into equation 3, one can rewrite the regulatory cost as follows:

$$k(A, f_0(q), \beta) = \frac{1}{\beta} e^q \left[ e^{\sqrt{2 \cdot A/f_0(q)}} - \left(1 + \sqrt{2 \cdot A/f_0(q)}\right)\right]$$ \hspace{1cm} (8)

which is an increasing function of the area $A$.

Equation 8 can be easily calculated from the data if we observe the asset size distribution both with and without regulation. In situations in which the asset size distribution without

\footnote{This approximation is more accurate when the CDF $G(q)$ has a small local curvature.}
regulation is not observable, we can fit a flexible polynomial to the observed distribution, excluding data in a range around the threshold \( q \), and extrapolate the fitted distribution to the threshold.

### 4.2 Estimating regulatory costs with maximum likelihood

As a comparison to the sufficient statistic approach, we also estimate the full structural model using a maximum likelihood estimator. To do so, we need to specify the distribution assumptions for each of the unobservable random variables. We assume that undistorted log asset \( q \) and measurement error \( u \) are both normally distributed\(^6\). The core parameters in the model are the average and standard deviation of undistorted log asset \( \mu_q, \sigma_q \), the standard deviation of the measurement error \( \sigma \), the regulatory fixed cost \( k \), and the demand semi-elasticity \( \beta \) that is separately estimated (see section A). The average measurement error is assumed to be 0 as in Garicano et al. (2016). One advantage of this methodology is that we can control for changes in the parameters before and after the regulation, which we do by letting \( \mu_q \) be different across the two periods. As a result, the pre-regulation period is used as counterfactual, whereas in the sufficient statistics we assumed a locally uniform counterfactual distribution and didn’t use pre-regulation data.

The probability of observing asset \( a_i \) lower than \( a \) can be shown to be (see appendix B.2 and B.3 for derivations):

\(^6\)Recall that without cost, undistorted log asset is given by \( q = z - 1 \) where \( z \) is interpreted as bank’s productivity. Hence, assuming that the undistorted log asset is normally distributed implies that productivity \( z \) is normally distributed.
\[ P(a_i \leq a) = \int_{\frac{1}{\sigma}(a-q)}^{\infty} \Phi \left( \frac{a - \sigma v - \mu q}{\sigma_q} \right) \phi(v) \, dv \]

\[ + \Phi \left( \frac{\bar{q} - \mu q}{\sigma_q} \right) \left\{ \Phi \left( \frac{1}{\sigma} (a - q) \right) - \Phi \left( \frac{1}{\sigma} (a - \bar{q}) \right) \right\} \]

\[ + \int_{-\infty}^{\frac{1}{\sigma}(a-q)} \Phi \left( \frac{a - \sigma v - \mu q}{\sigma_q} \right) \phi(v) \, dv \]

where \( \Phi(\cdot) \) and \( \phi(\cdot) \) denote the cumulative and probability density function of the standard normal distribution. Figure 4 shows how the parameters \( k \) and \( \sigma \) affect the theoretical cumulative distribution of asset. \( k \) drives the amplitude of the kink around threshold whereas \( \sigma \) drives the slope of the deviation.

We compute the probability density function by taking the derivative of \( P(a_i \leq a) \), which gives us the following:

\[ f(a) = \]

\[ \frac{1}{\sigma} \phi \left( \frac{1}{\sigma} (a - q) \right) \left( \Phi \left( \frac{\bar{q} - \mu q}{\sigma_q} \right) - \Phi \left( \frac{q - \mu q}{\sigma_q} \right) \right) \]

\[ + \frac{1}{\sigma_q} \sqrt{\frac{\sigma^2_q}{\sigma^2 + \sigma^2}} \exp \left( -\frac{1}{2} \frac{(a - \mu q)^2}{\sigma^2 + \sigma^2_q} \right) \left\{ 1 - \Phi \left( \frac{1}{\sigma} (a - q) - \frac{a - \mu q}{\sqrt{\frac{\sigma^2_q}{\sigma^2 + \sigma^2_q}}} \right) \right\} \]

\[ + \frac{1}{\sigma_q} \sqrt{\frac{\sigma^2_q}{\sigma^2 + \sigma^2}} \exp \left( -\frac{1}{2} \frac{(a - \mu q)^2}{\sigma^2 + \sigma^2_q} \right) \Phi \left( \frac{1}{\sigma} (a - q) - \frac{a - \mu q}{\sqrt{\frac{\sigma^2_q}{\sigma^2 + \sigma^2_q}}} \right) \]

Finally, the maximum log-likelihood problem becomes:

\[ \max_{\theta} \mathcal{L}(\theta) = \sum_{t=1}^{T} \sum_{i=1}^{J_t} \ln f(a_{it}; \mu_{q}^{pre}, \mu_{q}^{post}, \sigma_q, \sigma, k) \]  

(9)
where $\mu_q, \sigma_q$ are to the mean and standard deviation of productivity respectively, $\sigma$ is standard deviation of the measurement error, $k$ is the fixed cost of regulation, $T$ is the number of quarters, $J_t$ is the number of bank in each period, and $a_{it}$ is the total log asset of bank $i$ in quarter $t$. We allow for a different average productivity $\mu_q$ for the pre and post Dodd–Frank periods as mentioned before.

4.3 Estimating regulatory costs with reduced-form regressions

To compare with our sufficient statistic approach, we also derive two reduced-form estimators for the regulatory cost in our setting. The first one is a DID estimator:

\[-k = (E[\hat{\pi} | a \geq q, R] - E[\hat{\pi} | a \geq q, N]) - (E[\hat{\pi} | a < q, R] - E[\hat{\pi} | a < q, N])\]  \hspace{1cm} (10)

where $\hat{\pi} = \pi - \frac{1}{\beta}e^a$ is the residual profits after controlling for observable characteristics (in our case, the asset size).

\[
p^* = \begin{cases} 
\frac{1}{\beta} & q \in (-\infty, q] \\
\frac{1}{\beta} (1 + q - q) & q \in (q, q^* (q)] \\
\frac{1}{\beta} & q \in (q^* (q), \infty) 
\end{cases} \hspace{1cm} q^* (q) = \begin{cases} 
q & q \in (-\infty, q] \\
q & q \in (q, q^* (q)] \\
q & q \in (q^* (q), \infty) 
\end{cases}\]  \hspace{1cm} (11)

The second one is a RDD estimator

\[-k = \lim_{a \to q^+} E[\pi (a)] - \lim_{a \to q^-} E[\pi (a)]\]
where $\pi(a)$ is the profit of a bank with observed asset, $a$. These two estimators give unbiased estimates of the regulatory cost only if banks do not endogenously respond to the regulation. Instead, they choose the undistorted log assets even after the regulation is in place. In this case,

$$E [\hat{\pi} | a \geq q, R] = -k$$

$$E [\hat{\pi} | a \geq q, N] = E [\hat{\pi} | a < q, R] = E [\hat{\pi} | a < q, N] = 0$$

$$\lim_{a \to q^-} E [\pi | a] = \frac{1}{\beta} e^a$$

$$\lim_{a \to q^+} E [\pi | a] = \frac{1}{\beta} e^a - k$$

However, if banks endogenously respond to the regulation, these two estimators give biased estimates of the regulatory cost unless one finds a valid instrument.

### 4.4 Simulation

Before applying the closed-form expression in 8 to the real data, we check how it performs on a simulated dataset. To do so, we simulate $N = 5,000$ observed bank assets which corresponds approximately to the number of bank-quarter in the post Dodd–Frank period. To simulate observed log-asset $a = q^*(q, k) + u$, we first simulate $q$ from a normal distribution (such that assets are log-normally distributed) and an exponential-logarithmic distribution (such that the assets are exponentially distributed). These two distributions are widely used to model company sizes and allows us to test the sensitivity of the estimation to distribution
assumptions. With the simulated values $q$, we compute the optimal chosen asset $q^*(q, k)$ for three different regulatory costs $k = $1, $10, $50 millions and simulate the measurement error from a normal distribution with three different levels of standard deviation $\sigma = 1, 2.5, 5\%$. As a result we have 9 estimated regulatory costs for each estimation method that we can compare to the true. We compute bootstrapped standard error from the standard deviation of the estimate from 500 simulations. Finally, we calibrate the respective distribution parameters of $q$ to fit the empirical moment of observed assets in the data post Dodd–Frank (see Table 1) and set the demand semi-elasticity $\beta = 33$ from our separate estimation in Appendix A. The results are shown in Table 4.

The first panel corresponds to the Sufficient Statistic (SS) method. We first restrict to a small interval of asset size distribution which contains most of the excess mass, $[q_{\text{min}}, q_{\text{max}}]$. Computing the bunching mass in a smaller interval reduces noise in the regions far way from the regulatory threshold. After we choose the asset size interval, $[q_{\text{min}}, q_{\text{max}}]$, we then compute $A$ by numerically integrating the area between the cumulative distribution function with and without regulation cost. Then we compute the probability density at the threshold $f_0(q)$. The SS estimates perform very well across different parameter values. When error in measurement is very low ($\sigma = 1\%$), the SS estimates differ from true values by a maximum of 1%. When the regulatory cost is higher and the endogenous bunching is more salient, the SS estimator tends to generate more accurate estimates. This property makes the SS estimator a good complement to reduced-form estimators such as DID and RD because they usually perform worse when the endogenous bunching is more salient. In the right hand side of the table, we also use exponential distributions for the data generating process and find the SS estimator produces robust estimates. This is not surprising because the SS estimator is derived under general distribution assumptions, thus its performance will not be sensitive to the distributions of the data generating process.

For the normal distribution, we set $\mu_q = 2$ and $\sigma_q = 0.7$, and for the exponential distribution, we set $\lambda = 0.1$. 

---

7 For the normal distribution, we set $\mu_q = 2$ and $\sigma_q = 0.7$, and for the exponential distribution, we set $\lambda = 0.1$.
In the next panel we show the MLE estimator for the case of log-normal assets\(^8\). Since the likelihood is correctly specified and corresponds exactly to the data generating process, the MLE provides unbiased and efficient estimates of the regulatory cost. Nevertheless, the MLE estimator is subject to its usual criticism: it requires the econometrician to specify the exact distribution and takes considerable time to implement. Indeed, in the right hand side of the table, we apply the MLE estimator derived with a normal distribution to the simulated data using exponential distributions. We find that the performance of the MLE deteriorates, especially as the error in measurement volatility increases.

In the third panel, we present the results based on the DID estimator in section 4.3. The DID estimator substantially underestimate the regulatory cost. The performance of the DID estimator deteriorates when the regulatory cost is high because the endogenous bunching because more severe. Note that the performance of DID improves if we drop the observations near the regulatory threshold as shown in Table A2.

In the forth panel, we present the results based on the RD estimator in section 4.3. The RD estimator performs even worse than the DID estimator because its identification is entirely based on the local area around the regulatory threshold, which is significantly distorted by the endogenous bunching behavior.

To summarize, the simulation results show that the sufficient statistic estimator exhibits superior performance compared to DID and RD estimators when endogenous selection around the regulatory threshold is severe and there is no valid instruments. The sufficient statistic estimator is also robust to alternative assumptions of the data generating process. In the following section, we will implement our estimator in the real data.

\(^8\)This is the case for which the log-likelihood has an analytical solution (see section 9)
5 Estimation

We now illustrate our approach to estimate the regulatory costs imposed by the Dodd–Frank Act on banks whose assets exceed the $10 billion threshold. We first use the sufficient statistic approach. Then we estimate the full model using the maximum likelihood estimator. Finally, we consider an extension of the model which allows heterogeneous regulatory costs across banks.

5.1 Results using sufficient statistic

Table 5 presents the main estimate of the regulatory cost using the distribution of bank-quarter assets in the post Dodd–Frank period. To compute the sufficient statistic directly from the data, we first choose the interval \([q_{\text{min}}, q_{\text{max}}] = [2.2, 6.5]\) which contains most of the excess mass. This corresponds to the dollar value of the assets from $9 billion to $14.2 billion. Then, we find the empirical distribution of bank asset size within \([q_{\text{min}}, q_{\text{max}}]\) is quite close to a uniform distribution (zero degree polynomial) so we approximate the distribution of log-asset without regulation using a uniform distribution.\(^9\) Next, we can compute the area between the CDF of the asset size distribution with and without regulation, \(A = 1.05.\)

Next, we can compute the area between the CDF of the asset size distribution with and without regulation, \(f_0 (q) = 1/ (2.65 - 2.2).\) Finally, we set \(\beta = 33\) which is separately estimated in logit demand estimation shown in Appendix A.

The estimate for the regulatory cost is $1.45 million per year, which is considerably smaller than the survey estimates reported in Table 2. For a bank with $10 billion asset this cost represents around 1.6% of the return on assets each year. To assess if this result is statistically significant, we randomly select 500 samples of bank-quarter assets (with replacement) and re-compute the sufficient statistic. This gives us a standard error of 0.7 for

\(^9\)Alternatively, we can also use the pre-regulation asset size distribution directly. The result is quite similar.
the regulatory cost, which is significantly different from 0 at 5% confidence. We conducted a battery of robustness checks by changing the asset intervals as well as assuming alternative counterfactual distribution (normal, exponential). All results ranged between $0.6 million and $3 million.

5.2 Results using maximum likelihood

Results of the MLE estimation can be found in Table 6, the point estimate for the fixed regulatory cost \( k \) is $1.066 million, which is slightly lower to the sufficient statistic result. In other words, crossing the regulatory threshold yields a decrease of $1.066 million in annual profit. To put this number in perspective, the regulatory cost amounts to approximately 1.1% of average annual returns on assets.

Additionally, the model is able to generate a similar increase in the in the cumulative distribution function of bank-quarter with assets between $9 and $11 billion. Figure 5 compares the simulated CDF of bank-quarter assets (top) with the empirical CDF. We notice that the estimated parameters shown in Table 6 are able to generate a similar distortion in the distribution after the introduction of the Dodd–Frank act.

Since the regulatory cost is directly linked to the demand semi-elasticity, we present in Figure 6 the sensitivity of our cost estimate \( k \) with respect to \( \beta \). This figure is obtained by estimating the MLE each time with a different parameter \( \beta \). Unsurprisingly, the regulation cost decreases exponentially with \( \beta \).

5.3 Heterogeneous cost of regulation

So far we have proposed a model to estimate the average cost of the regulation. Here we generalize our methodology to the case in which the regulatory cost is a linear function of
certain bank attributes. To do so, we re-estimate the model by assuming the fixed cost \( k \) to be:

\[
k(x_i) = \gamma_0 + x_i'\gamma_1
\]

with \( x_i \) the vector of demeaned bank characteristics, \( \gamma_1 \) the cost sensitivity associated with these characteristics, and \( \gamma_0 \) the fixed cost for the bank with average characteristics. Since some characteristics could be themselves affected by the regulation, we populate \( x_i \) with values taken as of 2010 Q2, the last quarter before the implementation of the Dodd-Frank regulation. For banks that appear after the introduction of the Dodd-Frank act, we use the cross-sectional average characteristics of 2010 Q2, which does not add any new information.

We focus on two characteristics that are likely to drive the cost of regulation for a bank; the tier 1 capital ratio, and profitability. The tier 1 capital ratio is the ratio of bank’s core equity capital divided by risk-weighted assets. This ratio is directly reported by banks (item number 7204) and follows the guidelines set in the Basel Accords. Our hypothesis is that the regulatory cost is higher for a bank with low capital ratio (i.e. \( \gamma_1 < 0 \)). The rationale is that capital-constrained banks might not pass the Dodd–Frank stress test without costly adjustments. Our second characteristic, profitability, is measured as operating income divided by equity. Although not directly linked to a provision of Dodd–Frank, bank’s profitability is likely to affect the cost of implementing any regulation. Our hypothesis in this case is that highly profitable banks incur a lower cost of regulation (\( \gamma_1 < 0 \)), as profitable banks have potentially better staff and and better existing compliance processes to face the new regulatory framework.

Table 7 provides the results of the MLE estimation when the fixed cost \( k \) is a function of bank’s tier 1 capital ratio and profitability. The coefficients associated to banks productivity \((\mu_\zeta, \sigma_\zeta)\) and measurement error \((\sigma)\) are all equal to the homogeneous cost estimation of
table 6. Similarly, the constant regulation cost \( \gamma_0 \) corresponding to a bank with average capital and profitability is equal to $1.026 million per year. This is again almost identical to the fixed cost computed without heterogeneity ($1.066 million per year). Regarding the coefficients of interest \( \gamma_1 \) and \( \gamma_2 \), both coefficients are negative, confirming our hypothesis laid out in section 5.3. However, only the coefficient associated to the tier 1 capital ratio is significantly different from 0. In terms of economic significance, a one percentage point decrease in tier 1 capital ratio is associated with an increase in the fixed regulatory cost of $13,250 per year \((= −$1.325M \cdot 1\% \cdot 1,000,000)\). This strongly suggests that the Stress test provision in Dodd–Frank is a major cost driver for banks, consistent with bank CEOs interviews in McCormick (2017).

6 Conclusion

A data-driven evaluation of regulatory costs plays a crucial role for regulatory rule-making. However, such a task is often difficult to accomplish because of information asymmetry between regulators and regulated parties. In this paper, we propose a revealed preference approach to address this challenge. Our methodology takes advantage of the endogenous bunching behavior of regulated parties around regulatory thresholds. We show that the profits that regulated parties are willing to give up to avoid regulation reveal the private compliance cost. Our approach complements the existing methods in several dimensions. By focusing on the actions of regulated parties, our approach circumvents the incentive problem which may distort survey-based estimates. Moreover, by explicitly modeling the endogenous responses to regulation, our approach can apply to the situations in which reduced-form regressions are less applicable. Finally, by using a sufficient statistic approach, our approach sidesteps the difficulty of a full structural estimation. We illustrate our approach using the U.S. bank data. We find that the Dodd–Frank Act imposes an annual compliance cost of
$1 million for banks whose assets are above the $10 billion threshold. Our estimates are significantly lower than those reported in surveys.

References


Figure 1: Cumulative Distribution of Bank Assets

Note: The top panel presents the proportion of banks in asset bins of $2 billion increments, the bottom panel shows the cumulative distribution of bank-quarter assets between 9 and 11 billions, before and after the Dodd–Frank act. Pre-Dodd–Frank period ranges from 2000Q3 to 2007Q3, and Post-Dodd–Frank period ranges from 2010Q4 to 2017Q3.
Figure 2: Average Expenses over Time by Bank Size

Note: This figure presents the evolution of average cost-related variables for banks in different group size. Banks are classified as “Below” (“Above”) if their total consolidated assets is between $6 and $10 ($10 and $14) billion at the start of Dodd–Frank in 2010. Each item is taken from the end-of-year bank financial statements reported in the Call Reports (commercial banks) and FYR-9C forms (bank holding companies). Only highest holding banks (subject to regulation) are selected for the analysis.
Figure 3: The Effect of Regulation on Assets and Profit

**Note:** The top panel shows the level of assets implied by the model as a function of bank’s undistorted log-asset, when regulatory fixed cost is 0 (orange dashed line) or $0.01 billion (blue solid line). The bottom panel shows the loss due to regulatory cost as a function of bank’s undistorted log-asset.
Note: The plots exhibit the cumulative distribution function of log-assets for different regulatory fixed cost $k$ and measurement error volatility $\sigma$. The regulatory cost $k$ can be recovered from area $A$ and the slope of the CDF at $q$: $f(q)$. 
Figure 5: Simulated Cumulative Distribution of Bank Assets

Note: The top panel plots the cumulative distribution of bank-quarter simulated assets pre- and post-Dodd–Frank. The data are simulated using the parameters $(\mu_q, \sigma_q, \sigma, k)$ estimated from the bank-level data from 2000 - 2017. The bottom panel plots the actual cumulative distribution of bank-quarter assets.
Figure 6: Sensitivity of Regulation Cost to Elasticity of Demand

Note: The plot shows the value of the parameter $k$ in USD millions as a function of the semi-elasticity of demand $\beta$. 
Table 1: Summary statistics

<table>
<thead>
<tr>
<th></th>
<th>2001Q3-2007Q3</th>
<th></th>
<th>2010Q3-2017Q3</th>
<th></th>
<th>Whole sample</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std.</td>
<td>Mean</td>
<td>Std.</td>
<td>Mean</td>
<td>Std.</td>
</tr>
<tr>
<td>Assets</td>
<td>11,214</td>
<td>10,552</td>
<td>9,579</td>
<td>8,318</td>
<td>10,072</td>
<td>9,080</td>
</tr>
<tr>
<td>Cash</td>
<td>3.3</td>
<td>5.3</td>
<td>2.9</td>
<td>4.8</td>
<td>3.2</td>
<td>4.9</td>
</tr>
<tr>
<td>Securities for sale</td>
<td>18.9</td>
<td>18.3</td>
<td>16.2</td>
<td>17.9</td>
<td>17.3</td>
<td>18.2</td>
</tr>
<tr>
<td>Trading assets</td>
<td>0.5</td>
<td>1.6</td>
<td>0.3</td>
<td>1.2</td>
<td>0.4</td>
<td>1.4</td>
</tr>
<tr>
<td>Equity</td>
<td>8.9</td>
<td>8.8</td>
<td>11.1</td>
<td>9.6</td>
<td>10.1</td>
<td>9.3</td>
</tr>
<tr>
<td>Deposits</td>
<td>53.1</td>
<td>44.5</td>
<td>60.9</td>
<td>50.4</td>
<td>58.0</td>
<td>47.5</td>
</tr>
<tr>
<td>Loans</td>
<td>62.3</td>
<td>58.7</td>
<td>66.5</td>
<td>58.7</td>
<td>64.9</td>
<td>58.5</td>
</tr>
<tr>
<td>Operating income</td>
<td>1.8</td>
<td>2.1</td>
<td>1.7</td>
<td>2.1</td>
<td>1.7</td>
<td>2.2</td>
</tr>
<tr>
<td>Net income</td>
<td>1.2</td>
<td>1.3</td>
<td>0.9</td>
<td>1</td>
<td>0.9</td>
<td>1.2</td>
</tr>
<tr>
<td>Deposit rate</td>
<td>2.4</td>
<td>1.6</td>
<td>0.3</td>
<td>0.3</td>
<td>1.2</td>
<td>1.4</td>
</tr>
<tr>
<td>Loan rate</td>
<td>7.0</td>
<td>2.3</td>
<td>4.9</td>
<td>2.0</td>
<td>5.8</td>
<td>2.3</td>
</tr>
<tr>
<td>Spread</td>
<td>4.5</td>
<td>2.0</td>
<td>4.6</td>
<td>1.7</td>
<td>4.5</td>
<td>1.8</td>
</tr>
<tr>
<td># of employees</td>
<td>3,138</td>
<td>3,402</td>
<td>1,638</td>
<td>1,430</td>
<td>2,181</td>
<td>2,422</td>
</tr>
<tr>
<td># of branches</td>
<td>117</td>
<td>115</td>
<td>84</td>
<td>84</td>
<td>97</td>
<td>99</td>
</tr>
<tr>
<td># of banks per quarter</td>
<td>122</td>
<td>10</td>
<td>182</td>
<td>38</td>
<td>155</td>
<td>41</td>
</tr>
<tr>
<td># of bank-quarter</td>
<td>3,508</td>
<td>5,030</td>
<td>10,219</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of banks</td>
<td>206</td>
<td>307</td>
<td>421</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Summary statistics of highest holding banks in the sample. Assets are in $ million, and Cash, Securities for sales, Trading assets, Equity, Deposits, Loans, Operating income, and Net income are all expressed as a percentage of average asset. Both Operating Income and Net income are annual. The deposit rate, loan rate, and spread are computed using the Call Reports for commercial banks.
Table 2: Measuring the Impact of Dodd–Frank on Banks

<table>
<thead>
<tr>
<th>Source</th>
<th>Target of study</th>
<th>Methodology</th>
<th>Cost per bank per year ($ million)</th>
<th>Date</th>
<th>Main assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bank Director Magazine</td>
<td>Banks around $10B in assets</td>
<td>Quotes from approx. 10 CEOs</td>
<td>9.9</td>
<td>Jan-17</td>
<td>Average of the 4 numbers cited. Assumed net income of $50 million</td>
</tr>
<tr>
<td>American Action Forum</td>
<td>All institutions</td>
<td>Estimation from Federal Register</td>
<td>9.3</td>
<td>Jul-16</td>
<td>Total cost of $36 billion for 6 years. Cost per year of $6 billion. 648 highest holding banks corresponds to $9.3 billion per bank.</td>
</tr>
<tr>
<td>Federal Reserve Bank of Minneapolis</td>
<td>Banks below $1B in assets</td>
<td>Estimation of cost of new hires</td>
<td>1.1</td>
<td>Mar-13</td>
<td>Average decrease in ROA for a $1B bank under baseline and alternative scenario</td>
</tr>
<tr>
<td>Bloomberg BNA</td>
<td>All institutions</td>
<td>Estimation from Federal Register</td>
<td>16.0</td>
<td>Feb-17</td>
<td>Total cost of $10.4 billion. 648 highest holding banks corresponds to $16 billion per bank.</td>
</tr>
<tr>
<td>JPMorgan and Citigroup</td>
<td>JPMorgan and Citigroup</td>
<td>Direct quotes from the two banks</td>
<td>123.8</td>
<td>2012, 2014</td>
<td>Average of the two numbers cited by the banks</td>
</tr>
<tr>
<td>Bank Director and Grant Thornton LLP Survey</td>
<td>Banks above $500M in assets</td>
<td>Survey of 130 senior executives</td>
<td>Qualitative</td>
<td>Jun-13</td>
<td></td>
</tr>
<tr>
<td>KPMG 2013 Community Banking Survey</td>
<td>Banks between $1B and $20B in assets</td>
<td>Survey of 100 senior executives</td>
<td>Qualitative</td>
<td>Oct-13</td>
<td></td>
</tr>
<tr>
<td>Florida Chamber Fundation</td>
<td>Banks and credit unions with less than $5B in assets</td>
<td>Survey of 75 banks</td>
<td>Qualitative</td>
<td>Jul-12</td>
<td></td>
</tr>
<tr>
<td>Mercatus Center’s Small Bank Survey</td>
<td>Banks below $10B in assets</td>
<td>Survey of 200 banks</td>
<td>Qualitative</td>
<td>Feb-14</td>
<td></td>
</tr>
<tr>
<td>Risk Management Association Survey</td>
<td>Banks below $5B in assets</td>
<td>Survey of 230 executives</td>
<td>Qualitative</td>
<td>Mar-13</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** This table presents a non-exhaustive list of sources that have looked into the impact of Dodd–Frank. We selected sources that focused on the costs for banks. For sources with quantitative estimates, we translate the estimate into a total cost per bank per year. The assumptions behind this transformation are listed in the last column. The URL of each source can be found in the online appendix.
Table 3: Difference in Differences of Administrative Costs

<table>
<thead>
<tr>
<th></th>
<th>(1) # of employees</th>
<th>(2) Salaries</th>
<th>(3) Legal</th>
<th>(4) Data processing</th>
<th>(5) Advisory</th>
<th>(6) Printing</th>
<th>(7) Auditing</th>
<th>(8) Communications</th>
<th>(9) Total expenses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treat</td>
<td>1.138**</td>
<td>26.35**</td>
<td>0.382</td>
<td>0.576</td>
<td>0.238</td>
<td>0.279**</td>
<td>0.0685</td>
<td>0.254</td>
<td>28.15**</td>
</tr>
<tr>
<td></td>
<td>[0.444]</td>
<td>[12.83]</td>
<td>[0.268]</td>
<td>[1.109]</td>
<td>[0.331]</td>
<td>[0.131]</td>
<td>[0.0795]</td>
<td>[0.178]</td>
<td>[13.77]</td>
</tr>
<tr>
<td>Post</td>
<td>0.358*</td>
<td>12.93***</td>
<td>0.669***</td>
<td>0.897**</td>
<td>1.473***</td>
<td>0.165**</td>
<td>0.104**</td>
<td>0.0696</td>
<td>16.31***</td>
</tr>
<tr>
<td></td>
<td>[0.204]</td>
<td>[4.018]</td>
<td>[0.150]</td>
<td>[0.346]</td>
<td>[0.530]</td>
<td>[0.0625]</td>
<td>[0.0479]</td>
<td>[0.0602]</td>
<td>[4.812]</td>
</tr>
<tr>
<td>Treat * Post</td>
<td>-0.0492</td>
<td>4.334</td>
<td>0.505</td>
<td>0.115</td>
<td>-0.773</td>
<td>0.197</td>
<td>-0.202</td>
<td>0.256</td>
<td>4.433</td>
</tr>
<tr>
<td></td>
<td>[0.286]</td>
<td>[6.309]</td>
<td>[0.322]</td>
<td>[1.235]</td>
<td>[0.597]</td>
<td>[0.230]</td>
<td>[0.131]</td>
<td>[0.176]</td>
<td>[7.520]</td>
</tr>
<tr>
<td>Constant</td>
<td>1.524***</td>
<td>24.52***</td>
<td>0.286***</td>
<td>2.012***</td>
<td>0.265***</td>
<td>0.160***</td>
<td>0.0694**</td>
<td>0.215***</td>
<td>27.52***</td>
</tr>
<tr>
<td></td>
<td>[0.108]</td>
<td>[1.575]</td>
<td>[0.0564]</td>
<td>[0.313]</td>
<td>[0.0607]</td>
<td>[0.0451]</td>
<td>[0.0290]</td>
<td>[0.0507]</td>
<td>[1.730]</td>
</tr>
<tr>
<td>Observations</td>
<td>572</td>
<td>572</td>
<td>572</td>
<td>572</td>
<td>572</td>
<td>572</td>
<td>572</td>
<td>572</td>
<td>572</td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.127</td>
<td>0.144</td>
<td>0.095</td>
<td>0.025</td>
<td>0.046</td>
<td>0.068</td>
<td>0.007</td>
<td>0.070</td>
<td>0.150</td>
</tr>
</tbody>
</table>

Note: This table presents the results of a difference-in-difference regression for a sample of banks with assets between 6 and 14 billions, covering the period 2005-2016. All expenses are in USD millions, as of the end of the year. Number of employees in the first column is in thousands. Post is a dummy that takes 1 for all years after the Dodd-Frank act (after 2010). Treat is a dummy equal to 1 if the bank total consolidated asset is between 10 and 14 billion as of 2010, and 0 otherwise. Robust standard errors clustered at the bank-level are showed in parenthesis.
Table 4: Validity of Regulatory Cost Estimations

<table>
<thead>
<tr>
<th>k</th>
<th>$\sigma$</th>
<th>$\log(Q) \sim N(2,0.7)$</th>
<th>$\log(Q) \sim \exp(\lambda=0.1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0</td>
<td>1.2</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>(0.1)</td>
<td>(0.3)</td>
<td>(0.7)</td>
</tr>
<tr>
<td>10</td>
<td>10.1</td>
<td>11.4</td>
<td>11.1</td>
</tr>
<tr>
<td></td>
<td>(0.2)</td>
<td>(0.6)</td>
<td>(1.2)</td>
</tr>
<tr>
<td>50</td>
<td>51.0</td>
<td>57.6</td>
<td>58.5</td>
</tr>
<tr>
<td></td>
<td>(0.5)</td>
<td>(1.4)</td>
<td>(2.6)</td>
</tr>
<tr>
<td>1</td>
<td>1.0</td>
<td>1.0</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>(0.1)</td>
<td>(0.3)</td>
<td>(0.3)</td>
</tr>
<tr>
<td>10</td>
<td>10.0</td>
<td>10.0</td>
<td>9.8</td>
</tr>
<tr>
<td></td>
<td>(0.2)</td>
<td>(0.4)</td>
<td>(0.5)</td>
</tr>
<tr>
<td>50</td>
<td>50.1</td>
<td>50.2</td>
<td>49.9</td>
</tr>
<tr>
<td></td>
<td>(0.6)</td>
<td>(1.2)</td>
<td>(1.3)</td>
</tr>
<tr>
<td>1</td>
<td>1.1</td>
<td>1.3</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>(0.2)</td>
<td>(1.0)</td>
<td>(1.1)</td>
</tr>
<tr>
<td>10</td>
<td>8.5</td>
<td>10.3</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td>(0.2)</td>
<td>(1.1)</td>
<td>(1.2)</td>
</tr>
<tr>
<td>50</td>
<td>27.5</td>
<td>31.6</td>
<td>36.3</td>
</tr>
<tr>
<td></td>
<td>(0.7)</td>
<td>(1.4)</td>
<td>(1.3)</td>
</tr>
<tr>
<td>1</td>
<td>1.8</td>
<td>3.4</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>(0.2)</td>
<td>(1.2)</td>
<td>(1.3)</td>
</tr>
<tr>
<td>10</td>
<td>3.4</td>
<td>16.9</td>
<td>16.7</td>
</tr>
<tr>
<td></td>
<td>(0.1)</td>
<td>(0.7)</td>
<td>(0.9)</td>
</tr>
<tr>
<td>50</td>
<td>3.8</td>
<td>18.7</td>
<td>19.0</td>
</tr>
<tr>
<td></td>
<td>(0.1)</td>
<td>(0.6)</td>
<td>(0.6)</td>
</tr>
</tbody>
</table>

**Note:** Using simulated data, this table exhibits the estimated fixed regulatory cost $k$ using equation estimation methods exposed in section 4.4.
Table 5: Estimation of Regulatory Cost $k$ with Sufficient Statistic

<table>
<thead>
<tr>
<th>Stat.</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>1.45</td>
<td>0.70</td>
</tr>
<tr>
<td>Area</td>
<td>1.05</td>
<td>0.51</td>
</tr>
<tr>
<td>$q_{\text{min}}$</td>
<td>2.20</td>
<td></td>
</tr>
<tr>
<td>$q_{\text{max}}$</td>
<td>2.65</td>
<td></td>
</tr>
<tr>
<td>$f_0(q)$</td>
<td>2.25</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>33</td>
<td></td>
</tr>
</tbody>
</table>

Note: Using actual data, this table exhibits the estimated fixed regulatory cost $k$ using equation (3): $k = \frac{1}{\beta} e^{2 \left[ e^{\Delta q} - (1 + \Delta q) \right]}$, where the bunching range $\Delta q$ is approximated with the sufficient statistics of proposition 1: $\Delta q \approx \sqrt{2 \int_a^b (F(a) - F_0(a)) \, da / f_0(q)}$. We choose the bunching interval that is consistent with an annual regulatory cost of $10M per bank. This number is taken from anecdotal evidence from banks around the $10B threshold from McCormick (2017). Given this assumption, 95% of the bunching region will be between $q_{\text{min}} = 2.2$ and $q_{\text{max}} = 2.65$ ($Q_{\text{min}} = 9$ and $Q_{\text{max}} = 14.2$). Additionally, we assume the counterfactual distribution of log-asset $f_0(q)$ to be uniform in the bunching region such that $f_0(q) = 1/(2.65 - 2.2)$. Finally, we set $\beta = 33$ which is separately estimated in Appendix A.
Table 6: Parameters Estimates

<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>Estimated value</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Productivity (pre DF)</td>
<td>$\mu_{\zeta}^{preDF}$</td>
<td>3.097</td>
<td>0.012</td>
</tr>
<tr>
<td>Productivity (post DF)</td>
<td>$\mu_{\zeta}^{postDF}$</td>
<td>2.991</td>
<td>0.010</td>
</tr>
<tr>
<td>Productivity std.</td>
<td>$\sigma_{\zeta}$</td>
<td>0.716</td>
<td>0.005</td>
</tr>
<tr>
<td>Measurement error volatility</td>
<td>$\sigma$</td>
<td>0.040</td>
<td>0.004</td>
</tr>
<tr>
<td>Cost of regulation ($\text{million}$)</td>
<td>$k$</td>
<td>1.066</td>
<td>0.234</td>
</tr>
</tbody>
</table>

Note: Parameters from the MLE estimation. Standard errors are obtained via the inverse of the Hessian matrix.

Table 7: Parameters Estimates with Heterogeneous Cost of Regulation

<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>Estimated value</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Productivity (pre DF)</td>
<td>$\mu_{\zeta}^{preDF}$</td>
<td>3.097</td>
<td>0.012</td>
</tr>
<tr>
<td>Productivity (post DF)</td>
<td>$\mu_{\zeta}^{postDF}$</td>
<td>2.991</td>
<td>0.010</td>
</tr>
<tr>
<td>Productivity std.</td>
<td>$\sigma_{\zeta}$</td>
<td>0.716</td>
<td>0.005</td>
</tr>
<tr>
<td>Measurement error volatility</td>
<td>$\sigma$</td>
<td>0.039</td>
<td>0.004</td>
</tr>
<tr>
<td>Constant ($\text{million}$)</td>
<td>$\gamma_0$</td>
<td>1.026</td>
<td>0.280</td>
</tr>
<tr>
<td>Tier 1 capital ratio</td>
<td>$\gamma_1$</td>
<td>-1.325</td>
<td>0.280</td>
</tr>
<tr>
<td>Profitability</td>
<td>$\gamma_2$</td>
<td>-0.479</td>
<td>4.043</td>
</tr>
</tbody>
</table>

Note: Parameters from the MLE estimation. Standard errors are obtained via the inverse of the Hessian matrix. The model is estimated assuming that the fixed cost of regulation is given by $k_i = \gamma_0 + \gamma_1 \cdot (\text{Tier1CapitalRatio}_{i,2010Q2} - \text{Tier1CapitalRatio}_{i,2010Q2}) + \gamma_2 \cdot (\text{Profitability}_{i,2010Q2} - \text{Profitability}_{i,2010Q2})$. The tier 1 capital ratio is directly taken from the bank reports (item number 7204), and profitability is measured as operating income divided by equity.
## Appendix

### Table A1: Elasticity of Assets to Spread

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-DF</td>
<td>Post-DF</td>
<td>Whole sample</td>
<td>Without crisis</td>
</tr>
<tr>
<td>Spread</td>
<td>-22.33***</td>
<td>-32.75***</td>
<td>-31.69***</td>
<td>-32.33***</td>
</tr>
<tr>
<td></td>
<td>[3.460]</td>
<td>[4.302]</td>
<td>[3.794]</td>
<td>[4.227]</td>
</tr>
<tr>
<td># of employees per branch</td>
<td>0.488</td>
<td>1.845**</td>
<td>1.024</td>
<td>0.872</td>
</tr>
<tr>
<td></td>
<td>[0.353]</td>
<td>[0.767]</td>
<td>[0.708]</td>
<td>[0.662]</td>
</tr>
<tr>
<td>Concentration</td>
<td>-0.621***</td>
<td>-1.066***</td>
<td>-1.312***</td>
<td>-1.314***</td>
</tr>
<tr>
<td></td>
<td>[0.207]</td>
<td>[0.279]</td>
<td>[0.247]</td>
<td>[0.265]</td>
</tr>
<tr>
<td>Share of total # of branches</td>
<td>7.675</td>
<td>60.34**</td>
<td>2.189</td>
<td>1.389</td>
</tr>
<tr>
<td></td>
<td>[7.748]</td>
<td>[26.27]</td>
<td>[7.142]</td>
<td>[7.066]</td>
</tr>
<tr>
<td>Bank FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>6,651</td>
<td>5,833</td>
<td>12,484</td>
<td>11,050</td>
</tr>
</tbody>
</table>

**Note:** This table presents the results of the 2SLS regression of log(assets) on interest spread (instrumented with salaries per assets and expenses on premises per assets), number of employees per branch, concentration (sum of squared deposit market shares in each MSA), bank share of total number of branches, and bank and year fixed effects. The pre-DF period covers 2000 - 2009, the post-DF period covers 2010-2017, the whole sample covers 2000-2017, and the sample with crisis drops 2008 and 2009.
Table A2: DID without the bunching range

<table>
<thead>
<tr>
<th></th>
<th>( \sigma = 1 )</th>
<th>( \sigma = 2.5 )</th>
<th>( \sigma = 5 )</th>
<th>( \sigma = 1 )</th>
<th>( \sigma = 2.5 )</th>
<th>( \sigma = 5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>same</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>DID</td>
<td>(0.2)</td>
<td>(0.6)</td>
<td>(1.3)</td>
<td>(0.2)</td>
<td>(0.6)</td>
<td>(1.3)</td>
</tr>
<tr>
<td>DID</td>
<td>10.0</td>
<td>10.1</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.1</td>
</tr>
<tr>
<td>DID</td>
<td>(0.3)</td>
<td>(0.8)</td>
<td>(1.8)</td>
<td>(0.3)</td>
<td>(0.8)</td>
<td>(1.6)</td>
</tr>
<tr>
<td>DID</td>
<td>50.1</td>
<td>50.1</td>
<td>50.2</td>
<td>50.0</td>
<td>50.1</td>
<td>50.1</td>
</tr>
<tr>
<td>DID</td>
<td>(0.5)</td>
<td>(1.3)</td>
<td>(2.9)</td>
<td>(0.4)</td>
<td>(1.1)</td>
<td>(2.4)</td>
</tr>
</tbody>
</table>

Note: Using simulated data, this table exhibits the estimated fixed regulatory cost \( k \) using equation estimation methods exposed in section 4.4. We remove banks with assets in the bunching region (between \( q \) and \( \bar{q} \)).

A Estimation of demand semi-elasticity \( \beta \)

The elasticity of the demand is an important parameter in our model. It pins down the cost of regulation \( k \) when solving for the upper bound of productivity \( z \) above which banks decide to incur the fixed cost of regulation (and stop bunching towards the regulatory threshold \( \bar{A} \)).

Our objective is to estimate the parameter \( \beta \) in the following demand equation:

\[
\ln q_i = - \beta p_i + \ln z_i + \ln u_i
\]

where \( q_i \) is the total asset of bank \( i \), \( \zeta_i \) a bank-specific productivity, and \( u_i \) an error term (unobserved to the econometrician and banks). As a measure for the price of intermediation \( p_i \) charged by banks, we use the spread between loan and deposit interest rates. To control for the bank-specific unobserved productivity, we include a bank fixed effect in the regression, leveraging the panel dimension of the data. This allows us to control for unob-
served time-invariant bank characteristics correlated with the spread. There are however several unobserved time-varying bank characteristics that might correlate with the spread and impact the demand (e.g. quality of customer service, reputation of the bank, ...). To alleviate this concern, we follow the literature (e.g. Dick (2008), Ho and Ishii (2011) and Xiao (2018)) and use bank “cost shifters” as instrument variables for the spread. We use two main cost-shifters: total amount of salaries scaled by assets and expenses on premises scaled by assets. These two variables arguably do not affect the demand directly, but have an impact on the pricing of the bank. As shown in Figure 1 of the online appendix, the two instruments have a positive conditional correlation with bank spreads. Having defined the instrument, we proceed to estimate the following two-stage-least-square regression

\[
\ln q_{it} = \alpha_i + \alpha_t - \beta_{IV} \widehat{Spread}_{i,t} + x_{i,t}' \gamma + \varepsilon_{it} \tag{12}
\]

\[
Spread_{i,t} = \delta_i + \delta_t + \theta_1 Salaries_{i,t} + \theta_2 FixedExpenses_{i,t} + x_{i,t}' \delta + \nu_{i,t}
\]

where \(Spread_{i,t}\) corresponds to the interest rate earned on loans minus the interest rates paid on deposits, \(\alpha_i\) and \(\alpha_t\) are bank and year fixed effects, and \(x_{i,t}\) contains a set of control variables such as bank age, number of employees per branch, concentration (sum of squared deposit market shares in each Metropolitan Statistical Area (MSA)), and the bank’s share of the total number of branches in the United states.

Using the sample of commercial banks for the period from 2000 Q3 to 2017 Q3, we present the estimates of regression 12 in Table A1. Our coefficient of interest, the demand semi-elasticity \(\beta\), is highly significant and relatively stable across different periods, ranging from 22.33 in the pre-Dodd–Frank to 32.75 in the post-Dodd–Frank period. For our baseline estimation of the regulatory cost, we take the value from the whole sample excluding the crisis; \(\beta = 32.33\). This means that for a 1 percentage point increase in the spread, the quantity of assets decreases by approximately 32.3%. This number can also be informative about the average markup that banks charge their customers. From the assumed demand,
banks should set their markup equal to \( \left( \frac{\partial \ln q}{\partial p} \right)^{-1} = \beta^{-1} \approx 3.1\% = (1/32.3) \).

B Derivations and proofs

B.1 Proof of Proposition 1 and Corollary ??

Proposition 1

The area between the two CDF is given by:

\[
\int_a^b (F(a) - F_0(a)) \, da = \int_a^b \int_{a-q}^{a-q} (G(\bar{q}) - G(a-u)) \, dH(u) \, da
\]

Assuming \( H(u) \) is \( C^1 \),

\[
\int_{a-q}^{a-q} (G(\bar{q}) - G(a-u)) \, dH(u) \approx \int_{a-q}^{a-q} \left( G(\bar{q}) - \frac{1}{2} (G(\bar{q}) - G(q)) \right) \, dH(u)
\]

\[
= \int_{a-q}^{a-q} \frac{1}{2} (G(\bar{q}) - G(q)) \, dH(u)
\]

\[
= \frac{1}{2} \left( G(\bar{q}) - G(q) \right) (H(a-q) - H(a-q))
\]

\[
\approx \frac{1}{2} \left( G(\bar{q}) - G(q) \right) h \left( a - \frac{1}{2} (\bar{q} + q) \right) \Delta q
\]

Hence:

\[
\int_a^b (F(a) - F_0(a)) \, da \approx \frac{1}{2} \left( G(\bar{q}) - G(q) \right) \Delta q \int_a^b \left( a - \frac{1}{2} (\bar{q} + q) \right) \, da
\]

\[
= \frac{1}{2} \left( G(\bar{q}) - G(q) \right) \Delta q
\]
To calculate $\Delta q$, we need two empirical moments: 1) $\int_a (F(a) - F_0(a)) \, da$, the area between $F(a)$ and $F_0(a)$, 2) $f_0(q)$, the slope of $F_0(a)$ at $a = q$.

Note that

$$\int_a (F(a) - F_0(a)) \, da \approx \frac{1}{2} (G(q) - G(q)) \Delta q \approx \frac{1}{2} g(q) (\Delta q)^2$$

Note that

$$f_0(a) = \frac{dF_0(a)}{da} = \int_{-\infty}^{+\infty} \frac{dG(a-u)}{da} \, dH(u) = \int_{-\infty}^{+\infty} g(a-u) \, dH(u)$$

$$f_0(q) \approx g(q)$$

Therefore, we have

$$\Delta q = \sqrt{\frac{2 \int_a (F(a) - F_0(a)) \, da}{f_0(q)}}$$

### B.2 Cumulative distribution function of assets $\mathbb{P}(a_i \leq a)$

Result:

$$\mathbb{P}(a_i \leq a) = \int_{\frac{a - \bar{q}}{\sigma q}}^{\infty} \Phi \left( \frac{a - \mu q}{\sigma q} \right) \phi(v) \, dv$$

$$+ \Phi \left( \frac{\bar{q} - \mu q}{\sigma q} \right) \left\{ \Phi \left( \frac{1}{\sigma} (a - q) \right) - \Phi \left( \frac{1}{\sigma} (a - \bar{q}) \right) \right\}$$

$$+ \int_{-\infty}^{\frac{a - \bar{q}}{\sigma q}} \Phi \left( \frac{a - \sigma v - \mu q}{\sigma q} \right) \phi(v) \, dv$$

43
Derivation:

We have shown that the asset takes the following form:

\[
a_i = q(q_0, u) = \begin{cases} 
q_0 + \sigma v & q_0 \in (-\infty, \bar{q}] \\
q + \sigma v & q_0 \in (\bar{q}, \bar{\bar{q}}] \\
q_0 + \sigma v & q_0 \in (\bar{\bar{q}}, \infty)
\end{cases}
\]

where \( q_0 = z - \beta c - 1 \) is the undistorted quantity chosen by the bank in absence of regulation (with associated undistorted optimal price \( p = c + \beta^{-1} \)). For ease of notation we omit the subscript \( i \) in the derivation. Note that since \( c \) and \( \beta \) are constant across banks, the distribution of \( q_0 \) is pinned down by the distribution of unobserved bank productivity \( z \). So we denote \( q_0 \sim N(\mu_q, \sigma_q) \). The error in measurement is written as \( u = \sigma v \), where \( v \sim N(0, 1) \). Let us now compute \( \mathbb{P}(q(q_0, v) < a) \):

If \( a \) is small

\[
a < q + \sigma v,
\]

and hence:

\[
\mathbb{P}(q(q_0, v) < v) = \mathbb{P}(q_0 + \sigma v \leq a|v) \\
= \mathbb{P}(q_0 \leq -\sigma v|v) \\
= \Phi \left( \frac{a - \sigma v - \mu_q}{\sigma_q} \right)
\]

If \( a \) is in the middle region

\[
\bar{q} + \sigma v \leq a < \bar{\bar{q}} + \sigma v
\]

\[
\mathbb{P}(q(q_0, v) \leq a|v) = \Phi \left( \frac{\bar{q} - \mu_q}{\sigma_q} \right)
\]
If \( a \) is in high region

\[
a \geq q_0 + \sigma v
\]

\[
\mathbb{P}(q_0 + \sigma v < v) = \Phi\left(\frac{a - \sigma v - \mu_q}{\sigma_q}\right)
\]

Then the distribution can be obtained by

\[
\mathbb{P}(a_i \leq a) = \mathbb{E}\left[\mathbb{P}(q_0 + \sigma v \leq a | v)\right]
= \int_{\frac{1}{\sigma}(a - q)}^{\infty} \Phi\left(\frac{a - \sigma v - \mu_q}{\sigma_q}\right) \phi(v) \, dv
+ \Phi\left(\frac{\bar{q} - \mu_q}{\sigma_q}\right) \left\{ \Phi\left(\frac{1}{\sigma}(a - q)\right) - \Phi\left(\frac{1}{\sigma}(a - \bar{q})\right) \right\}
+ \int_{-\infty}^{\frac{1}{\sigma}(a - q)} \Phi\left(\frac{a - \sigma v - \mu_q}{\sigma_q}\right) \phi(v) \, dv
\]

**B.3 Probability distribution function of assets \( f(a) \)**

Result:

\[
f(a) =
\frac{1}{\sigma} \phi\left(\frac{1}{\sigma}(a - q)\right) \left( \Phi\left(\frac{\bar{q} - \mu_q}{\sigma_q}\right) - \Phi\left(\frac{q - \mu_q}{\sigma_q}\right) \right)
+ \frac{1}{\sigma_q} \sqrt{\frac{\sigma_q^2}{\sigma^2 + \sigma_q^2}} \exp\left(-\frac{1}{2} \frac{(a - \mu_q)^2}{\sigma^2 + \sigma_q^2}\right) \left\{ 1 - \Phi\left(\frac{1}{\sigma}(a - q) - \sigma \frac{a - \mu_q}{\sigma^2 + \sigma_q^2}\right) \right\}
+ \frac{1}{\sigma_q} \sqrt{\frac{\sigma_q^2}{\sigma^2 + \sigma_q^2}} \exp\left(-\frac{1}{2} \frac{(a - \mu_q)^2}{\sigma^2 + \sigma_q^2}\right) \Phi\left(\frac{1}{\sigma}(a - \bar{q}) - \sigma \frac{a - \mu_q}{\sigma^2 + \sigma_q^2}\right)
\]
Derivation:

\[ f(a) = \frac{d}{da} \mathbb{P}(a_i \leq a) = \frac{d}{da} \int_{-\infty}^{\infty} \mathbb{P}(g_0 + \sigma v \leq a|v) \phi(v) \, dv \]

\[ = \frac{d}{da} \int_{\frac{a}{\sigma} (a-q)}^{\infty} \Phi\left( \frac{a - \sigma v - \mu_q}{\sigma} \right) \phi(v) \, dv \]

\[ + \frac{d}{da} \left[ \Phi\left( \frac{\bar{q} - \mu_q}{\sigma_q} \right) \left\{ \Phi\left( \frac{1}{\sigma} (a - \bar{q}) \right) - \Phi\left( \frac{1}{\sigma} (a - \bar{q}) \right) \right\} \right] 

+ \frac{d}{da} \left[ \int_{-\infty}^{\frac{1}{\sigma} (a-q)} \Phi\left( \frac{a - \sigma v - \mu_q}{\sigma} \right) \phi(v) \, dv \right] v \]

The second term is

\[ \frac{d}{da} \left[ \Phi\left( \frac{\bar{q} - \mu_q}{\sigma_q} \right) \left\{ \Phi\left( \frac{1}{\sigma} (a - \bar{q}) \right) - \Phi\left( \frac{1}{\sigma} (a - \bar{q}) \right) \right\} \right] = \frac{1}{\sigma} \Phi\left( \frac{\bar{q} - \mu_q}{\sigma_q} \right) \{ \phi\left( \frac{1}{\sigma} (a - \bar{q}) \right) - \phi\left( \frac{1}{\sigma} (a - \bar{q}) \right) \}

For the first and third terms, we use the Libnitz Formula

\[ \frac{d}{da} \int_{h(a)}^{g(a)} f(a, u) \, du = f(a, g(a)) g'(a) - f(a, h(a)) h'(a) + \int_{h(a)}^{g(a)} f_a(a, u) \, du \]
The first term is

\[ \frac{d}{da} \int_{\frac{1}{\sigma}(a-q)}^{\infty} \Phi \left( \frac{a - \sigma v - \mu_q}{\sigma q} \right) \phi(v) \, dv \]

\[ = - \frac{1}{\sigma} \Phi \left( \frac{q - \mu_q}{\sigma q} \right) \frac{1}{\sigma} (a - q) + \frac{1}{\sigma q} \int_{\frac{1}{\sigma}(a-q)}^{\infty} \Phi \left( \frac{a - \sigma v - \mu_q}{\sigma q} \right) \phi(v) \, dv \]

\[ = - \frac{1}{\sigma} \Phi \left( \frac{q - \mu_q}{\sigma q} \right) \frac{1}{\sigma} (a - q) \]

\[ + \frac{1}{\sigma q} \sqrt{\frac{\sigma^2}{\sigma^2 + \sigma^2_q}} \exp\left( - \frac{1}{2} \frac{(a - \mu_q)^2}{\sigma^2 + \sigma^2_q} \right) \left\{ 1 - \Phi \left( \frac{\frac{1}{\sigma} (a - q) - \frac{\sigma - \mu_q}{\sigma^2 + \sigma^2_q}}{\sqrt{\frac{\sigma^2}{\sigma^2 + \sigma^2_q}}} \right) \right\} \]
since:

\[
\int_{\frac{a}{\sigma} (a-q)}^{\infty} \phi \left( \frac{a - \sigma v - \mu_q}{\sigma_q} \right) \phi(v) \, dv
\]

\[
= \frac{1}{2\pi} \int_{\frac{a}{\sigma} (a-q)}^{\infty} \exp \left( -\frac{1}{2} \left( \frac{a - \sigma v - \mu_q}{\sigma_q} \right)^2 - \frac{1}{2} v^2 \right) \, dv
\]

\[
= \frac{1}{2\pi} \int_{\frac{1}{\sigma} (a-q)}^{\infty} \exp \left( -\frac{1}{2} \frac{(\sigma^2 + \sigma_q^2) v^2 - 2(a - \mu_q) \sigma v + (a - \mu_q)^2}{\sigma^2} \right) \, dv
\]

\[
= \frac{1}{2\pi} \int_{\frac{1}{\sigma} (a-q)}^{\infty} \exp \left( -\frac{1}{2} \frac{(\sigma^2 + \sigma_q^2) (v - \frac{a-\mu_q}{\sigma^2 + \sigma_q^2} \sigma)^2 - (a-\mu_q)^2 \sigma^2 + (a - \mu_q)^2}{\sigma^2} \right) \, dv
\]

\[
= \exp \left( \frac{(a-\mu_q)^2}{(\sigma^2 + \sigma_q^2) \sigma^2} - (a - \mu_q)^2 \right) \frac{\sqrt{\sigma^2}}{\sqrt{2\pi}}
\]

\[
\cdot \frac{1}{\sqrt{2\pi}} \int_{\frac{1}{\sigma} (a-q)}^{\infty} \exp \left( -\frac{1}{2} \left( \frac{v - \frac{a-\mu_q}{\sigma^2 + \sigma_q^2} \sigma}{\sqrt{\frac{\sigma^2}{(\sigma^2 + \sigma_q^2)}}} \right)^2 \right) \, dv
\]

\[
= \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \frac{(a - \mu_q)^2}{\sigma^2 + \sigma_q^2} \right)
\]

\[
\cdot \frac{1}{\sqrt{2\pi}} \int_{\frac{1}{\sigma} (a-q)}^{\infty} \exp \left( -\frac{\sigma^2 + \sigma_q^2}{2\sigma_q^2} \left( v - \frac{a - \mu_q}{\sigma^2 + \sigma_q^2} \right)^2 \right) \, dv
\]

\[
= \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \frac{(a - \mu_q)^2}{\sigma^2 + \sigma_q^2} \right) \left\{ 1 - \Phi \left( \frac{1}{\frac{a-q}{\sigma} - \frac{\mu_q}{\sigma^2 + \sigma_q^2}} \sqrt{\frac{\sigma^2}{\sigma^2 + \sigma_q^2}} \right) \right\}
\]

The third term is computed similarly.