

# RISE | RICE INITIATIVE *for the* STUDY *of* ECONOMICS

## RISE Working Paper 15-016

“McMillan Palgrave Handbook of post crisis financial modeling, 2015

Analyzing Bank Efficiency: Are “too-big-to-fail” Banks Efficient?”

by Hulusi Inanoglu, Michael Jacobs, Jr., Junrong Liu and Robin Sickles



# RICE

Department of Economics

Baker Hall, MS22

6100 Main Street, Houston, Texas 77005

<https://economics.rice.edu>

**McMillan Palgrave *Handbook of post crisis financial  
modeling, 2015***

**Analyzing Bank Efficiency: Are “too-big-to-fail” Banks Efficient?**

Hulusi Inanoglu

Federal Reserve Board

Banking Supervision & Regulation

Quantitative Risk

Michael Jacobs, Jr.

Pricewaterhouse Coopers LLC

Financial Services Advisory

Junrong Liu

IFE Group

Robin Sickles

Rice University

Department of Economics

March 31, 2015

**Abstract** This paper analyzes the provision of banking services—the multi-output/multi-input technology that is utilized by banks in their role in the provision of banking services, including both balance-sheet financial intermediation businesses and off-balance-sheet activities. We focus on the largest financial institutions in the U. S. banking industry. We examine the extent to which scale efficiencies exist in this subset of banks in part to address the issue of whether or not there are economic justifications for the notion that these banks may be “too-big-to-fail.” Our empirical study is based on a newly developed set data based on Call Reports from the FDIC for the period 1994-2013. We contribute to the post-financial crisis "too-big-to-fail" debate concerning whether or not governments should bail-out large institutions under any circumstances, risking moral hazard, competitive imbalances and systemic risk. Restrictions on the size and scope of banks may mitigate these problems, but may do so at the cost of reducing banks' scale efficiencies and international competitiveness. Our study also utilizes a suite of econometric models and assesses the empirical results by looking at consensus among the findings from our various econometric treatments and models in order to provide a robust set of inferences on large scale banking performance and the extent to which scale economies have been exhausted by these large financial institutions. The analyses point to a number of conclusions. First, despite rapid growth over the last 20 years, the largest surviving banks in the U.S. have decreased their level of efficiency. Second, we find no measurable returns to scale across our host of models and econometric treatments and in fact find negative correlation between bank size and the efficiency with which the banks take advantage of their scale of operations. In addition to the broad policy implications of our analysis our paper also provides an array of econometric techniques, findings from which can be combined to provide a set of robust consensus-based conclusions that can be a valuable analytical tool for supervisors and others involved in the regulatory oversight of financial institutions.

**Keywords:** Banking productivity; panel data models; quantile regression, distance functions, economies of scale and scope

**JEL:** C14; C21; C23; G28

## 1. Introduction

The recent financial crisis has given rise to a reexamination by regulators and academics of the conventional wisdom regarding the implications of the spectacular growth of the financial sector of the economy. In the pre-crisis era, there was a widespread common wisdom that "bigger is better". The arguments underpinning this view ranged from potential economies of scale and scope, to a better competitive stance at the international level. However, in the post-crisis world the common wisdom has been altered somewhat as large banks have come to be viewed as problematic for policy makers and regulators, for various reasons. One reason often given is that economic agents who are insured have the incentive to take on too much ex ante risk, also known as the moral hazard problem. Second, there is the "too-big-to-fail" problem, the fear that large and interconnected financial institutions may become a source of systemic risk if allowed to go out of business, especially in a "disorderly" fashion (Bernanke, 2009). Support for or against large banking institutions turns on the central issue of whether or not efficiencies of scale and scope are economically and statistically significant and are positively associated with bank size. If they are positively associated with bank size then the expected benefits of the cost savings generated by increased efficiencies passed on to consumers in terms of better services or reduced banking service fees are traded off with the expected costs implicit in the moral hazard and systemic risk arguments. In this paper we attempt to shed some light on this question through an empirical analysis that investigates the relationship between measures of the efficiency of a bank's operation on the one hand, and the size of the institution on the other.

More recently, regulatory features added by the Dodd-Frank Act (DFA)<sup>1</sup> introduced a variety of new policy levers, including capital surcharges, resolution plan requirements, consideration of systemic risk effects in mergers which specifically increased the emphasis on understanding of economies of scale and scope in large financial firms. That is, DFA requires the review of whether a proposed merger would

---

<sup>1</sup> Public Law 111-203 Dodd-Frank Wall Street Reform and Consumer Protection Act

lead to greater concentrated risks to financial stability. Regulators have encouraged researchers to better understand the social utility of the largest, most complex financial firms (Tarullo, 2011).

Some elaboration on what we mean by "too-big-to-fail" (TBTF) banks is also in order. During times of financial crisis banking supervisors have strong incentives to forestall the failure of large and highly interconnected financial firms due to the damage that such an event could pose to both the financial sector as well as the real economy. Unfortunately, as market participants anticipate that a particular firm may be protected in this way, this has the perverse yet highly rational effect of undermining market discipline and encouraging excessive risk-taking by the firm. Furthermore, it establishes economically unjustified incentives for a bank to become larger in order to reap this benefit. This results in a competitive advantage for such a large bank over its smaller competitors who may be perceived as lacking this implicit government safety net. Public sector bailouts are costly and politically unpopular and this issue has emerged as an enormous problem in the wake of the recent crisis. Therefore, as a tactical matter the state of the financial system has left supervisors with little choice but to use government resources to avoid failures of major financial institutions and accompanying destabilization of the financial sector. However, on a prospective basis supervisors have been directed to better address this issue through improved monitoring of systemically critical firms, with a view to preventing excessive risk-taking, and by strengthening the resilience of the financial system in order to minimize the consequences of a large firm being unwound.

A series of reforms have been proposed to address these problems. They include increasing capital requirements and limits upon leverage (e.g., Basel III), capping the size of banks, limiting the scope of banking activities, subjecting bank mergers and acquisitions to additional scrutiny, prescribing that banks draft "living wills" to plan their orderly unwinding, and requiring the federal government to proactively break up selected banks. These measures are not without their detractors, however. Feldman

(2010), for example, casts doubt on the reforms focusing on size<sup>2</sup> by arguing even if such reforms could address TBTF, reforms that take aim at bank size directly might be bad policy because their costs could exceed their benefits. Moreover, the size of a bank may be positively related to other benefits. Large banks could offer cost advantages that would ultimately benefit society by taking advantage of scale economies in their service production processes. Wheelock and Wilson (2012), for example, concluded that most U.S. banks faced increasing returns to scale using their highly parameterized local linear estimator of banking services.

However, there may be problems with this perceived wisdom that large banks are large because of such scale economies for at least three reasons. First, some of the econometric work on economies of scale for banking, as in Hughes and Mester (1998), Hughes, Mester and Moon (2001), etc. find such benefits at all sizes of banks. Hughes and Mester (2008) summarize the extensive research findings in this regard. Second, we may simply not yet know very much about the presence of scale economies for today's unprecedentedly large banks. DeYoung (2010) emphasizes this point by arguing that the unique nature of today's large banks makes it difficult to apply statistical techniques to historical data to divine the extent of scale economies. It is clear that the financial sector has grown enormously in recent years. The question is why. Banks indeed contribute to economic output through intermediation and have performed this economically useful function in many countries for hundreds of years, but value-added intermediation does not necessarily justify a large banking sector or banks whose current size is enormous by any historical standards. There are reasons to think that this sector may have become too big in the sense that too many of society's resources are allocated to it and may continue to contribute to a distortion in rents paid to those employed in the financial sector. Perceptions by creditors of banks that the government will protect them can lead the sector to grow inefficiently large as TBTF guarantees attract excessive funding to banks. These creditors understand that

---

<sup>2</sup> Feldman argued that "...I am skeptical that reforms focused on size per se will achieve their stated purpose of addressing TBTF; I have more confidence in reforms that identify and address features that produce spillovers in the first place..."

their bank investments are implicitly subsidized by the assurance of government bailouts should the bank begin to fail. For example, Tracey and Davies (2012) argues that there exists an “implicit funding subsidy” for TBTF banks<sup>3</sup>. Another point about the limits of our knowledge concerning the scale economies of large banks is that analysts face real challenges in measuring the “output” produced by banks. Since the banking sector provides loans deposit and liquidity services it is a challenge to ensure that cross-firm comparisons are made controlling for these various service provisions, when economies of scale for the multi-output banking services technology is analyzed. Still another point is that the debate about TBTF and scale economies often presents the two in contradiction, when in fact they may complement one another. Some activities of a bank such may rely heavily on automation and thus may benefit from scale economies that enhance that bank’s TBTF status.<sup>4</sup> The average cost of the large investments on these automated systems could be driven down by the increasing in the volume of goods and services produced. Such automation-dependent products and services can generate a substantial portion banking revenues. Hence, greater scale activity could come with higher TBTF cost. The presence of economies of scale, from this perspective, suggests that policymakers sharpen their focus on fixing TBTF, see Feldman (2010).

The question of bank efficiency amongst the leading banking organizations in the US is important as the banks must too comply with the stress test and capital plan requirements outlined by the Federal Reserve’s Comprehensive Capital Analysis and Review (CCAR)<sup>5</sup>. For estimating the impact of given stress testing scenarios, large banks have been relying statistical models in order to quantify potential losses. The problem with this paradigm is that although it captures the social cost element it fails

---

<sup>3</sup> They conclude that scale economies appear to increase with bank size for large banks from a standard model of bank production that does not control for any TBTF funding cost advantage, while using an adjustment for the price of debt using the implicit funding subsidy they find evidence of constant returns to scale and possible scale diseconomies for large banks.

<sup>4</sup> Note that greater automation could imply greater operational risk, which is an implicit element of cost, but that is beyond the scope of the current empirical treatment.

<sup>5</sup> Also see BCBS 2009b.

to capture the potential social benefits of bank scale and scope economies, as banks generally cannot incorporate these potential gains into their risk models. Our research contributes to a balanced analysis of this by considering efficiency measures.

Our paper analyzes the provision of banking services—the multi-output/multi-input technology that is utilized by banks in their role in the provision of banking services, including both balance-sheet financial intermediation businesses and off-balance-sheet activities. We focus on large banks, in particular the largest 50 financial institutions in the U. S. banking industry. The combined total assets of the largest 50 U.S. banks is close to 80% of the total assets of U.S. banking system<sup>6</sup>. We examine the extent to which scale efficiencies exist in this subset of banks in part to address the issue of whether or not there are economic justifications for the notion that these banks may be “too-big-to-fail”. Our empirical study is based on a newly developed dataset based on Call Reports from the FDIC for the period 1994-2013. We contribute to the post-financial crisis "too-big-to-fail" debate concerning whether or not governments should bail-out large institutions under any circumstances, risking moral hazard, competitive imbalances and systemic risk. Restrictions on the size and scope of banks may mitigate these problems, but may do so at the cost of reducing banks' scale efficiencies and international competitiveness. Our study also utilizes a suite of econometric models and assesses the empirical results by looking at consensus among the findings from our various econometric treatments and models in order to provide a robust set of inferences on large scale banking performance and the extent to which scale economies have been exhausted by these large financial institutions. The analyses point to a number of conclusions. First, despite rapid growth over the last 20 years, the largest surviving banks in the U.S. have decreased in their level of efficiency. Second, we find no measurable returns to scale across our host of models and econometric treatments and in fact find negative correlation between bank size and the efficiency with which the banks take advantage of their scale of operations. In addition to the broad policy implications of our analysis our paper also provides an

---

<sup>6</sup> As of 3Q2013, the total assets of all U.S. Insured Commercial banks is \$13.5 trillion.



array of econometric techniques, findings from which can be combined to provide a set of robust consensus-based conclusions that can be a valuable analytical tool for supervisors and others involved in the regulatory oversight of financial institutions.

The preceding section has provided a short discussion addressing previous studies related to our work. Section 2 describes the econometric models that will be estimated. In section 3 we provide a description of our data-set. A discussion of our empirical findings is presented in section 4. Section 5 concludes.

## 2. Econometric Models

In this section, we review our estimating framework. We will estimate second order approximations in logs (translog) to a multi-output/multi-input distance function, (see Caves, Christensen and Diewert (1982) and Coelli and Perelman (1996)). The models we consider are linear in parameters. As our banking data constitute a balanced panel of banks and we are interested in a set of robust and consistent inferences from a wide variety of modeling approaches, we consider a number of different panel data estimators and assess the comparability of inferences from them. Our many treatments for various forms of unobserved heterogeneity can be motivated with the following classical model for a single output banking technology estimated with panel data assuming unobserved bank effects:

$$y_{it} = x_{it}\beta + \eta_i + u_{it} \quad i = 1, \dots, N; t = 1, \dots, T \quad (0.0)$$

Here  $y_{it}$  is the response variable (e.g. some measure of bank output),  $\eta_i$  represents a bank specific fixed effect,  $x_{it}$  is a vector of exogenous variables and  $u_{it}$  is the error term.

In the classical Fixed Effects (FE) model for panel data, individual unobserved effects  $\eta_i$  are assumed to be correlated with the regressors  $x_{it}$ , while in the classical Random Effects (RE) model individual unobserved effects  $\eta_i$  are assumed to be

uncorrelated with the regressors  $x_{it}$ . We also consider the Hausman and Taylor (H-T) (1981) panel estimator. The H-T estimator distinguishes between regressors that are uncorrelated with the individual effects ( $x_{it}^1$ ) and regressors that are correlated with the effects ( $x_{it}^2$ ). As we have no time-invariant regressors in our study, the model becomes:

$$y_{it} = x_{it}^1 \beta_1 + x_{it}^2 \beta_2 + \eta_i + u_{it} \quad i = 1, \dots, N; t = 1, \dots, T \quad (0.0)$$

We may interpret (0.0) or (0.0) as log-linear regressions, transformed from a Cobb-Douglas or translog function that is linear in parameters. In what follows, we do not distinguish between the  $x$ 's that are or are not allowed to be correlated with the effects in order to reduce notational complexity. We do, however, make clear what these variables are in the empirical section. In order to move from a single to the multi-output technology considered in our empirical work we specify the multi-output distance function in the following way. Let the  $m$  outputs be  $Y_{it} = \exp(y_{it})$  and the  $n$  inputs  $X_{it} = \exp(x_{it})$ . Then express the  $m$ -output,  $n$ -input deterministic distance function  $D_o(Y, X)$  as a Young index, described in Balk (2008):

$$D_o(Y, X) = \frac{\prod_{j=1}^m Y_{it}^{\gamma_j}}{\prod_{k=1}^n X_{it}^{\delta_k}} \leq 1 \quad (0.0)$$

The output-distance function  $D_o(Y, X)$  is non-decreasing, homogeneous, and convex in  $Y$  and non-increasing and quasi-convex in  $X$ . After taking logs and rearranging terms we have:

$$-y_{1,it} = \eta_i + \sum_{j=2}^m \gamma_j y_{jit}^* + \sum_{k=1}^n \delta_k x_{kit} + u_{it}, i = 1, \dots, N; t = 1, \dots, T \quad (0.0)$$

where  $y_{jit}^* = \ln(Y_{jit} / Y_{1it})$ . After redefining a few variables, the distance function can be written as

$$y = X\beta + Z\eta + u \quad (0.0)$$

Here  $y \in R^{NT}$  stacks the response variables across banks and time, the matrix  $Z = I_N \otimes i_T \in R^{NT \times N}$  distributes the bank specific fixed effects (or the "incidence matrix" that identifies N distinct entities in a sample) that are stacked in the vector  $\eta = (\eta_1, \eta_2, \dots, \eta_N) \in R^N$ , while  $X = [x_{NT \times n}, y_{NT \times (m-1)}^*]$  contains both exogenous and endogenous variables and  $U = (u_{it})^T \in R^{NT}$  is the stacked vector of error terms  $u_{it}$ .

However, the Cobb-Douglas specification of the distance function (Klein, 1953) has been criticized for its assumption of separability of outputs and inputs and for incorrect curvature as the production possibility frontier is convex instead of concave. On the other hand, as pointed out by Coelli (2000), the Cobb-Douglas remains a reasonable and parsimonious first-order local approximation to the true function<sup>7</sup>. We also consider the translog output distance function, where the second-order terms allow for greater flexibility, proper local curvature, and lift the assumed separability of outputs and inputs. If the translog technology is applied, the distance function takes the form:

$$\begin{aligned} -y_{lit} = & \eta_i + \sum_{j=2}^m \gamma_j y_{jit}^* + 1/2 \sum_{j=2}^m \sum_{l=2}^m \gamma_{jl} y_{jit}^* y_{lit}^* + \sum_{k=1}^n \delta_k x_{kit} + 1/2 \sum_{k=1}^n \sum_{p=1}^n \delta_{kp} x_{kit} x_{pit} \\ & + \sum_{j=2}^m \sum_{k=1}^n \theta_{jk} y_{jit}^* x_{kit} + u_{it}, \quad i = 1, \dots, N; t = 1, \dots, T \end{aligned} \quad (0.0)$$

This can be written in the form of Eq. (0.0). Here X contains the cross-product terms as well as the own n input m-1 normalized output terms.  $X = [x_{NT \times n}, y_{NT \times (m-1)}^*, xx_{NT \times (n \times (n+1)/2)}, y^* y_{NT \times ((m-1) \times m/2)}, xy_{NT \times (m-1) \times n}^*]$ , the latter of which appear in their normalized form owing to the homogeneity of the output distance function.

In the translog specification, our focus should be on the following key derivatives, which correspond to the input and output elasticities. The derivatives are expressed as

<sup>7</sup> Therefore, we estimate the distance function under both Cobb-Douglas and translog specifications. We will discuss only for the translog distance function, as those for the Cobb-Douglas are qualitatively comparable. These results are available on request.

follows in Eq. (0.0) and Eq. (0.0).

$$s_p = \delta_p + \sum_{k=1}^n \delta_{kp} x_k + \sum_{j=2}^m \theta_{pj} y_j^*, \quad p = 1, 2, \dots, n \quad (0.0)$$

$$r_j = \gamma_j + \sum_{l=2}^m \gamma_{jl} y_l^* + \sum_{k=1}^n \theta_{kj} x_k, \quad j = 2, \dots, m \quad (0.0)$$

## 2.1 Frontier Estimation Methodology

In this subsection, we describe our estimation methodology utilizing the semiparametric efficiency estimators summarized in Sickles (2005). We utilize Eq. (0.0) and consider cases in which  $\mathbf{u}$  and  $(\eta, x_1, x_2)$  are independent but there is a level of dependency among the effects and the regressors. Eq. (0.0) can be reinterpreted as a stochastic panel production frontier model introduced by Pitt and Lee (1981) and Schmidt and Sickles (1984). Although we may be on somewhat solid footing by invoking a central limit argument to justify a Gaussian assumption on the disturbance term  $u_{it}$ , we may be far less justified in making specific parametric assumptions concerning the distribution of the  $\eta_i$  term, which in the stochastic frontier efficiency literature is interpreted as a normalized radial shortfall in a bank's performance relative to the best-practice performance it could feasibly attain. While we can be confident in restricting the class of distributions of the inefficiency term to those that are one-sided (see the inequality in Eq. (0.0), the heterogeneity terms are intrinsically latent and unobservable components and we encounter problems regarding identifiability of these parameters (Ritter and Simar, 1997). The additional model we use in our analyses is a semi-parametric efficient (SPE) estimator and is well-suited to provide us with robust point estimates and minimum standard errors when we are unwilling to use parametric assumptions for the distribution of the heterogeneity terms and their dependency with either all or some of the regressors. The general approaches to deriving such semiparametric efficient estimators is discussed at length in Newey (1990) and Pagan and Ullah (1999), as well as in a series of papers by Park, Sickles and Simar (1998, 2003, 2007). Interested readers can find the derivations for

the SPE panel stochastic frontier estimators we utilize in our empirical work below in the cited papers. The framework for deriving all of the estimators is somewhat straightforward and has much in common across the different stochastic assumptions on which the different SPE estimators are based.

We utilize a particular SPE estimator in our analyses. This estimator is detailed in Park et al. (1998). We refer to this as the PSS1 estimator and it is an extension of the estimator introduced in Park and Simar (1994), which assumed that the effects were assumed to be independent of all of the regressors. We assume in the specification (0.0) that the set of regressors  $x_{1,it}$  is conditionally independent of the individual unobserved random effects  $\eta_i$  given the set of correlated regressors  $x_{2,it}$ :

$$f(\eta, x_1, x_2) = h(\eta, x_2)g(x_1 | x_2) \quad (0.0)$$

Furthermore, it is assumed that  $\eta_i$  depends on  $x_{2,it}$  only through its long-run movement:

$$h(\eta_i, x_{2,it}) = h_M(\eta_i, \bar{x}_{2,it})p(x_{2,it}) \quad (0.0)$$

Here  $h_M(\eta_i, \bar{x}_{2,it})$  is a nonparametric multivariate density specified using kernel smoothers. We will discuss our strategy for selection of the variables that are partitioned into  $x_{1,it}$  and  $x_{2,it}$ .

In addition to the PSS1 SPE estimator, we consider an alternative approach that allows for time-varying heterogeneity, interpreted in the stochastic frontier literature as a normalized level of technical efficiency. The approach is parametric. Battese and Coelli (1992), henceforth BC, consider a panel stochastic frontier production function with an exponential specification of time-varying firm effects:

$$\begin{aligned} Y_{it} &= f(X_{it}, \beta) \exp(u_{it} - \eta_{it}) \\ \eta_{it} &= \{\exp[-\zeta(t - T)]\} \eta_i \end{aligned} \quad (0.0)$$

where  $u_{it} \sim NID(0, \sigma_u^2)$  and  $\eta_i \sim NID^+(0, \sigma_v^2)$  are normal i.i.d. and non-negative truncated normal i.i.d., respectively. Maximum likelihood estimators of the model

parameters can be derived and mean technical efficiency can be constructed.<sup>8</sup>

## 2.2 Quantile Regression

A final class of estimator we consider in our empirical analyses of banking performance is the panel quantile regression model. The  $\tau^{th}$  conditional quantile function of the response  $y_{it}$ , the analog to Eq.(0.0), can be written as:

$$Q_y(\tau | Z, X) = X\beta(\tau) + Z\eta + u \quad (0.0)$$

Note that in model (0.0) the effects  $\beta(\tau)$  of the covariates X are allowed to depend upon the quantile  $\tau$ . The vector  $\eta$  is intended to capture individual specific sources of unobserved heterogeneity that are not adequately controlled for by other covariates. The estimates of the individual specific effects ( $\eta$ 's) are restricted to be invariant with respect to the quantile but are allowed to be correlated with the  $x$ 's as they are modeled as fixed effects. As pointed out in Galvao (2011), in settings in which the time series dimension is relatively large allowing quantile specific fixed effects is not feasible.

Koenker (1984) considered the case in which only the intercept parameter was permitted to depend upon the quantile and the slope parameters were constrained to be identical over selected quantiles. The slope parameters are estimated as regression L-statistics and the individual effects are estimated as discretely weighted L-statistics. The model we apply in this paper is the quantile regression fixed effects model for panel data developed in Koenker (2004), which solves the following convex minimization problem:

$$(\hat{\beta}, \hat{\eta})^T = \arg \min_{\beta, \eta} \left\{ \sum_{k=1}^K \sum_{i=1}^N \sum_{t=1}^T v_k \rho_{\tau}(y_{it} - x_{it}\beta(\tau_k) - \eta_i z_{it}) \right\} \quad (0.0)$$

where  $k$  indexes the  $K$  quantiles  $\{\tau_1, \tau_2, \dots, \tau_k\}$ ,  $\rho_{\tau}(u) \equiv u(\tau - I_{u < 0})$  is a piecewise linear quantile loss function as defined in Koenker and Bassett Jr (1978), and  $v_k$  are weights

---

<sup>8</sup> Alternatives to the BC specification of time varying heterogeneity, which has the same pattern but different intercepts for different firms, such as the Cornwell et al. (1990) estimator, required too much temporal variation in efficiency scores than the sample contained and we were unable to implement this estimator in our translog specification.

that control the influence of the quantiles on the parameter estimates. The choice of the latter are analogous to discreetly weighed L-statistics (Mosteller, 1946), a common choice of which is Tukey's trimean (Koenker, 1984).

### **3. Data**

The bank sample is from the top 50 U.S. banks by total book value of assets (TBVA), as of the third quarter of the year 2013, from quarterly Call Reports. More precisely, we have quarterly data from 1Q1994 to 3Q2013, obtained from the “Consolidated Reports of Condition and Income for a Bank with Domestic and Foreign Offices - FFIEC 031” regulatory reports, expressed on a pro-forma basis that go back in time to account for mergers. In order to illustrate, if a bank in 2008 is the result of a merger in 2008, pre-2008 data is merged on a pro-forma basis (i.e., the other non-surviving bank’s data will be represented as part of the surviving bank going back in time). The rationale behind this methodology is to create a long historical data-set that controls for survival bias, and also that does not exhibit a distorted measure of Banks' growth. U.S. bank regulators use this data in order to estimate risk measurement models, such as the Bank Charge-off at Risk Model (Frye and Peltz, 2008), which is the basis of risk dashboards used for centralized bank supervision. While this sample design is not a common practice amongst academics, this does reflect methodologies used by banks in calibrating credit risk models, such as those used for Basel III and for CCAR.<sup>9</sup>

Although we intended to analyze the top 50 U. S. commercial banks, due to missing and questionable data entry, we ended up using 44 of these banks in our analyses. The five output and six input variables used to estimate the distance function using both stochastic frontier analysis and quantile regression are:

---

<sup>9</sup> For more discussion of this issue the use of similar data in models for risk aggregation see Inanoglu and Jacobs (2009).

Real Estate Loans (REL)  
Commercial and Industrial Loans (CIL)  
Consumer Loans (CL)  
Securities (SC)  
Off-Balance-Sheet Activities (OFF)  
Premises & Fixed Assets (PFA)  
Number of Employees (NOE)  
Purchased Funds (PF)  
Savings Accounts (SA)  
Certificates of Deposit (CD)  
Demand Deposits (DD).

The risk proxies are:

CREDIT RISK: Gross Charge-off Ratio (CR)

LIQUIDITY RISK: Liquidity Ratio (LR)

MARKET RISK: Trading Revenue Deviation to Trading Book Ratio (MR)

Before further providing the descriptive statistics on our variables, we would like to draw the attention to our contribution to the banking efficiency in terms of a control variable, i.e. Market Risk (MR) proxy, which we have used in our analyses. Market risk results from holding or taking positions in interest rates, foreign exchange, equities, commodities, and credit spreads. While the core function of traditional banking is to accept deposits and make loans, large banks also take market risk on their trading books and make trading revenues. Loosely speaking, the banking book comprises lending activities, whereas the trading book comprises trading securities, over-the-counter (OTC) derivatives<sup>10</sup> and market making activities. Notwithstanding

---

<sup>10</sup> OTC derivatives are financial contracts which derive their values from underlying assets and market conditions. OTC derivatives create counterparty credit risk due to the risk of insolvency of one party before the settlement of the transactions. It is very difficult –if not impossible- to incorporate counterparty credit risk measures in an efficiency framework as counterparty credit risk measures are



the fact that, the 2007-08 financial crisis was initiated by a U.S. housing crisis, OTC derivatives which are mainly reported on banks' trading books contributed to amplifying various problems and provided channels for systemic risk to propagate (Gregory, 2014 p.3). The key differences between the trading and banking book relate to holding intent, liquidity and mark-to-market valuation. It has been evidenced that traditional banking business of accepting deposits and making loans has declined significantly in the U.S. (Allen and Santomero, 2001). The evidence continues to prevail in the ratio of the size of the trading book to total loans (i.e. traditional lending business) for top U.S. banks even after the 2007-08 financial crisis (Fig. 1).

Regulatory capital requirements for the banking and trading books differ significantly. As trading book positions are daily marked-to-market and actively hedged by the banks, they are not intended to be held for an extended period of time. Hence, the regulatory capital charges for such positions have been based on the price volatility. The first market risk regulatory capital requirements to recognize this fact were introduced in 1996 (Basel I Amendment). 1996 Amendment required banks to estimate a risk measure so called Value-at-Risk (VaR) for the trading book assets over a ten-day time horizon. However, during the 2007-08 financial crisis, losses in many banks' trading books have been significantly higher than the minimum capital requirements under the market risk rules (BCBS 2009a). Across global banks, trading book losses totaled over \$900 billion over 2007-2009 (Haldane 2009). The explanation was straightforward; when markets remain liquid and asset prices rose, banks gained from mark-to-market trading book valuations, but when asset prices fell during a financial crisis, market maker banks lost billion dollar losses on their trading books. This was clearly the case for major U.S. banks. Before the crisis, top five U.S. banks rarely reported quarterly trading losses but incurred multiple billion dollar losses during the crisis quarters (Fig. 2).

---

forward looking and constructed from “exposure profiles”. See Jacobs (2014) for regulatory requirements for counterparty credit risk measurement.

In response to the financial crisis, the Basel Committee on Banking Supervision (BCBS) introduced incremental changes to the current VaR based trading book framework in 2009 (also known as Basel 2.5, BCBS 2009a)<sup>11</sup>. The short term fix was to recognize the credit risk in the trading book with an incremental risk capital charge (IRC) for unsecured credit products, and comprehensive risk measure (CRM) for tranching credit products. Additionally, BCBS required banks to calculate a stressed VaR taking into account a one-year observation period relating to significant losses, which must be calculated in addition to the Value-at-Risk based on the most recent one-year observation period. The additional stressed value-at-risk requirement was incorporated to help reduce the procyclicality of the minimum capital requirements for market risk.

While these additional measures were meant to capture the real risk exposures of trading books, BCBS had agreed that the additional measures were not sufficient and planned to carry out a more fundamental review of the market risk framework, including the use of VaR estimates as the basis for the minimum capital requirement. The initial proposal<sup>12</sup> was released in May 2012 (BCBS 2012) which focused on key areas such as the trading book/banking book boundary, expected shortfall (ES) measure as an alternative to VaR, and a comprehensive incorporation of the risk of market illiquidity among other things. The importance of incorporating the risk of market illiquidity is a key consideration in banks' regulatory capital requirements for trading portfolios. Before the introduction of the Basel 2.5 changes, the entire market risk framework was based on an assumption that trading book risk positions were liquid, i.e. that banks could liquidate these positions over a 10-day horizon. The recent crisis proved this assumption to be false. That is, during the financial crisis, banks experienced significant illiquidity in a wide range of credit products held in the

---

<sup>11</sup> See Inanoglu, H., Jacobs, Jr., M., and A.K., Karagozoglu, (2014) for an impact analysis of Basel 2.5 on banks' regulatory capital for trading portfolios.

<sup>12</sup> A second consultative document was published in October 2013.

<http://www.bis.org/publ/bcbs265.htm>

trading book, hence, they were forced to retain exposures for prolonged periods of time.

Having stated the problems encountered for banks' trading portfolios during the crisis and the recent regulatory responses to the trading book related issues, we included a market risk proxy in our efficiency models in order to recognize the risk exposure of banks' trading books. To the best of our knowledge, this is the first paper which uses the variability of *unexpected* trading revenue as the market risk proxy in the banking efficiency literature. Ideally, one should use the quarterly Value at Risk (VaR)<sup>13</sup> which is an average of daily reported VaR's in a given quarter to proxy a bank's market risk exposure, however, as daily VaR's are not available to us, we follow Jorion (2002) who demonstrated that a bank's expected absolute value of "*unexpected* trading revenue" is proportional to the dispersion of Value at Risk (VaR) if the trading revenue is distributed symmetrically around zero. Following Jorion, we remove an estimate of the mean of the trading revenue (i.e. moving average of the last 4 quarters) in order to calculate the variability of trading revenue which is proxied as the absolute value of unexpected trading revenue. We then divide the absolute value of unexpected trading revenue by the gross sum of trading assets and trading liabilities to calculate the market risk proxy. That is;

$$MR = \frac{|\text{Deviation from the moving average of last 4 quarters of trading revenue}|}{(\text{Trading Assets} + \text{Trading Liabilities})}$$

Returning to descriptive statistics, Table 1 summarizes key variables as of 3Q2013, from the Call Reports for the top nationally chartered banks in the U.S. by *total book* value of assets (TBVA) at this time. We display details on the Top 10 out of 50 by TBVA in descending order (JP Morgan Chase, Bank of America, Wells Fargo,

---

<sup>13</sup> We note the deficiency of VaR measure especially after the crisis, but VaR is still the industry standard in measuring market risk.

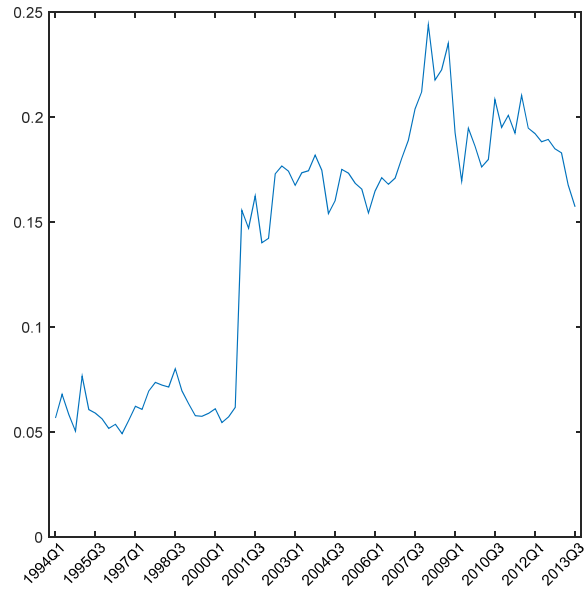
Citigroup, US Bank, Capital One, Bank of N.Y. - Mellon, PNC, State Street and HSBC) and distributional statistics on the Top 50. The data is extremely skewed in terms of size as measured by TBVA, with the average of the top 4 in TBVA each in excess of the 95th percentile of \$1.45 Trillion, and the Top 10 comprising \$8.14 Trillion (or 79.27%) out of the 10.27\$ Trillion total, as compared to median (average) TBVA of \$81.35 (\$233.47) Billion. There is similar extreme skew by the value of Total Loans (TL), with the average of the top 4 in TBVA in excess of the \$751.44 Billion 95th percentile of TL, and the Top 10 comprising \$3.81 Trillion or (74.08%) out of the \$5.14 Trillion total, as compared to median (average) TL of \$40.64 (\$116.80) Billion. We observe more extreme skew than even TBVA in the value of trading revenue deviation, with the average of top 4 in significant excess of the \$242.05 Million 95th percentile of trading revenue deviation, and the Top 10 comprising \$1.84 Billion or (90.12%) out of the \$2.04 Billion total, as compared to median (average) trading revenue deviation of \$46.33 Million (\$3.16 Million). Similarly, total gross charge-offs are skewed toward the largest banks, with the average of top 4 in TBVA each in excess of the \$1.81 Billion 95th percentile of gross charge-offs, and the Top 10 comprising \$9.93 Billion (or 83.99%) out of the \$11.82 Billion total, as compared to median (average) gross charge-offs of \$48.74 Million (\$268.69 Million). Finally for the dollar measures, total cash balances are very concentrated in the largest banks, with the average of top 4 in TBVA in excess of the \$244.53 Billion 95th percentile of total cash balances, and the Top 10 comprising \$ 1.66 Trillion (or 87.53%) out of the \$ 1.89 Trillion total, as compared to median (average) total cash balances of \$5.45 (\$42.99) Billion. Gross Charge-off ratios (CR) for many of the top 10 are on the high side relative to the center of the distribution, 8 of them above (ranging in 0.11%-0.67%) the median (average) in the broader sample of 0.12% (0.16%). There is a similar pattern with respect to liquidity ratios (LR), with many of the top 10 on the high side relative to the center of the distribution, 8 of them above (ranging in 10.19%-49.73%) the median (average) in the broader sample of 7.41% 13.04%.) Figures 3 through 7 represent several of these measures in time series on from the 1st quarter of 1994 until the 3th quarter of 2013.

Table 1: Characteristics of Top 50 Banks by Total Book Value of Assets as of 3Q2013  
(Call Report Data 1994Q1-2013Q3)

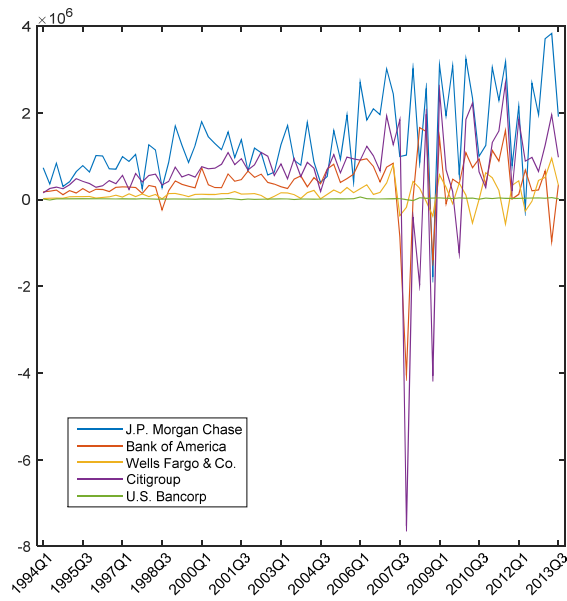
Bank	Book Value of Assets	Banking Book Loans	Gross Charge-offs	Cash Balances	Trading Revenue Deviation	Trading Book	Charge-off Ratio	Liquidity Ratio	Trading Revenue Deviation Ratio
Details on Top 10									
J.P. Morgan Chase	2,122,287,068	725,503,268	1,642,129	642,077,208	1,120,000	371,149,000	0.23%	30.25%	0.30%
Bank of America	1,621,455,000	925,878,000	2,205,000	165,301,000	298,395	113,693,000	0.24%	10.19%	0.26%
Wells Fargo & Co.	1,380,697,455	811,970,656	1,323,356	211,042,671	128,750	48,215,000	0.16%	15.29%	0.27%
Citigroup	1,346,413,607	603,523,317	2,594,637	322,674,064	217,909	186,016,138	0.43%	23.97%	0.12%
U.S. Bancorp	356,590,456	233,535,765	450,397	15,085,200	19,259	1,319,346	0.19%	4.23%	1.46%
Capital One Financial Corp	313,154,981	192,463,829	1,296,423	30,757,816	12,205	790,465	0.67%	9.82%	1.54%
Bank of New York Mellon	309,488,944	38,396,960	2,457	153,895,620	5,000	13,514,000	0.01%	49.73%	0.04%
PNC Financial	298,485,621	195,566,120	340,744	15,712,887	19,478	4,643,815	0.17%	5.26%	0.42%
State Street	212,689,010	15,636,947	1	47,486,430	12,903	10,728,408	0.00%	22.33%	0.12%
HSBC	181,762,250	64,552,314	73,301	51,422,421	3,159	37,506,127	0.11%	28.29%	0.01%
Statistics on Top 50									
Minimum	19,301,507	9,305,195	1	283,329	0	1	0.00%	1.38%	0.00%
5th Percentile	20,889,429	10,523,202	1	766,956	5	1,401	0.00%	1.63%	0.01%
25th Percentile	25,174,400	15,707,171	12,304	1,756,586	775	114,623	0.08%	4.29%	0.20%
Median	81,348,361	40,637,447	48,741	5,446,157	3,160	577,647	0.12%	7.41%	0.41%
Average	233,465,451	116,801,277	268,668	42,985,668	46,327	18,400,566	0.16%	13.04%	1.56%
75th Percentile	171,570,765	74,088,496	148,414	20,443,619	13,257	1,934,628	0.18%	16.77%	1.08%
95th Percentile	1,452,924,719	751,443,484	1,810,990	244,532,089	242,054	135,389,941	0.50%	47.13%	10.93%
Maximum	2,122,287,068	925,878,000	2,594,637	642,077,208	1,120,000	371,149,000	0.76%	49.73%	17.67%
Standard Deviation	461,032,705	217,045,601	592,341	112,345,609	175,235	63,699,072	0.16%	12.92%	3.52%
Skewness	10.4004	9.1999	9.5389	20.3433	33.6645	23.6458	8.7609	4.4204	14.4414
Kurtosis	2.9058	2.7271	2.7461	4.0299	5.4930	4.4868	2.2786	1.5191	3.4713
Grand Total	8,143,024,392	3,807,027,176	9,928,445	1,655,455,317	1,837,056	787,575,299	0.26%	20.33%	0.23%

Fig. 1 shows the ratio of the trading book to total loans across the US top 44 out of 50 banks from 1994Q1 to 2013Q3. This ratio fluctuates from 5% to 8% in the 1990's, and sharply surge up to 15% in early 2000's. It reaches the peak of around 25% in 2007 and drops to 17% in less than 2 years. The ratio continues decreasing in most recent years. Fig. 2 displays the trading revenue trend for the top 5 banks. These banks show similar fluctuations in time trend though some banks have greater variations than others do. These banks rarely experience negative trading revenues but incurred significant amount of dollar losses during the crisis quarters. Fig. 3 shows the TBVA across the U.S. 44 out of 50 largest banks over time, reflecting the growth in the banking industry overall as well as of the largest banks, with TBVA increasing smoothly from around just under \$4 Trillion in the early 1990's, to about \$10 Trillion during the recent financial crisis and declining about 1 Trillion until 2010 and then bouncing over \$10 Trillion recently. Fig. 4 shows the quarterly TL from over this

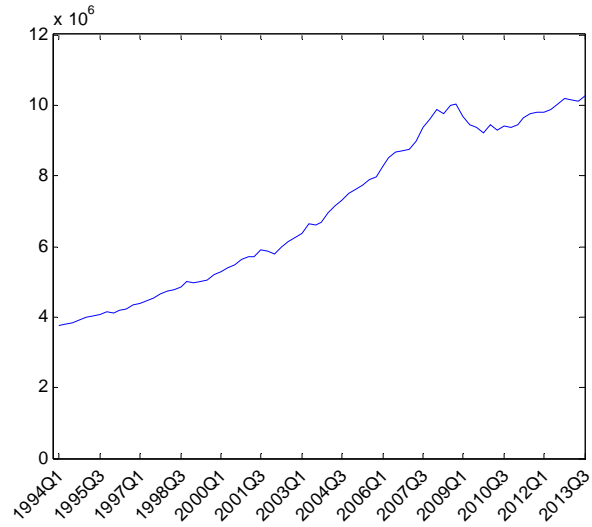
period, which shows a similar trend to TBVA, a secular upward trend of growth (from about \$2.5 to nearly \$5.5 Trillion in 2008), as well the financial crisis, reflected dips of about \$1 Trillion in the period 2008 to 2009, and increased slowly since then. In Fig. 5, the time series of CRs clearly reflects the credit cycle, with previous peaks of 0.4% around early 2000s, and alarmingly near 1% by the end of 2009. On the other hand, in Fig. 6, LRs display a markedly different pattern over time as compared to CRs, a secular decline from around 10% at the beginning of the sample period to around 6% from 1997 to 2001, and reaching up to about 16% after 2007 and fluctuating since then to 17% the end of the sample period. Finally, in Fig. 7, we see the ratio of deviation of trading revenue from the moving average of the previous 4 quarters to the trading book displaying yet another different pattern to the other risk measures; it shows one mode around year 2000 and another peak at the year 2007 and sharp decline since then). In Fig. 8 through Fig. 12 we show the distributions of the 5 measures analyzed in Table 1 across the largest banks as of 3Q2013. The right skewness in all of these variables is evident.



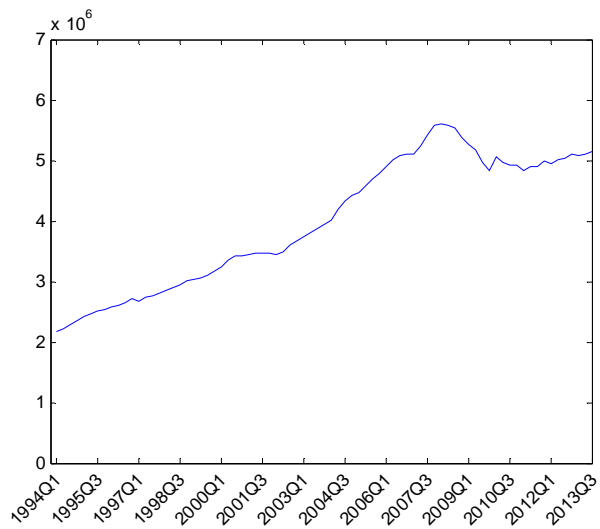
**Fig. 1:** Ratio of Trading Book to Total Loans as of 2013Q3



**Fig. 2:** Trading Book Revenue for Top 5 US Banks as of 2013Q3 (in Thousand \$)

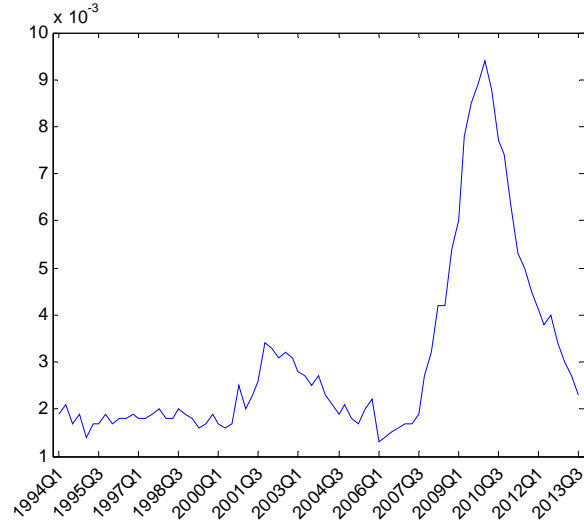


**Fig. 3:** Total Book Value of Assets (in Million \$)

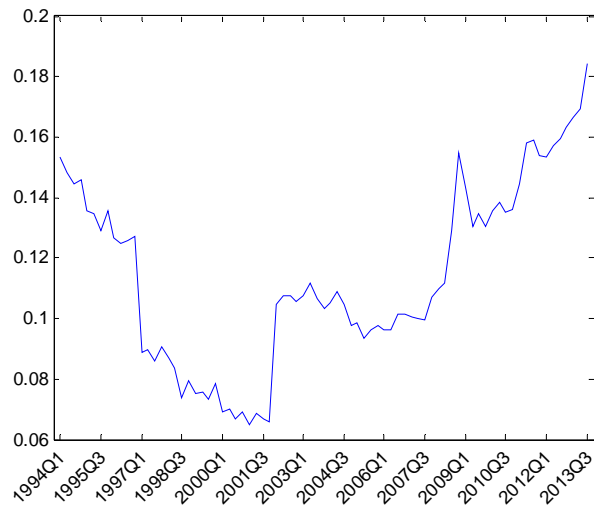


**Fig. 4:** Total Loans (in Million \$)

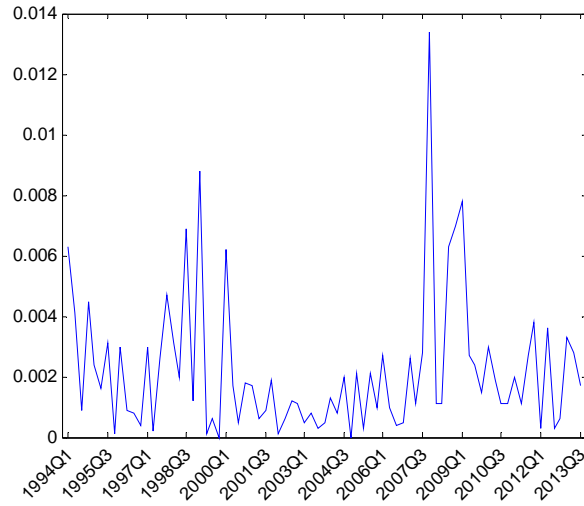




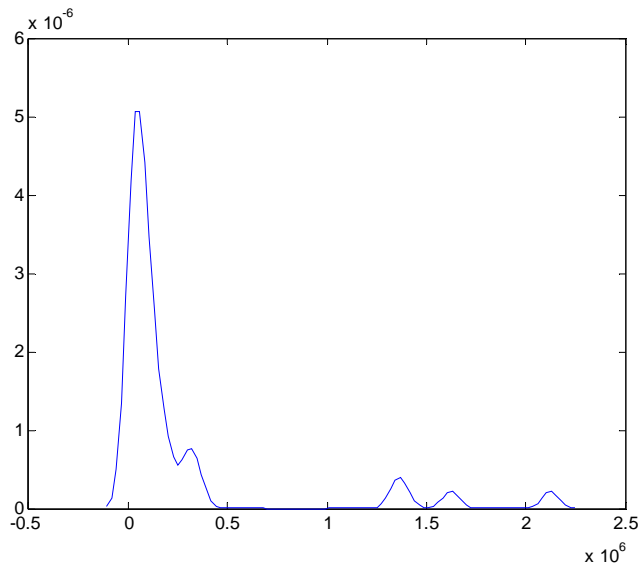
**Fig. 5:** Average Ratio of Total Charge-off to Total Loans



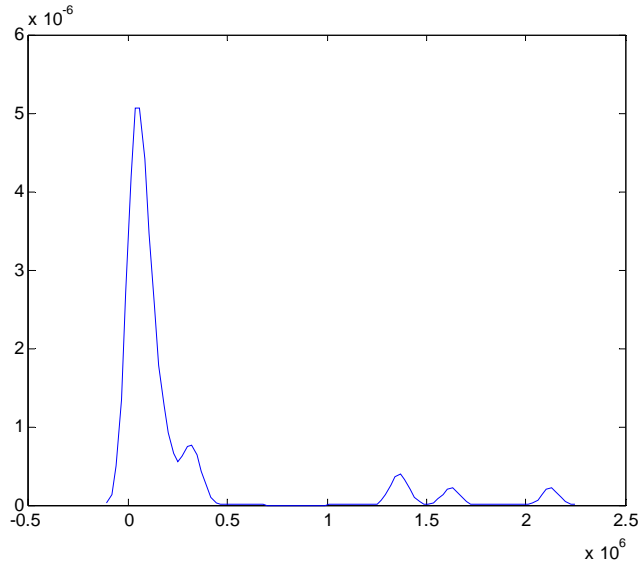
**Fig. 6:** Average Liquidity Ratios



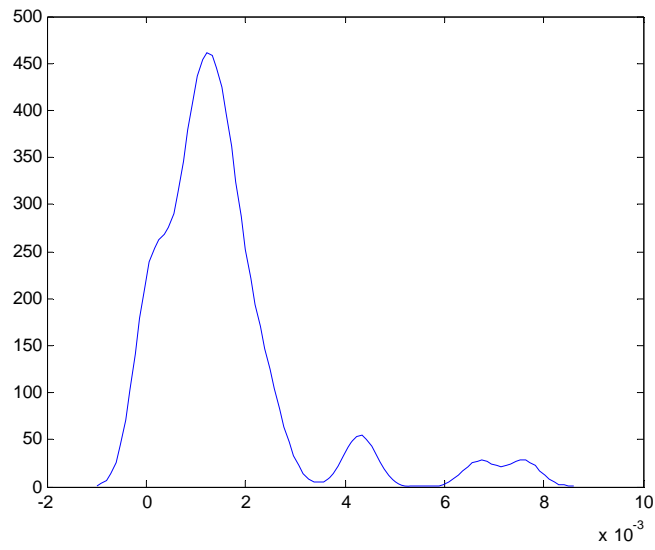
**Fig. 7:** Average Trading Revenue Deviation to Trading Book



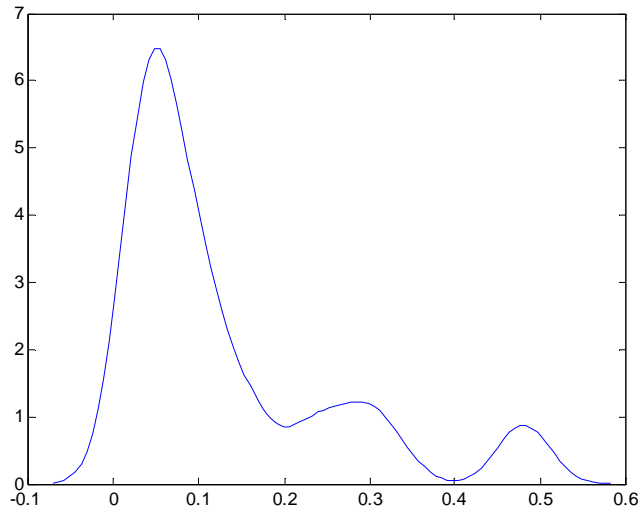
**Fig. 8:** Distribution of Total Book Value of Assets as of 2013Q3



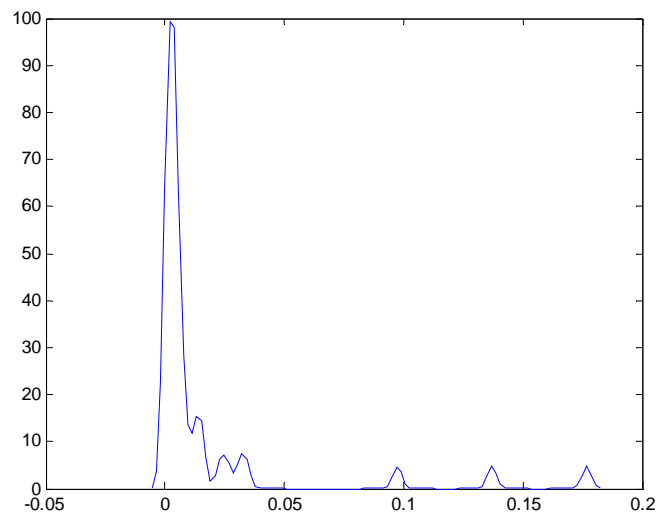
**Fig. 9:** Distribution of Total Loans as of 2013Q3



**Fig. 10:** Distribution of Total Charge-off to Total Loan Ratios as of 2013Q3



**Fig. 11:** Distribution of Liquidity Ratios as of 2013Q3



**Fig. 12:** Distribution of Trading Revenue Deviation to Trading Book as of 2013Q3

#### **4. Estimation Results**

Our specifications of the translog output distance functions are based on the intermediation interpretation of banking services wherein banks utilize deposits and other input factors to provide loan services as their outputs, see Sealey and Lindley (1977). The alternative production approach views deposits as outputs as opposed to inputs proposed by Baltensperger (1980).

Anticipating the discussion to follow, the overall conclusion of our empirical analyses is that the largest surviving banks - in spite of tremendous growth in the last 20 years - have experienced a diminished capacity to provide loan services as they took on increasing levels of risk. This is reflected in a decline in efficiency as implied by the econometric models that allow efficiency levels to vary temporally. In addition, larger banks have lower scale efficiency levels. There is no evidence of scope economies. Finally, there is no evidence of economies of scale for the large banks in our sample.

The elasticities of six inputs and three outputs are evaluated at the sample mean of the data points, in Table 2, where the standard errors are reported in parentheses. We utilize a non-parametric bootstrap following Efron and Tibshirani (1986), which is implemented through 1,000 iterations where in each run, 44 banks are chosen with replacement and 79 quarters are chosen with replacement, and the model is re-estimated. Since our dataset is mean deflated prior to estimating the distance function, the first derivatives expressed in Eq.(0.0) and Eq.(0.0) will simply be equal to the first order coefficients when evaluated at the sample mean.

Table 2: The Elasticity Estimates Evaluated at Sample Mean

	FE	RE	FEIV	REIV	HT	PSS1	BC	QR(50%)
PFA	-0.0486 (0.0501)	-0.0519 (0.0506)	-0.0903 (0.0490)	-0.0875 (0.0809)	-0.0500 (0.0485)	-0.0437 (0.0433)	-0.0253 (0.0562)	-0.0640 (0.0405)
NOE	-0.1745 (0.0637)	-0.2121 (0.0595)	-0.0978 (0.0614)	-0.1571 (0.0728)	-0.1839 (0.0650)	-0.1601 (0.0858)	-0.2116 (0.0749)	-0.1204 (0.0603)
PF	-0.0215 (0.0039)	-0.0202 (0.0030)	-0.0224 (0.0049)	-0.0206 (0.0049)	-0.0212 (0.0051)	-0.0224 (0.0051)	-0.0141 (0.0033)	-0.0195 (0.0029)
SA	-0.5519 (0.0400)	-0.5582 (0.0489)	-0.5453 (0.0401)	-0.5529 (0.0576)	-0.5532 (0.0440)	-0.5611 (0.0376)	-0.5526 (0.0644)	-0.5905 (0.0415)
CD	-0.0586 (0.0135)	-0.0569 (0.0136)	-0.0443 (0.0119)	-0.0423 (0.0170)	-0.0583 (0.0127)	-0.0606 (0.0132)	-0.0712 (0.0192)	-0.0778 (0.0120)
DD	-0.0828 (0.0286)	-0.0984 (0.0363)	-0.0894 (0.0327)	-0.1197 (0.0487)	-0.0861 (0.0314)	-0.0971 (0.0277)	-0.1553 (0.0468)	-0.1086 (0.0242)
REL	0.4028 (0.0495)	0.3793 (0.0480)	0.4247 (0.0548)	0.3878 (0.0348)	0.3982 (0.0595)	0.4125 (0.0635)	0.3029 (0.0606)	0.4823 (0.0348)
CIL	0.2105 (0.0400)	0.2172 (0.0366)	0.2254 (0.0471)	0.2283 (0.0333)	0.2117 (0.0360)	0.2139 (0.0446)	0.2266 (0.0228)	0.1810 (0.0287)
CL	0.0817 (0.0253)	0.0814 (0.0186)	0.0495 (0.0206)	0.0581 (0.0208)	0.0819 (0.0237)	0.0737 (0.0303)	0.0707 (0.0183)	0.0719 (0.0217)
SC	0.2604 (0.0270)	0.2700 (0.0307)	0.2704 (0.0309)	0.2829 (0.0201)	0.2622 (0.0291)	0.2678 (0.0366)	0.3242 (0.0497)	0.2462 (0.0220)
OFF	0.0446 (0.0140)	0.0521 (0.0203)	0.0299 (0.0109)	0.0428 (0.0128)	0.0461 (0.0122)	0.0320 (0.0125)	0.0757 (0.0240)	0.0185 (0.0103)
RST	0.9379 (0.0661)	0.9978 (0.0281)	0.8895 (0.0826)	0.9801 (0.0400)	0.9527 (0.0545)	0.9451 (0.0690)	1.0301 (0.0247)	0.9809 (0.0457)

The elasticity estimates shown in Table 2 are consistent with the monotonicity assumption. The six inputs' elasticities have negative signs, and the three outputs' elasticities have positive signs. Alternatively, all of the input variables (Premises and Fixed Assets, Number of Employees, Purchased Funds, Savings Accounts, Certificates of Deposit and Demand Deposits) contribute positively to the output, albeit varying in magnitude. Compared with the other inputs, SA and DD have the greatest impact. NOE is also an important input source albeit it has less impact than SA and DD; while the estimates of PFA and CD are similar in magnitude. PF has the smallest impact in all the inputs.

Across most models, our estimates suggest no evidence of increasing returns to scale since the numbers are varying closely around 1.

Turning our attention to the controls for risk, which are displayed in the last three rows in Table 4 and

Table 5 in the Appendix, we observe that in all have generally positive signs on coefficient estimates, which have the interpretation that all else equal, risk taking activities decrease output, as more risk is detrimental and reduces the capacity of the banks to make loans. The magnitudes of the coefficient estimates of Credit Risk (CR) are around 10 times smaller than Liquidity Risk (LR). As LR is proxied by the liquidity ratio (cash balance/total assets) one might first expect a negative sign on the coefficient since the positive signs indicated by all of the estimators indicates that increases in the LR reduce the level of intermediation services provided by the bank. It is clear from our estimates that these banks are not managing their liquidity optimally, controlling for market and credit risk. The positive sign for coefficient estimates of Market Risk (MR) suggest that as banks move from traditional banking (i.e. lending business) to trading book activities, banks have become less efficient in lending.

Coefficient estimates on all of the three risk proxies are generally the same across models using both stochastic frontier analysis and quantile regression. The positive signs on the coefficient estimates indicate that greater LR, CR or MR inhibits output. The estimates on MR are generally much less substantial across models; the estimates on LR consistently have more substantial across models than the other two risk proxies. These results regarding LR and MR support the policy argument that banks should be restricted from engaging in highly risky activities, such as proprietary trading, and encouraged to maintain an appropriate liquidity ratio. More generally, our results taken in totality lead to the sensible implication that banks which stray from their core competencies will provide less intermediation services and should shrink over time.

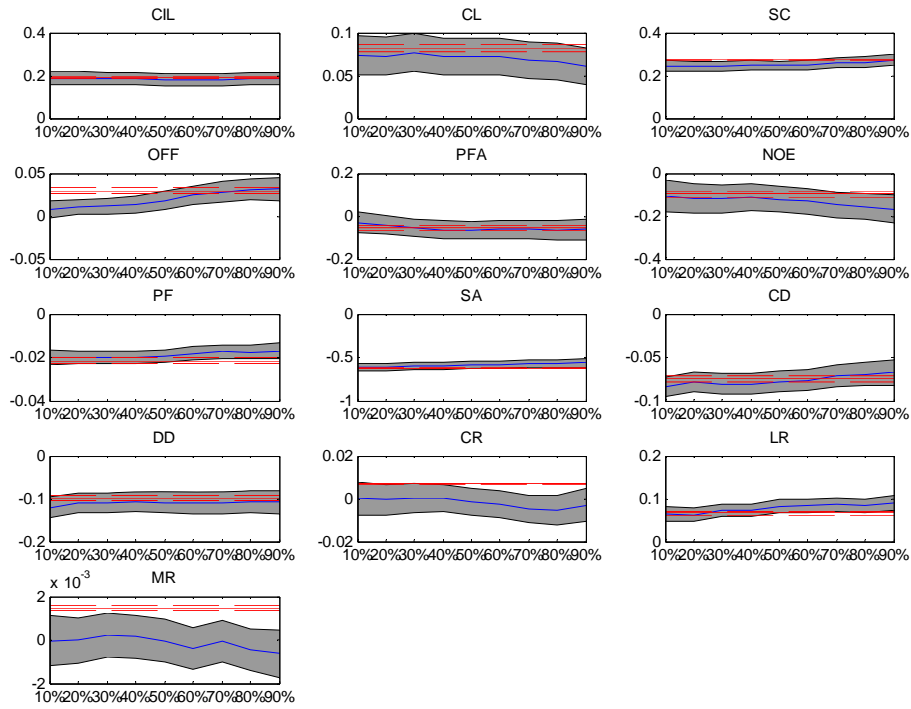
In Fig. 13 and

Table 5, we summarize the estimation results of the quantile regression fixed effects model for panel data. We estimate these models in the R statistical programming language (R Core Development Team, 2010) using the quantreg package by Koenker (2009), which the authors adapt and extend in order to produce longitudinal data results as well as to produce more reliable statistical inference. From the figure below, we can see that the quantile regression estimates on the elasticities, represented in black lines, are compatible with those from Fixed Effect model, which are denoted in the red lines. The elasticity estimates are not varying significantly across quantiles, but the estimates on Credit Risks and Liquidity Risks have displayed a distinctive increasing pattern.<sup>14</sup>

---

<sup>14</sup> The linearity of covariate effects across different quantiles is consistent with the standard interpretation of technical efficiency in the stochastic frontier paradigm as a radial measure.





**Fig. 13:** Panel Data Quantile Regression Elasticity Estimates

Economies of scope, displayed in Table 3 below, are constructed following Hajargasht, Coelli and Rao (2008), who derive the expression for economies of scope in terms of the derivative of the distance functions utilizing the duality between the cost and input distance functions. The economies of scope between outputs  $i$  and  $j$  can be calculated using the derivatives of the output distance function as follows.

$$C_{yy} / C = D_y D_y' - D_{yy} + D_{yx} [D_{xx} + D_x D_x']^{-1} D_{xy} \quad (0.0)$$

Our dataset is centered on the geometric mean of all observations. Results are essentially the same when we center at the median time period as well. This enables us to more transparently interpret the translog results. Economies of scope evaluated at the sample geometric means for the median time period can be calculated following this formula in Eq.(0.0). A positive sign represents scope diseconomies.

$$D_y D_y' - D_{yy} + D_{yx} [D_{xx} + D_x D_x']^{-1} D_{xy} = \begin{bmatrix} \gamma_1 - \gamma_{11} & \cdots & -\gamma_{1m} \\ \vdots & \ddots & \vdots \\ -\gamma_{m1} & \cdots & \gamma_m - \gamma_{mm} \end{bmatrix} + \quad (0.0)$$

$$\begin{bmatrix} \delta_1 \gamma_1 + \theta_{11} & \cdots & \delta_n \gamma_1 + \theta_{n1} \\ \vdots & \ddots & \vdots \\ \delta_1 \gamma_m + \theta_{1m} & \cdots & \delta_n \gamma_m + \theta_{nm} \end{bmatrix} \begin{bmatrix} 2\delta_1^2 + \delta_{11} - \delta_1 & \cdots & 2\delta_1 \delta_n + \delta_{1n} \\ \vdots & \ddots & \vdots \\ 2\delta_n \delta_1 + \delta_{n1} & \cdots & 2\delta_n^2 + \delta_{nn} - \delta_n \end{bmatrix}^{-1} \begin{bmatrix} \delta_1 \gamma_1 + \theta_{11} & \cdots & \delta_1 \gamma_m + \theta_{1m} \\ \vdots & \ddots & \vdots \\ \delta_n \gamma_1 + \theta_{n1} & \cdots & \delta_n \gamma_m + \theta_{nm} \end{bmatrix}$$

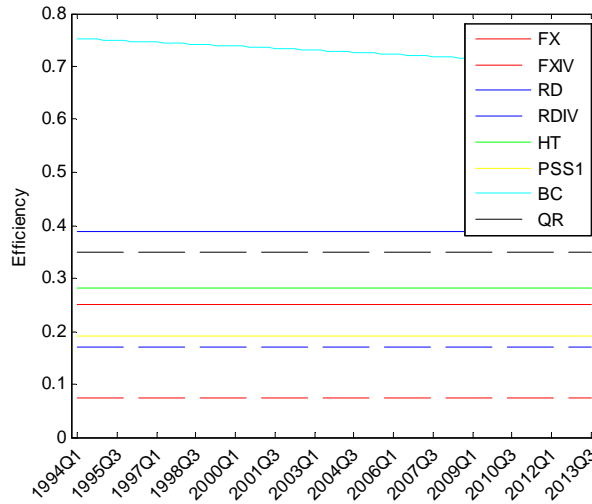
For the standard errors of the scope economy measures, we bootstrapped 1000 times within our dataset. Based on sample measures, it is suggested that there is no evidence of economies of scope across all models among the three different types of loans evaluated at the sample mean point. Our results are consistent with the findings of Hughes and Mester (1993). They base their analysis on the translog cost dual to our primal output distance function. We both find no evidence of scale economies for the largest banks or significant scope economies. It is not clear that alternative nonparametric approaches such as the local linear approximations utilized by Wheelock and Wilson (2012) are directly comparable to our results given their focus on banks of varying sizes and the substantial differences in number of parameters for such models. Constructing tests for the regularity conditions of the dual cost function from such innovative nonparametric approaches is a research issue that requires more study.

Table 3: The Scope Economies Estimates

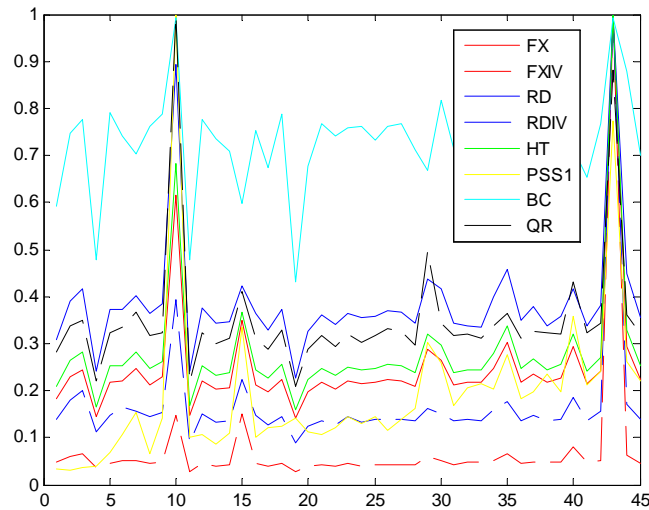
	FE	RE	FEIV	REIV	HT	PSS1	BC	QR(50%)
REL-CIL	0.0106 (0.0613)	0.0108 (0.0267)	0.0399 (0.0376)	0.0278 (0.0430)	0.0107 (0.0299)	0.0212 (0.0751)	0.0298 (0.0303)	0.0160 (0.1134)
REL-CL	0.0266 (0.0400)	0.0273 (0.0161)	0.0307 (0.0292)	0.0456 (0.0516)	0.0267 (0.0144)	0.0257 (0.0137)	0.0237 (0.0147)	0.0267 (0.0289)
REL-SC	0.0354 (0.0855)	0.0322 (0.0157)	0.0178 (0.0459)	0.0037 (0.0781)	0.0353 (0.0228)	0.0360 (0.0212)	0.0125 (0.0477)	0.0290 (0.0450)
REL-OFF	0.0083 (0.0307)	0.0125 (0.0050)	0.0431 (0.0101)	0.0601 (0.0581)	0.0092 (0.0143)	0.0131 (0.0359)	0.0192 (0.0176)	0.0120 (0.0094)
CIL-CL	-0.0030 (0.0907)	-0.0034 (0.0129)	-0.0368 (0.0321)	-0.0502 (0.0411)	-0.0030 (0.0143)	-0.0094 (0.0295)	0.0020 (0.0219)	-0.0044 (0.0201)
CIL-SC	0.0514 (0.0401)	0.0563 (0.0414)	0.0677 (0.0457)	0.0872 (0.0444)	0.0522 (0.0335)	0.0544 (0.0652)	0.0625 (0.0219)	0.0557 (0.0565)
CIL-OFF	0.0013 (0.0150)	0.0020 (0.0097)	-0.0391 (0.0189)	-0.0533 (0.0266)	0.0015 (0.0123)	-0.0066 (0.0210)	0.0043 (0.0087)	-0.0098 (0.0226)
CL-SC	0.0273 (0.0268)	0.0283 (0.0167)	0.0073 (0.0395)	0.0329 (0.0528)	0.0277 (0.0214)	0.0281 (0.0261)	0.0002 (0.0293)	0.0349 (0.0292)
CL-OFF	0.0055 (0.0093)	0.0078 (0.0069)	0.0029 (0.0200)	-0.0011 (0.0171)	0.0059 (0.0062)	0.0066 (0.0110)	0.0175 (0.0066)	0.0053 (0.0129)
SC-OFF	-0.0054 (0.0233)	-0.0085 (0.0088)	0.0066 (0.0211)	0.0161 (0.0157)	-0.0061 (0.0107)	-0.0063 (0.0171)	-0.0157 (0.0233)	-0.0017 (0.0039)

Fig. 14 below summarizes the results of the stochastic frontier estimation in terms of average efficiencies across the different estimators in each quarter. Efficiency levels range between about 0.10 to 0.4 using time-invariant estimators and with a downward trend using the BC model, whose specification requires that the temporal pattern is linear and monotonic and thus the decline in average efficiency over the sample period from 75% to 70%. This trend is probably due to the substantial downturns in the recent period of the Great Recession and the financial meltdown.

The relationship between efficiency levels and bank sizes is also explored. From Fig. 15, we can see that the largest banks do not necessarily have highest technical efficiencies; instead, the efficiency levels are fluctuating as bank sizes change.



**Fig. 14:** Estimated Efficiencies using all Stochastic Frontier Models



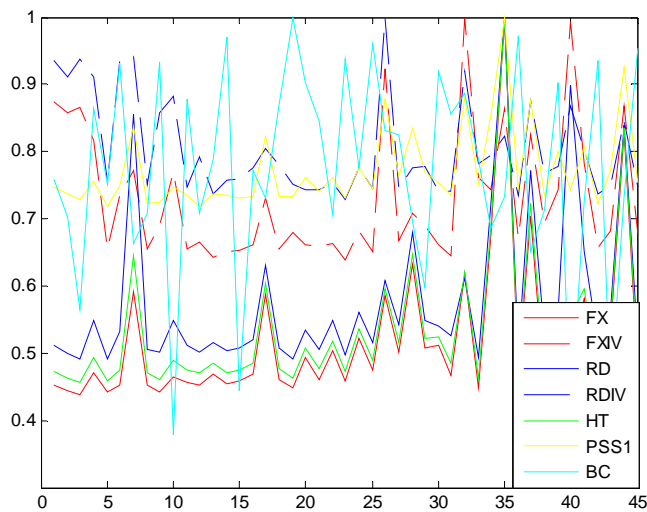
**Fig. 15:** Efficiency Levels and Bank Sizes

We further analyze the relationship between bank sizes and the Output Scale Efficiency ("OSE"). The derivation of this estimator follows Balk (2001).

$$OSE(x, y) = \frac{\overset{\vee}{D}'_o(x, y)}{D'_o(x, y)} = \frac{\overset{\vee}{OTE}'_o(x, y)}{OTE'_o(x, y)} \quad (0.0)$$

where the  $\overset{\vee}{OTE}'_o(x, y)$  is the output efficiency using cone technology (i.e., constant returns to scale - "CRS".) As we can see in Fig. 16, which plots this OSE versus size ranking, the scale efficiencies estimated using time-invariant estimators are increasing

with fluctuations as bank sizes decrease (the ranking numbers increase). The scale efficiency level using BC estimator<sup>15</sup>, although displays a more fluctuating pattern than those using time-invariant estimator, still suggests that large banks do not necessarily have higher scale efficiency levels.



**Fig. 16:** Scale Efficiency Plots using Time-invariant Estimators

## 5. Conclusion and Directions for Future Research

This study represents a contribution to the recent dialogue that has arisen in the wake of the recent financial crisis, a reexamination amongst regulators, practitioners and academicians of the conventional wisdom regarding the implications of the spectacular growth of the financial sector of the economy. Previously, there was a widespread belief the "bigger is better", with arguments underpinning this view ranging from potential economies of scale and scope, to a better competitive stance at the international level. We have seen this logic reversed in the post-crisis world to some degree, as for several reasons large banks have come to be viewed as a source of trouble and concern for policy makers and regulators.

<sup>15</sup> For the BC estimator, we use the average-over-time scale efficiency level.

We have addressed this controversy through an empirical analysis of the efficiency of U.S. banks with respect to their size and scope. This study utilized a new data-set of bank history, a panel of financial measures derived from supervisory Call Reports in the period 1994-2013, from which we construct the variables used in both the frontier estimation and quantile regression analyses (inputs and outputs, as well as controls for 3 major risk types - credit, market and liquidity.) In this exercise we have been able to develop both policy implications and also evaluate potential analytical tools for supervisors.

The conclusion of the stochastic frontier estimation is that in spite of growing, the largest U.S. surviving banks have decreased technical efficiency over the last 20 years. This has occurred as they took on increasing types of risk, and this is reflected in an overall decline in efficiency since early, as implied by the econometric model that allow this to vary temporally. The estimation results also revealed no evidence on increasing returns to scale or scope across models. According to the time-invariant estimators, there is no positive correlation between bank size and technical efficiencies, and neither exists such a relationship between size and scale efficiencies. We found that credit, liquidity and market risks are deleterious to efficiency, which has implications for the argument that banks should be restricted to traditional banking activities in their zone of competence. The panel quantile regression results were generally consistent with the stochastic frontier estimation, albeit with estimates not varying greatly across quantiles. Furthermore, the implied efficiencies here are uniformly lower in the quantile regressions, than for the other time-invariant frontier estimators.

This paper has both policy implications and also evaluates various econometric techniques as potentially valuable analytical tools for supervisors. First, our results highlight the importance of the prudential supervisory role in controlling the level of risk in the banking sector (also reducing incentive for regulatory arbitrage between the banking and trading books), as we have documented that the elevation in risk measures coupled with the growth of the sector has resulted in declining measures of

efficiency, a result that is robust to several econometric specifications. The policy implication is that we may want a better capitalized and somewhat smaller banking system, as this is likely to imply a more efficiently functioning banking industry. Second, the finding that market and liquidity risk dominate the influence of credit risk implied in the Volcker Rule debate, that regulators may wish not only to consider restricting banks from dangerous activities such as speculative proprietary trading, but also closely monitor the OTC exposures and their use of hedging some market risks instead of market making purposes and consequently encourage insured commercial banks to focus on their core competency of making loans. There are several fruitful avenues of extension for this research program. We may pursue alternative data-sets, such as other financial service types of firms (e.g., insurers, brokers), or data from other jurisdictions. We may expand our set of explanatory variables, with alternative controls (e.g., size, leverage, capitalization), or an expanded set of inputs (e.g., a measure of technological change.) Finally, we may expand our suite of alternative models, thereby seeking out further robust tools for the use by supervisors.







## References

- Allen and Santomero (2001) What do financial intermediaries do?, *Journal of Banking and Finance* 2001 Vol: 25:271-294
- Balk, B.M., 2001, Scale efficiency and productivity change. *Journal of productivity analysis* 15, 159-183.
- Balk, B.M., 2008, Price and quantity index numbers: models for measuring aggregate change and difference. Cambridge University Press, Cambridge.
- Baltensperger, E., 1980, Alternative approaches to the theory of the banking firm. *Journal of Monetary Economics* 6, 1-37.
- Battese, G.E., T.J. Coelli, 1992, Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. *Journal of productivity analysis* 3, 153-169.
- BCBS 2009a: "Revisions to the Basel II market risk framework" updated December 2010.
- BCBS, 2009b, "Principles for Sound Stress Testing Practices and Supervision - Consultative Paper, May (No. 155).
- BCBS, 2012, "Fundamental review of the trading book", <http://www.bis.org/publ/bcbs219.pdf>
- Bernanke, B., 2009, Financial reform to address systemic risk. Council on Foreign Relations, Washington, DC March 10, 2009.
- Caves, D.W., L.R. Christensen, W.E. Diewert, 1982, The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica* 50, 1393-1414.
- Coelli, T., 2000, On the econometric estimation of the distance function representation of a production technology. Center for Operations Research & Econometrics, Universite Catholique de Louvain.
- Coelli, T., S. Perelman, 1996, Efficiency measurement, multiple-output technologie and distance functions: With application to European Railways. CREPP Working Paper 96/05, University of Liege.
- DeYoung, R., 2010, Scale Economies Are a Distraction. The Region, Federal Reserve Bank of Minneapolis.
- Efron, B., R. Tibshirani, 1986, Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science* 1, 54-75.
- Feldman, R., 2010, Size and Regulatory Reform in Finance: Important but Difficult Questions. The Region, Federal Reserve Bank of Minneapolis.
- Frye, J. and E. Pelz, 2008, BankCaR (Bank Capital-at-Risk): US commercial bank chargeoffs, Working paper, Federal Reserve Bank of Chicago (2008:3).
- Galvao, A.F., 2011, Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics* 164, 142-157.
- Gregory, J., 2014, Central Counterparties: Mandatory clearing and bilateral margin requirements for OTC derivatives
- Hajargasht, G., T. Coelli, D. Rao, 2008, A dual measure of economies of scope. *Economics Letters* 100, 185-188.
- Haldane, 2009, "Banking on the state", <http://www.bis.org/review/r091111e.pdf>
- Hausman, J.A., W.E. Taylor, 1981, Panel data and unobservable individual effects. *Econometrica* 49, 1377-1398.

Hughes, J.P., L.J. Mester, 1998, Bank capitalization and cost: Evidence of scale economies in risk management and signaling. *Review of economics and statistics* 80, 314-325.

Hughes, J.P., L.J. Mester, 2008, Efficiency in banking: theory, practice and evidence. FRB of Philadelphia Working Paper No. 08-1.

Hughes, J.P., L.J. Mester, C.G. Moon, 2001, Are scale economies in banking elusive or illusive?: Evidence obtained by incorporating capital structure and risk-taking into models of bank production. *Journal of Banking & Finance* 25, 2169-2208.

Inanoglu, H., Jacobs, Jr., M., and A.K., Karagozoglu, 2014 (Spring), Empirical analysis of bank capital and new regulatory requirements for risks in trading portfolios, *Journal of Fixed Income*, 23:4, 71-88.

Jacobs, Jr., Michael, 2014, Supervisory requirements and expectations for portfolio level counterparty credit risk measurement and management, *Journal of Financial Regulation and Compliance*, 22-3, 253-270.

Jorion, P., 2002, How informative are value-at-risk disclosures. *The Accounting Review* 77 (4), 911-931

Klein, L.R., 1953, A textbook of econometrics. Row, Peterson & Company, Evanston, IL.

Koenker, R., 1984, A note on L-estimates for linear models. *Statistics & probability letters* 2, 323-325.

Koenker, R., 2004, Quantile regression for longitudinal data. *Journal of Multivariate Analysis* 91, 74-89.

Koenker, R., 2009, Quantreg: quantile regression. R package version 4.

Koenker, R., G. Bassett Jr, 1978, Regression quantiles. *Econometrica* 46, 33-50.

Mosteller, F., 1946, On Some Useful "Inefficient" Statistics. *The Annals of Mathematical Statistics* 17, 377-408.

Newey, W.K., 1990, Semiparametric efficiency bounds. *Journal of applied econometrics* 5, 99-135.

Pagan, A., A. Ullah, 1999, Nonparametric econometrics. Cambridge University Press, Cambridge.

Park, B.U., R.C. Sickles, L. Simar, 1998, Stochastic panel frontiers: A semiparametric approach. *Journal of Econometrics* 84, 273-301.

Park, B.U., R.C. Sickles, L. Simar, 2003, Semiparametric-efficient estimation of AR (1) panel data models. *Journal of Econometrics* 117, 279-309.

Park, B.U., R.C. Sickles, L. Simar, 2007, Semiparametric efficient estimation of dynamic panel data models. *Journal of Econometrics* 136, 281-301.

Park, B.U., L. Simar, 1994, Efficient semiparametric estimation in a stochastic frontier model. *Journal of the American Statistical Association* 89, 929-936.

Pitt, M.M., L.F. Lee, 1981, The measurement and sources of technical inefficiency in the Indonesian weaving industry. *Journal of Development Economics* 9, 43-64.

R Core Development Team, 2010, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ritter, C., L. Simar, 1997, Pitfalls of normal-gamma stochastic frontier models. *Journal of productivity analysis* 8, 167-182.

Schmidt, P., R.C. Sickles, 1984, Production frontiers and panel data. *Journal of Business & Economic Statistics* 2, 367-374.

Sealey, C.W., J.T. Lindley, 1977, Inputs, outputs, and a theory of production and cost at depository financial institutions. *The Journal of Finance* 32, 1251-1266.

Sickles, R.C., 2005, Panel estimators and the identification of firm-specific efficiency levels in parametric, semiparametric and nonparametric settings. *Journal of Econometrics* 126, 305-334.

Tarullo, D.K., 2011, Industrial Organization and Systemic Risk: An agenda for further research. Conference on Regulating Systemic Risk, Washington, DC Sept. 15, 2011.

Tracey, B., R. Davies, 2012, Too big to be efficient? The impact of implicit funding subsidies on scale economies in banking. Working Paper.

Wheelock, D.C., P.W. Wilson, 2012, Do Large Banks have Lower Costs? New Estimates of Returns to Scale for US Banks. *Journal of Money, Credit and Banking* 44, 171-199.