Article:
Glass, A.J. orcid.org/0000-0002-5984-7666, Kenjegalieva, K. and Weyman-Jones, T. (2020) The effect of monetary policy on bank competition using the Boone index. European Journal of Operational Research, 282 (3). pp. 1070-1087. ISSN 0377-2217

https://doi.org/10.1016/j.ejor.2019.10.022

The Effect of Monetary Policy on Bank Competition using the Boone Index

Anthony J. Glass[*,†], Karligash Kenjegalieva[‡] and Thomas Weyman-Jones[§]

October 2019

Abstract

An interesting strand of the theoretical literature on measuring competition posits that when competition increases in an industry, output is reallocated to more efficient firms. Our first contribution is on the methodology for the empirical implementation of this theoretical test of a change in competition. This contribution moves from the relationship between a change in competition and a single all-encompassing efficiency, to a set of relationships between a change in competition and multiple efficiencies that measure different components of economic performance. Our second contribution is to apply our empirical methodology to large U.S. banks. The results suggest that competition intensified between these banks during the financial crisis and beyond $(2008-15)$, vis-à-vis our pre-crisis period $(1994-07)$. This points to an increase in competition that has exogenous origins such as the decrease in the loan-deposit rate spread, which represented the collateral damage to banks from monetary policy to moderate the Great Recession.

Key words: Productivity and competitiveness; Competitive regimes; Network efficiency; Large U.S. banks; Exogenous change in competition.

[*]Corresponding author

[†]School of Business and Economics and Centre for Productivity and Performance, Loughborough University, Loughborough, Leics, UK, LE11 3TU. Email: a.j.glass@lboro.ac.uk

[‡]School of Business and Economics and Centre for Productivity and Performance, Loughborough University, Loughborough, Leics, UK, LE11 3TU. Email: k.a.kenjegalieva@lboro.ac.uk

[§]School of Business and Economics and Centre for Productivity and Performance, Loughborough University, Loughborough, Leics, UK, LE11 3TU. Email: t.g.weyman-jones@lboro.ac.uk

1 Introduction

It is well-established that industries with a high degree of competition are associated with a number of consumer benefits including, to name but a couple, greater product choice and lower prices. Robust methods of measuring competition are therefore imperative to provide an insight into the relationship between changes in the level of competition and changes in these consumer benefits. Given the magnitude and number of benefits from a high degree of competition a vast literature has accrued on measuring the degree of competitiveness in an industry.[1]

A further promising strand of the aforementioned literature posits that a change in the degree of competitiveness alters the relationship between the efficiency and profitability of firms in an industry. To illustrate, when there is an increase in competition in an industry this literature posits that output will be reallocated to more efficient firms, which highlights how a change in competition is related to firms' efficiencies. Boone (2008) presents a theoretical model of this relationship between the level of competition and a single all-encompassing measure of efficiency. In this paper we focus on extending and applying the empirical methodology to operationalize this theoretical relationship.

The empirical Boone literature is a small but emerging area (Duygun *et al.*, 2015; Bolt and Humphrey, 2010; 2015a; 2015b) that consists of two approaches. The approach in Duygun *et al.* tests the hypothesis that competition differs between two regimes, and in the papers by Bolt and Humphrey a frontier approach is used to calculate competition efficiencies. Both approaches use efficiency estimates in an innovative way, which is important as it helps efficiency measurement transcend other fields. Our first contribution extends the method that Duygun *et al.* develop to implement the Boone test empirically. This contribution draws on advances in efficiency measurement in the stochastic frontier literature to progress from the relationship between a change in competition and a single all-encompassing efficiency, to a set of relationships between a change in competition and multiple efficiencies that measure different components of economic performance. We distinguish between different sources of a change in competition using multiple efficiencies with different features. The different features of the efficiencies are time-variance, time-invariance, internalization to the firm and network effects as firms compete via their location choices for, among other things, network linkages. Our second contribution is to apply our extended empirical Boone test to large U.S. banks, as this is the size category that has the biggest impact on competition in the industry. Both our contributions are timely, as the Boone index was recently highlighted by the Bank for International Settlements (2018) as an empirical tool to measure changes in bank competition. Additionally, since the literature on bank efficiency covers a wide range of countries (e.g., European countries (Fiordelisi and Molyneux, 2010, Liu *et al.*, 2013, and Casu *et al.*, 2016, to name but a few), G7 countries (Goddard and Wilson, 2009) and countries in Latin America (Yeyati and Micco, 2007)), there is scope for wider application of our extended methodology to banks in other countries and to firms in other industries.

[1] Commonly used measures of competitiveness particularly in the banking literature include the Herfindahl-Hirschman index (HHI), the Lerner index and the H-statistic (Panzar and Rosse, 1987). For more discussion of these measures in a banking context see Claessens and Laeven (2004), DeGryse *et al.* (2009) and Bolt and Humphrey (2015a; 2015b).

As our two contributions relate to a second stage parametric analysis, for clarity we set out how these contributions fit within the broader context of our two-stage framework. The first stage of our analysis consists of the following two parts, which are adapted to the setting of bank branch networks from the spatial stochastic frontier literature on geographical areas (e.g., countries (Glass *et al.*, 2016a) and states (Glass *et al.*, 2016b)).

(i) Estimate the structural form of our network stochastic frontier model which includes own time-varying and own time-invariant inefficiencies, which we refer to as net inefficiencies as they are net of time-(in)variance. The resulting own net efficiencies are then used to calculate a combined own time-varying efficiency, which in line with our net terminology we refer to as a gross measure. These net and gross efficiencies are standard as they are net of any efficiency spillover across the network. We estimate a network stochastic profit frontier in our application as it is more closely aligned to our second stage Boone test. This is because, across firms in an industry, this is a test of a change in the relationship between a firm's relative profit difference (RPD) and its relative efficiency difference (RED).

(ii) Transform the estimated structural form of our spatial frontier into its reduced form to relate the own net and gross efficiencies to three further efficiency measures (internal, network and overall), which are all partially / entirely made up of an efficiency spillover across the network. We discuss these three further efficiency measures in detail in due course. In brief at this juncture, own and internal efficiencies are similar in that they relate to an individual firm, but at the same time they differ as internal efficiency is own efficiency plus efficiency feedback, where the latter is efficiency that reverberates back to a firm from other firms in the network. Network efficiency is the sum of the efficiency spillovers that gravitate to a firm from all the other firms in the network, and overall efficiency is a composite measure of the internal and network efficiencies.

The second stage of our analysis relates to our extended empirical Boone test and can be also be subdivided into two parts as follows.

(i) Use the net time-varying, net time-invariant and gross time-varying internal, network and overall efficiencies to calculate a series of RED measures. To avoid concerns about endogeneity calculate the RPD using an alternative measure of profitability to the first stage measure.

(ii) Identify two subsamples that represent different competitive regimes. For each subsample, first regress the RPD on the gross time-varying overall RED, and then regress the RPD on each departmentalization of this gross RED (e.g., net time-varying and net time-invariant overall REDs). By applying the Boone test to the regressions of RPD on the gross time-varying overall RED for the two subsamples, we determine if there is a difference in competition in the two regimes. This mirrors the extant literature on the empirical Boone test that uses a single efficiency measure. By applying our extended Boone test to the regressions of RPD on departmentalizations of the gross time-varying overall RED for the two subsamples, we determine the sources of a change in competition.

Having outlined our second stage contributions, we also note that our first stage stochastic frontier method builds on the HHI that Hirtle (2007) uses to account for differences in market competition in U.S. bank performance models. This bank level HHI is obtained by first calculating domestic deposit based HHIs for individual banks for each geographical market, where a market is taken to be a metropolitan statistical area (MSA) or non-MSA county where a bank operates. A branch deposit share weighted sum of these HHIs is then calculated for each bank. Since multiple banks operate in the same geographical market, branch deposits will be spatially correlated because branches face the same local and regional market conditions, which will be inherent in the deposit data. This HHI therefore captures the spatial correlation in branch deposits, but this is just one dimension of the spatial correlation between banks' variables. By not accounting for the other dimensions there is potentially an omitted variable bias, which may lead to biased efficiency estimates and, as a result, biased measurement of changes in competition. To capture the spatial correlation between banks' variables more fully we use spatial / network techniques that have been specifically designed for this purpose. These techniques account for inter-branch network interactions using a network linkage matrix, otherwise referred to in spatial econometrics as the spatial weights matrix. As we explicitly model the spatial correlation between banks' variables using a spatial / network stochastic frontier, we gain rich insights into the inter-bank spillovers and, in particular, the network efficiency spillovers between banks which play an important role in our second stage Boone analysis.

There are three parts to our empirical application of our extended Boone test to large U.S. banks. In the first part, for our entire sample of banks and two subsamples (globally and domestically systemically important (GDSI) banks, according to a classification in 2014 / 15, and the remaining banks), we test for the sources of a change in competition during the financial crisis and beyond ($2008-15$), vis-à-vis our pre-crisis period ($1994-07$). Among other things, we find that there was an increase in competition in $2008-15$ for the entire sample, GDSI banks and non-GDSI banks. This is intuitive because it is consistent with the profits of GDSI banks, and to lesser extent the profits of non-GDSI banks, declining during $2008-15$, which occurred for a variety of reasons that we discuss in the empirical application.[2]

A change in competition can be endogenous, which is actionable by the antitrust authorities, or exogenous, which the antitrust authorities can do little, if anything, about. In our empirical application crisis induced changes are consistent with an exogenous increase in competition. There are three such crisis induced phenomena- the U.S. economy going into recession; the decrease in the loan-deposit rate spread (i.e., loan returns fell more than deposit rates without the latter dropping below zero), which represented the collateral damage to banks from monetary policy to moderate the recession; and the tightening of bank regulation to reduce risk. It is not possible though to test whether there has been an endogenous change in competition in our empirical application because the exogenous influences on bank profits have not remained constant over our study period. What we do know though is that in the latter period in the first part, crisis induced phenomena forced a decline in profits upon the banking industry. In the second part therefore, we consider the origins of the exogenous increase in competition in $2008-15$ that we observe in the first part, by considering competition over the business cycle.[3] We examine this by analyzing for the pre-crisis period and the period covering the crisis and beyond, whether the expected increase in competition in the recession (i.e., the first portion of each period) is cancelled

[2] We thank an anonymous reviewer for highlighting the consistency between developments in the industry and the finding from the first part of our Boone analysis that competition intensified.

[3] We thank an anonymous reviewer for suggesting this research direction.

out by the expected decrease in the expansion phase (i.e., the second portion of each period).[4,5] For both periods we find no change in competition over the business cycle. This suggests that the exogenous increase in competition we observe for $2008 - 15$ has two origins- the collateral damage to banks from monetary policy to moderate the recession that the crisis triggered, and the post-crisis tightening of bank regulation to reduce risk.

In the first part of our Boone analysis we consider rather long subperiods, and the GDSI banks are limited in number and are based on a 2014 / 15 classification so they are fixed over time. We therefore in the third and final part test for the sources of a change in competition between successive shorter subperiods for our entire sample of banks and two subsamples of banks that change annually. Our first subsample comprises banks with total assets in the top third of our sample in each year and our second subsample consists of the remaining banks.[6] Moreover, in all three parts of our empirical Boone analysis, our decision to incorporate multiple efficiency measures into the test is vindicated by the additional findings.

The remainder of this paper is organized as follows. In section 2 we provide an overview of the role of the single all-encompassing measure of efficiency in the Boone theory. Section 3 discusses issues that arise in the empirical implementation of our Boone test methodology for multiple efficiencies. In section 4 we present the empirical application, which involves, among other things, setting out the network stochastic standard profit frontier we use to obtain the first stage efficiencies. Some concluding remarks are then made in section 5.

2 Overview of the Role of Efficiency in the Boone Theory

The setting for the Boone model is an industry comprising competing firms with different efficiencies. Competition is modeled using a two-stage game, where in the first stage firms choose between entering or not entering the market. Knowing which firms entered in the first stage, the firms that chose to enter strategically maximize their post-entry profits in the second stage. Boone identifies a subgame perfect equilibrium where a firm's profit is related to its efficiency, conditional on factors such as the aggressiveness of the firm's conduct in the market.

Following Boone, let $\pi(E)$ represent the profit level of a firm in an industry, which is a function of its efficiency level, E. Now consider three firms in an industry with different efficiency levels, $\min E \leq E \leq \max E$. Using terminology from the Boone paper, Eq. 1 then gives the RPD for a firm, ρ. The RPD measures the difference between the profits of a typical firm and the least efficient firm relative to the difference between the profits of the most and least efficient firms.

$$\rho = \frac{\pi(E) - \pi(\min E)}{\pi(\max E) - \pi(\min E)}. \tag{1}$$

[4] We only analyze our full sample of banks and non-GDSI banks for the business cycle in the pre-crisis period because the recession phase lasts for only one year. There is not therefore sufficient observations to analyze GDSI banks for this recession.

[5] For the period covering the crisis and beyond we consider an incomplete business cycle. This is because the expansion phase in the second portion of this period continues beyond the end of our study period. Even though this expansion phase is incomplete, it is quite long so there is no shortage of observations for GDSI banks.

[6] We thank an anonymous reviewer for this recommendation.

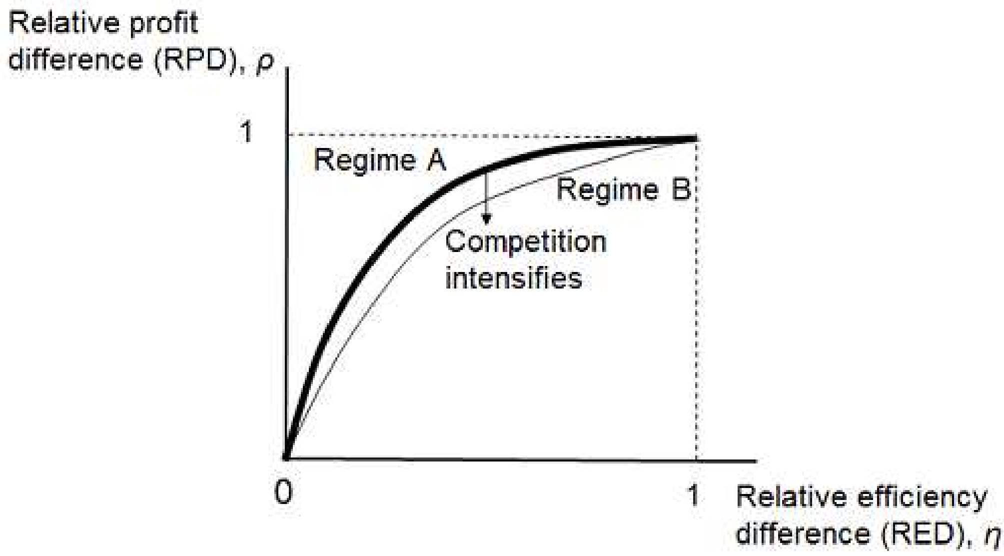


Figure 1: The relative profit difference function from the Boone theory

Boone's model is based on the premise that ρ falls (rises) when there is increase (decrease) in competition in the industry because the numerator in Eq. 1 will fall (rise) by more than the denominator. The intuition is- when an industry becomes more competitive, there is a progressively larger reallocation of output from a firm, the larger the gap between the firm's efficiency and the industry maximum, which results in a progressively larger fall in the firm's RPD. The explanation for this is- the lower a firm's competitiveness, the more the firm will be disciplined by the market when there is an increase in competition.

The RED, η, is calculated as

$$\eta = \frac{E - \min E}{\max E - \min E}. \tag{2}$$

We present the relationship that Boone established between the RED and RPD, $\rho(\eta)$, in figure 1 for two competitive regimes. The bold schedule represents regime A and the other schedule depicts regime B. These schedules are also a diagrammatic representation of Theorem 1 in the Boone paper, which states that when competition intensifies, $\rho(\eta)$ shifts down for all values of η. Boone compares the values of the integrals under different $\rho(\eta)$ schedules, $\int_0^1 \rho(\eta)\, d\eta$, to ascertain if there is a change in competition under different regimes. Since in figure 1 the value of the integral under the $\rho(\eta)$ schedule for regime B is smaller than that for regime A, we conclude that competition is more intense under regime B.

To summarize the features of the Boone index we compare it to the Lerner index. If efficiency is from a cost frontier, we can see from Eqs. 1 and 2 that the Boone index focuses on changes in relative costs (efficiency) and their effect on relative profits. The Lerner index, however, focuses on changes in relative revenues (via relative prices) and their likely effect on relative profits. In determining the level of competition and ultimately whether it is beneficial overall, the Boone index takes relative revenues and prices as "given", while the Lerner index treats relative costs as "given". As is also the case for the Boone index (see the introductory section), if the Lerner index indicates a change in competition, the change can be endogenous, which is actionable by the antitrust authorities, or exogenous, which the antitrust authorities can do little, if anything, about.[7]

3 Empirical Implementation of the Extended Boone Test

3.1 Profit Components, Multiple Efficiencies and Sources of Competition

To operationalize Boone's theory to empirically test for a difference in competition between two regimes, at the outset ρ must be regressed on η for each regime. For panel data, where in each cross-section there are N firms (indexed $i = 1, ..., N$) that operate over T periods (indexed $t = 1, ..., T$), we can write the RPD as $\rho_{it} = \frac{\pi(E_{it})-\pi(\min E_{it})}{\pi(\max E_{it})-\pi(\min E_{it})}$. A feature of the calculation of the RPD for panel data is that the least and most efficient firms can vary over time.

To calculate ρ_{it} some measure of profit is needed. In the empirical Boone test in Duygun *et al.* (2015), the shadow return on equity capital is used as the measure of profitability. Whatever data is used as the measure of profit, it is important to recognize that the data is made up of different profit components. Although we do not suggest that profit should be decomposed and the RPD test applied to the components, these components must be explained in the regression of ρ on η, otherwise there is a potential omitted variable bias in the measurement of changes in competition. To illustrate, in a network industry such as banking, some of a bank's profit can be attributed to the bank's network linkages because of, for example, inter-bank lending, as opposed to emanating from the bank in isolation. Only using in the RPD test therefore an own RED based on a conventional own efficiency measure from a standard non-network stochastic frontier model would not account for the relative efficiency spillover difference, which could potentially lead to biased measurement of changes in competition. This relative efficiency spillover difference arises because of the efficiency spillovers that gravitate to a firm from all the other firms in the same network. Hence why we refer to these efficiency spillovers as a bank's network efficiency and the associated relative efficiency spillover difference as a bank's network RED.[8]

We first address the above potential omitted variable bias in the RPD test by using an overall RED based on overall efficiency from a network stochastic frontier model, which incorporates a firm's internal and network efficiencies. We then draw on recent advances in efficiency measurement in the stochastic frontier literature to departmentalize the overall RED in the empirical Boone test into: (i) time-varying and time-invariant REDs; and (ii) time-varying and time-invariant internal and network REDs. We will discuss our use of the terminology 'departmentalized overall RED' with reference to Eq. 3, which we turn our attention to next.

Departmentalizing the overall RED into K types (indexed $k = 1, ..., K$) involves computing

[7] We thank an anonymous reviewer for suggesting that we compare the Boone and Lerner indices to highlight the features of the former.

[8] In our empirical application we focus on spatial bank networks. If banks have overlapping branch networks they are taken to be part of the same spatial network. This is because by operating in the same markets banks' variables will be spatially correlated. This must be accounted for using a spatial model such as the spatial stochastic frontier we use in our application. Using this model we obtain the inter-bank efficiency spillovers that feature in our second stage Boone analysis.

K RED measures. ρ_{it} is then regressed on K RED measures, where

$$\eta_{kit} = \frac{E_{kit} - \min E_{kit}}{\max E_{kit} - \min E_{kit}}. \tag{3}$$

A statistically significant difference between the value of the integrals pertaining to two sample scatters of points that represent different RPD schedules can therefore be related to the contribution of each of the K RED measures.[9] We interpret these contributions as K sources of a change in competition.

3.2 Testing for Sources of a Change in Competition

The theoretical test of a change in competition from the kth source is a sign criterion that compares the integrals for two competitive regimes. This test is therefore of the sign of the difference between the theoretical integrals in Eq. 4. If $\widehat{\Delta}_k$ is positive (negative), the kth source has led to competition under regime B being less (more) intense than under regime A.

$$\widehat{\Delta}_k = \int_0^1 \rho_{it}^B \left(\eta_{kit}^B\right) d\eta_k^B - \int_0^1 \rho_{it}^A \left(\eta_{kit}^A\right) d\eta_k^A. \tag{4}$$

To construct a sample scatter of points that represent the relationship between the RPD and the kth type of RED (i.e., $\rho_{it}\left(\eta_{kit}\right)$) we employ the following polynomial quantile regression (PQR).

$$\Pr\left(\rho_{it} \leq \rho_{it}\left(\eta_{kit}\right) = \sum_{m=1}^{m=M} \alpha_{km}\eta_{kit}^{m-1}\right) = q, \tag{5}$$

where the PQR is based on the parameter α_{km}; the degree of the polynomial M; and the probability of isolating the proportion of the sample on or below the quantile regression line q. We use PQR rather than, for example, OLS because PQR yields the empirical integral estimate in Eq. 6 which is a linear function of the quantile regression coefficients. It is therefore appropriate to use PQR to estimate the theoretical integrals in Eq. 4.

$$\int_0^1 \left[\sum_{m=1}^{m=M} \widehat{\alpha}_{km}\eta_{kit}^{m-1}\right] d\eta_k = \sum_{m=1}^{m=M} \frac{\widehat{\alpha}_{km}}{m} = \mathbf{r}'\widehat{\alpha}_k, \tag{6}$$

where $\widehat{\alpha}_k$ denotes the vector of estimated coefficients from the PQR and $\mathbf{r}'$ is the vector $\left(1, \frac{1}{2}, ..., \frac{1}{M}\right)$. From the variance matrix of $\widehat{\alpha}_k$ the standard error of the integral in Eq. 6 is

$$SE\left(\int_0^1 \left[\sum_{m=1}^{m=M} \widehat{\alpha}_{km}\eta_{kit}^{m-1}\right] d\eta_k\right) = SE\left(\sum_{m=1}^{m=M} \frac{\widehat{\alpha}_{km}}{m}\right) = \left(\mathbf{r}' var\left(\widehat{\alpha}_k\right)\mathbf{r}\right)^{1/2}. \tag{7}$$

Specifically, to test if there is a change in competition from the kth source under competitive regime B vis-à-vis regime A, the null and alternative hypotheses are as follows.

[9] We take care above to use the terminology 'departmentalized overall RED' because we are not able to decompose the overall RED into different RED measures in the usual sense. This is because each RED is calculated using different relatives.

$$H_0 : \Delta = \int_0^1 \left[\sum_{m=1}^{m=M} \alpha_{km}^B \eta_{kit}^{m-1} \right] d\eta_k - \int_0^1 \left[\sum_{m=1}^{m=M} \alpha_{km}^A \eta_{kit}^{m-1} \right] d\eta_k = 0,$$
$$H_1 : \Delta \neq 0. \tag{8}$$

To fix ideas, the empirical methodology to test for a change in competition from the kth source between two regimes, A and B, has two parts. First, we compute the values of the integrals pertaining to the kth source of competition for each competitive regime by splitting the sample and estimating the PQR in Eq. 5 for each regime. Second, for the kth source of competition, testing if the integrals differ between the two regimes involves pooling the data for the regimes and estimating the following PQR.

$$\Pr\left(\rho_{it} \leq \sum_{m=1}^{m=M} \alpha_{km} \eta_{kit}^{m-1} + \sum_{m=1}^{m=M} \beta_{km} \left(\eta_{kit}^{m-1} \times D_{it} \right) \right) = q, \tag{9}$$

where β_{km} is a coefficient and D_{it} is a competitive regime dummy variable that takes a value of 0 for regime A and 1 for regime B. In terms of this pooled PQR the null in Eq. 8 is $\Delta = \mathbf{r}'\beta_k = 0$, where β_k is the vector of β_{km} coefficients. We test this null via a Wald test with 1 restriction using the F-distributed test statistic in Eq. 10.

$$\Omega = \left(\mathbf{r}'\widehat{\beta}_k \right) \left[\mathbf{r}' var\left(\widehat{\beta}_k \right) \mathbf{r} \right]^{-1} \left(\mathbf{r}'\widehat{\beta}_k \right). \tag{10}$$

Following Duygun *et al.* (2015) we use the fitted PQR at the third quartile. The third quartile is chosen because it is a compromise between inclusivity of the sample points and avoidance of undue impact from outliers, as 75% of the sample scatter of points will lie on or below the fitted regression line.

4 Application to Large U.S. Banks

4.1 Network Stochastic Standard Profit Frontier

The general specification of the structural form of the panel data network stochastic standard profit frontier that we estimate is

$$\pi_t = \zeta\iota + TL\left(\mathbf{p}_t, \mathbf{w}_t, t\right) + \phi'\mathbf{L}_N\mathbf{p}_t + \omega'\mathbf{L}_N\mathbf{w}_t + \lambda\mathbf{L}_N\pi_t + \xi + \mathbf{v}_t - \delta - \mathbf{u}_t \tag{11}$$
$$\xi \sim \mathrm{N}\left(0, \sigma_\xi^2\right) \;\; \delta \sim \mathrm{N}^+\left(0, \sigma_\delta^2\right) \;\; \mathbf{v}_t \sim \mathrm{N}\left(0, \sigma_v^2\right) \;\; \mathbf{u}_t \sim \mathrm{N}^+\left(0, \sigma_u^2\right).$$

The variables are in log form and in each cross-section there are N firms (indexed $i = 1, ..., N$), which operate over T periods (indexed $t = 1, ..., T$). As the network linkage matrix $\mathbf{L}_N$ is key to Eq. 11 we present the model in vector form, where successively stacked cross-sections are denoted by dropping the i subscripts on the variables and error components. π_t denotes the vector of stacked cross-sectional observations in period t for normalized profit, ζ is the intercept, and as the model is in the form of stacked cross-sections, scalars are multiplied by the $N-$dimensional vector of ones ι.

$TL(\mathbf{p}_t, \mathbf{w}_t, t) = \varpi t\iota + \frac{1}{2}\varsigma t^2\iota + \varrho'\mathbf{p}_t + \varphi_t'\mathbf{w}_t + \frac{1}{2}\mathbf{p}_t'\boldsymbol{\Theta}\mathbf{p}_t + \frac{1}{2}\mathbf{w}_t'\boldsymbol{\Gamma}\mathbf{w}_t + \mathbf{p}_t'\boldsymbol{\Psi}\mathbf{w}_t + \kappa'\mathbf{p}_t(t\iota) + \gamma_t'\mathbf{w}_t(t\iota)$ is the translog approximation of the log of the standard profit function. $\mathbf{w}_t$ denotes the vector of stacked cross-sectional observations for the normalized input prices (indexed $g = 1, ..., G$); $\mathbf{p}_t$ is the vector of stacked cross-sectional observations for the output prices (indexed $h = 1, ..., H$); and non-neutral technical change is accounted for by a non-linear time trend and the interactions $\mathbf{p}_t(t\iota)$and $\mathbf{w}_t(t\iota)$. The objects in $TL(\mathbf{p}_t, \mathbf{w}_t, t)$ to be estimated are the ϖ and $\frac{1}{2}\varsigma$ coefficients; the vectors coefficients, ϱ', φ', κ' and γ'; and the matrices of coefficients, $\frac{1}{2}\boldsymbol{\Theta}$, $\frac{1}{2}\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$.

$\mathbf{L}_N$ is the exogenous $(N \times N)$ matrix of non-negative constant linkage weights l_{ij}.[10] $\mathbf{L}_N$ is specified before the estimation of the model and represents: (i) the arrangement of the cross-sectional firms in the network; and (ii) the strength of the linkages between these firms. Since a firm cannot be linked to itself, all the elements on the main diagonal of $\mathbf{L}_N$ are set to zero. As our model is more commonly framed as a spatial frontier, in the spatial context $\mathbf{L}_N\pi_t$ is the spatial lag of the dependent variable, otherwise known as the spatial autoregressive (SAR) variable. The SAR coefficient λ is bounded in the interval $(1/c_{\min}, 1/c_{\max})$, where $c_{\min}$ and $c_{\max}$ are the most negative and most positive real characteristic roots of $\mathbf{L}_N$. In the parlance of a network, we refer to $\mathbf{L}_N\pi_t$ as the network autoregressive (NAR) variable, which we construct by weighting the normalized profit observations of the other firms in the ith firm's network. These observations are weighted according to the specification of $\mathbf{L}_N$, which we discuss in detail in subsection 4.3. Despite our specification of $\mathbf{L}_N$ being exogenous $\mathbf{L}_N\pi_t$ is endogenous, which we account for in the estimation of the model and provide details of how we do so below. Following the spatial non-frontier literature (e.g., Anselin, 2003), since $\mathbf{L}_N\pi_t$ features explicitly in Eq. 11 we refer to this specification as the structural form of the model.[11]

The $\mathbf{L}_N\mathbf{p}_t$ and $\mathbf{L}_N\mathbf{w}_t$ variables are network / spatial lags of $\mathbf{p}_t$ and $\mathbf{w}_t$, and are constructed in the same way as $\mathbf{L}_N\pi_t$. We only include network lags of $\mathbf{p}_t$ and $\mathbf{w}_t$ for parsimony, and because from a behavioral perspective one can argue that a firm does not consider in its decision making network lags of higher order variables and network lags of interactions between first order variables. There

[10] We acknowledge an anonymous reviewer for the following point, which involves noting at the outset that translog frontier models for U.S. banks are typically estimated using large panel data sets that yield parameters with very narrow sampling distributions. With this type of dataset, first order (input and / or output) price variables are often found to have a lot of explanatory power. The squares of these variables and interactions with these variables are also often statistically significant. Here, using the same type of dataset, we investigate the significance of a network effect. To uncover the effect of network spillovers and test their significance, one can estimate a model with no squared or interaction terms, but which includes a variable (or variables) to capture the characteristics of the network. One such variable for U.S. banks is the HHI (e.g., Hirtle, 2007). As we noted in the introductory section, we do not use the HHI to capture the characteristics of the network as it would only relate to the spatial distribution of deposits, which is just one dimension of the spatial correlation between banks' variables. We account for the other dimensions and therefore mitigate the omitted variable bias by introducing a new spatial / network approach to the banking competition literature. This involves creating network lags of variables by pre-multiplying a variable by the bank network linkage matrix, $\mathbf{L}_N$. $\mathbf{L}_N$ though, in contrast to the HHI, is not a variable but a matrix of weights, so it is not possible to uncover the effect of $\mathbf{L}_N$. However, by using a spatial / network method that is explicitly designed to model the range of spatial correlations between banks, for variables and efficiency, we are able to distinguish between the own and network effects.

[11] As will become apparent the NAR variable does not feature explicitly in the reduced form of the model.

is though an important difference between the network dependence that $\mathbf{L}_N\pi_t$ models and $\mathbf{L}_N\mathbf{p}_t$ and $\mathbf{L}_N\mathbf{w}_t$ model. Via the network multiplier matrix, which as we will see features in the reduced form of our model, the NAR variable models endogenous global network dependence (i.e., profit spillovers from a firm's 1st order network, 2nd order network and so on and so forth). $\mathbf{L}_N\mathbf{p}_t$ and $\mathbf{L}_N\mathbf{w}_t$, on the other hand, model exogenous local network effects (i.e., only network spillovers from a firm's 1st order network), where ϕ' and ω' are the associated vectors of local network coefficients.

$\mathbf{L}_N\pi_t$, $\mathbf{L}_N\mathbf{p}_t$ and $\mathbf{L}_N\mathbf{w}_t$ shift the network standard profit frontier technology and therefore reflect the observed heterogeneity of the firms. There are some non-spatial banking studies, however, that estimate purely theoretical functional forms of frontier technologies and do not therefore include variables that shift the frontier (e.g., Wheelock and Wilson, 2012; 2018). In contrast, there is a vast non-spatial banking literature that estimates models that augment theoretical functional forms of frontier technologies with variables that shift the frontier. Some of these shifters are bank level variables, e.g., a variable that profiles bank risk, which would reflect, among other things, a bank's venture capital investments and loans to start-ups. Other shifters capture the macroeconomic and industry environment, e.g., variables that control for monetary policy and the tightening of bank regulation in response to the crisis.

Our model specification has a very interesting feature as it is a hybrid of the above two approaches in the banking literature. This is because it includes variables that shift the frontier, while also having a theoretical network / spatial functional form. We refer to our functional form as the network Durbin model, which is based on the spatial Durbin labelling in the spatial literature (e.g., LeSage and Pace, 2009; Glass *et al.*, 2016a). Our function is an extension of the NAR model, which is Eq. 11 with the local network lagged variables omitted. We retain these local network lagged variables as spatial lags of the exogenous regressors are often found to be important determinants in spatial applications.

By including as shifters spatially weighted neighboring firms' variables that make up their technologies we are following the empirical banking analysis by Tabak *et al.* (2013). Tabak *et al.* estimate a closely related geographically weighted stochastic frontier model for U.S. savings banks, which exploits the spatial correlation between variables that are part of banks' technologies.[12] Specifically, they use geographically weighted observations of these variables for other banks to effectively control for the environment that the *ith* bank faces. As a result, in their model and ours a relatively small number of spatial variables are used to account for a wide range of observed bank heterogeneities, which circumvents the potentially problematic task of choosing what could be a large set of environmental variables. For example, in our model a shift in the frontier technology due to a change in monetary policy would be, together with other impacts, incorporated within the effects of various spatial variables that shift the frontier, namely, the spatially correlated profit, cost of deposits and prices of loans and securities of other banks in the network.

The Boone type analyses of competition in Bolt and Humphrey (2010; 2015a; 2015b) are based on the distribution free approach (DFA) to frontier analysis (Berger, 1993). The DFA involves

[12]We use a different modeling approach to Tabak *et al.* to account for the spatial correlation because, among other things, their model only yields a measure of own efficiency. Our model, however, revolves around the network weights matrix $\mathbf{L}_N$, from which we obtain the network multiplier matrix and in turn the network efficiency spillovers we use in our second stage Boone analysis.

splitting panel data into a number of subperiods and estimating for each subperiod a frontier model with a composed error. This composed error for a profit frontier would be $\mathbf{v}_t - \mathbf{u}_t$, where $\mathbf{v}_t$ is the vector of stacked cross-sectional idiosyncratic errors and $\mathbf{u}_t$ is the vector of stacked time-varying inefficiencies. The DFA then assumes that the average composed error for each firm across the subperiods is an average inefficiency. This is because, on average, the idiosyncratic errors will be close to zero.

Our model has a more complex four component error structure, $\widetilde{\varepsilon}_t = \varepsilon + \varepsilon_t = \xi + \mathbf{v}_t - \delta - \mathbf{u}_t$, where $\varepsilon = \xi - \delta$ and $\varepsilon_t = \mathbf{v}_t - \mathbf{u}_t$ are the time-invariant and time-varying components. δ is the vector of stacked time-invariant inefficiencies due to, for example, rigidities in the internal organization of production and in slow to adjust factors such as fixed assets. ξ is the vector of stacked random effects to account for the unobserved firm heterogeneity that is not part of the observed firm heterogeneities that $\mathbf{L}_N\pi_t$, $\mathbf{L}_N\mathbf{p}_t$ and $\mathbf{L}_N\mathbf{w}_t$ capture.[13] $\mathbf{v}_t$ and $\mathbf{u}_t$ are as above, and the vector of stacked combined time-varying inefficiencies is $\delta + \mathbf{u}_t$.

Notwithstanding the benefit of the DFA, if we were to use the DFA to obtain a set of averages for each efficiency measure, we would substantially reduce the number of efficiency observations to calculate the RED regressors for our second stage econometric analysis. Since in our application we consider GDSI banks, which is a very small subsample of the large bank size category, we guard against a shortage of efficiency observations for our Boone analysis by obtaining annual efficiencies from the first stage. We obtain these efficiencies by using distributional assumptions to distinguish between the error components, which we now discuss in more detail.

As network / spatial stochastic frontier analysis is in its infancy, in line with the starting point for non-spatial stochastic frontier analysis (Aigner *et al.*, 1977), we assume that both δ and $\mathbf{u}_t$ have half-normal distributions. Since our approach is sufficiently general to incorporate other inefficiency distributions, a logical future direction for spatial stochastic frontier modeling would be to follow the evolution of the corresponding non-spatial literature by seeking more appropriate distributional assumptions for inefficiency (e.g., the gamma distribution (Greene, 1980)). As we will see though, our second stage Boone results are intuitive when both inefficiency measures are half-normally distributed.

We test the appropriateness of the error structure in Eq. 11 for our empirical application using the one-sided likelihood ratio test in Gouriéroux *et al.* (1982). The test is for each of the error components and the asymptotic distribution of the test statistic is a mixture of chi-squared distributions, $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$. For $\mathbf{b} \in \{\xi, \mathbf{v}_t, \delta, \mathbf{u}_t\}$, rejection of the null ($\widehat{\sigma}^2_{\mathbf{b}} = 0$) in favor of the alternative hypothesis ($\widehat{\sigma}^2_{\mathbf{b}} > 0$) constitutes evidence of the presence of the component.

The estimation procedure we use is set out in Glass *et al.* (2016b). In particular, it is an extension of the estimator for the pooled SAR stochastic frontier (Glass *et al.*, 2016a) to a setting where unobserved heterogeneity is accounted for and there are time-invariant and time-varying inefficiencies in the same model. As we focus on the development and application of the methodology to test for sources of a change in competition we only outline the procedure we use to

[13] As our pragmatic estimation procedure rests on the error components being independently distributed, we do not use fixed effects to account for unobserved heterogeneity and instead use random effects via an error component. This is because, in contrast to the situation with random effects, the time-varying errors would be correlated with the fixed effects and so these errors would not be independently distributed.

estimate Eq. 11. Due to the complexity of our model because of the presence of the endogenous NAR variable, we use a practically appealing maximum likelihood (ML) estimation procedure known as pseudo ML (PML). This appeal is because our consistent PML estimator simplifies the estimation by breaking it up into a number of steps, which also facilitates model convergence.

Our PML procedure comprises three steps and in each step a log-likelihood function is maximized. In step 1 we transform Eq. 11 into the corresponding non-frontier random effects network Durbin model. Estimating this model distinguishes between the time-invariant and time-varying components of $\widetilde{\varepsilon}_t$ (ε and ε_t). The log-likelihood function for step 1 includes the scaled logged determinant of the Jacobian of the transformation from $\widetilde{\varepsilon}_t^*$ to π_t^* ($T \log |\mathbf{I_N} - \lambda \mathbf{L}_N|$). $\mathbf{I_N}$ denotes the $(N \times N)$ identity matrix; $\widetilde{\varepsilon}_t^*$ and π_t^* are transformations of $\widetilde{\varepsilon}_t$ and π_t into a quasi-differenced form; and, as is standard in the spatial econometrics literature, the transformation from $\widetilde{\varepsilon}_t^*$ to π_t^* accounts for the endogeneity of the NAR variable (e.g., Elhorst, 2009). For details on transforming a model such as Eq. 11 into its non-frontier counterpart and the quasi-differenced transformations of $\widetilde{\varepsilon}_t$ and π_t see Glass *et al.* (2016b).

By maximizing the log-likelihood function for step 2 and using the Battese and Coelli (1988) panel data inefficiency estimator we split the time-varying error from step 1 into its constituent parts. In doing so, we compute the estimate of time-varying inefficiency, $\widehat{\mathbf{u}}_t$. In step 3, using the same approach as in step 2, we split the time-invariant error from step 1 into its two components and thereby obtain the estimate of time-invariant inefficiency, $\widehat{\delta}$. As $\mathbf{u}_t$ and δ form part of our logged model the corresponding efficiencies are $NVE = \exp(-\mathbf{u}_t)$ and $NIE = \exp(-\delta)$, where NVE and NIE denote net time-varying and net time-invariant efficiencies, which signifies that the former is net of time-invariance and the latter is net of time-variance. The combined time-varying efficiency is $GVE = \exp(-\delta - \mathbf{u}_t) = NIE \times NVE$, where GVE denotes gross time-varying efficiency.[14]

The above NIE, NVE and GVE measures from Eq. 11 are the typical type of efficiencies in the stochastic frontier literature as they are bounded in the interval $[0, 1]$. Even though these efficiencies are from a network stochastic frontier that controls for interdependence across the network, they are from the structural form of our model and, as a result, are the typical own efficiencies that relate to a firm in isolation. These efficiencies are therefore net of any network efficiency, i.e., they do not include any efficiency spillovers. From the reduced form of our model we compute, among other things, network efficiency, which is what we turn our attention to next.

4.2 Elasticities and Network Efficiencies

It is now well-established in the spatial literature that the coefficients from a model that contains the SAR variable, which we refer to as the NAR variable, cannot be interpreted as elasticities. This is because the elasticities for the variables are a function of the NAR coefficient. Unlike the simple interpretation of a standard non-network model in log form, to interpret our model we must apply the approach in the spatial literature and compute what we refer to as internal, network and overall elasticities. To a different degree all three of these elasticity measures include

[14] NIE, NVE and GVE should not be confused with the net and gross efficiencies in Coelli *et al.* (1999), as the interpretations of net and gross in their model are entirely different.

a spillover / linkage effect.

An internal elasticity is interpreted in the same way as a conventional elasticity from a non-network model, although an internal elasticity also includes a linkage effect which is due to feedback. This feedback is the effect of the change in a firm's independent variable that partially rebounds back to the firm's dependent variable via its effect on the dependent variables of the other firms in the sample. The network multiplier matrix, which we will discuss in more detail shortly, is at the heart of this feedback and the 'fading memory' property of this matrix across the network is the reason why the feedback is only a partial rebound effect. In contrast to an internal elasticity, a network elasticity is entirely made up of a linkage effect. There are two interpretations of a network elasticity which can be calculated in two ways giving the same value. The first interpretation is the average effect of a change in an independent variable for a firm on the dependent variables of all the other firms in the sample. The second interpretation is the average effect on the dependent variable of a firm due to a change in an independent variable for all the other firms in the sample. Summing the internal and network elasticities gives the overall elasticity.

We calculate the internal, network and overall elasticities and also the corresponding efficiencies from the reduced form of our structural network frontier. Having presented our structural frontier in terms of vectors of stacked cross-sections we can move directly to the reduced form in Eq. 12. We simply take $\lambda \mathbf{L}_N \pi_t$ in Eq. 11 to the left-hand side to obtain $\pi_t - \lambda \mathbf{L}_N \pi_t = (\mathbf{I_N} - \lambda \mathbf{L}_N)\,\pi_t$ and then divide throughout by $(\mathbf{I_N} - \lambda \mathbf{L}_N)$.

$$\pi_t = \ (\mathbf{I_N} - \lambda \mathbf{L}_N)^{-1} \begin{pmatrix} \zeta\iota + TL(\mathbf{p}_t, \mathbf{w}_t, t) + \phi' \mathbf{L}_N \mathbf{p}_t + \omega' \mathbf{L}_N \mathbf{w}_t + \\ \xi + \mathbf{v}_t - \delta - \mathbf{u}_t \end{pmatrix}, \tag{12}$$

where $(\mathbf{I_N} - \lambda \mathbf{L}_N)^{-1}$ plays a key role and denotes the network multiplier matrix that we referred to earlier, while everything else is as previously defined. Note that in line with the spatial literature, we refer to Eq. 12 as the reduced form because, in contrast to the structural form, the NAR variable does not explicitly feature (Anselin, 2003).

Although the general approach to calculate elasticities from Eq. 12 is standard in the spatial literature this is not the case in the OR literature. We therefore outline the approach to compute internal, network and overall elasticities in the OR context of our network profit frontier, which we demonstrate for an output price, $p_{h,t}$, at the sample mean. As is also the case with a simple non-network translog model, in our network setting when the data is mean adjusted the higher order and interaction terms in our model are zero at the sample mean. The coefficients on the higher order and interaction variables do not therefore feature in the calculation of the internal, network and overall elasticities for a variable at the sample mean.

Differentiating Eq. 12 with respect to $p_{h,t}$ yields the two matrices on the right-hand side of Eq. 13, which are independent of the time index.

$$\boldsymbol{\Upsilon} = \begin{bmatrix} \frac{\partial \pi_1}{\partial p_{h,1}} & \cdots & \frac{\partial \pi_1}{\partial p_{h,N}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \pi_N}{\partial p_{h,1}} & \cdots & \frac{\partial \pi_N}{\partial p_{h,N}} \end{bmatrix}_t = (\mathbf{I_N} - \lambda \mathbf{L}_N)^{-1} \begin{bmatrix} \varrho_h & \cdots & l_{1N}\phi_h \\ \vdots & \ddots & \vdots \\ l_{N1}\phi_h & \cdots & \varrho_h \end{bmatrix}. \tag{13}$$

The product of these matrices is the matrix $\boldsymbol{\Upsilon}$, which comprises internal elasticities on the main diagonal, while the off-diagonal elements are asymmetric network elasticities for each pair of firms. Following the approach in the spatial literature, we facilitate interpretation of this large number of elasticities by calculating mean elasticities (internal, network and overall). The mean internal elasticity is the average of the diagonal elements and the mean network elasticity is the mean row (column) sum of the off-diagonal elements. Although the two ways of calculating the mean network elasticity are numerically the same, they are interpreted as the mean spillover elasticity to (from) a firm from (to) all the other firms in the sample. We compute the $t-$statistics for the mean elasticities via Monte Carlo simulation of their distributions.

For each of the conventional own efficiency measures from the structural form of our model, we compute the corresponding efficiencies from the reduced form model, which in line with the reduced form elasticities, we refer to as internal, network and overall NIE, NVE and GVE. To provide the necessary intuition we define these efficiencies. As is the case with internal, network and overall elasticities, to different degrees the corresponding efficiencies all include a spillover / linkage efficiency effect. Internal efficiency is interpreted in the same way as own efficiency from the structural form of our model or a conventional non-network frontier. That said, internal efficiency is own efficiency plus the linkage efficiency we label as efficiency feedback. An example of efficiency feedback is the effect of a change in a firm's independent variable which affects the dependent variables and hence efficiencies of the other firms in its 1st and higher order networks. Through the network multiplier matrix this effect partially rebounds back to the dependent variable and thus efficiency of the firm that initiated the process.

In line with the network elasticity of a variable, network efficiency is entirely a spillover / linkage efficiency effect and there are two ways of calculating network efficiency which have different interpretations. The first measure is the sum of the efficiency spillovers that gravitate to a firm from all the other firms in the sample. The second measure is the sum of the efficiency spillovers that are transmitted in the opposite direction from a firm to all the other firms. Recall that asymmetric network elasticities for a variable are obtained for pairs of firms and averaging the row or column sums of these elasticities yields the same numerical value for the mean network elasticity. Similarly, the network efficiencies relating to each pair of firms are asymmetric, but the sample averages of the two measures of network efficiency will be equal. This has implications for overall efficiency because along the same lines as the overall elasticity of a variable is calculated, overall efficiency for a firm is the sum of its internal and network efficiencies. Thus, for individual firms asymmetric network efficiencies lead to asymmetric overall efficiencies, but the sample means of these two measures of overall efficiency will be equal. It is important to note, however, that in the empirical application we use the sum of the network efficiencies that a firm receives together with the resulting corresponding overall efficiency. This is because with these measures the direction of travel of the efficiency spillovers is appropriate for a firm level analysis of the sources of a change in competition.

In line with conventional own efficiencies from the non-network stochastic frontier literature and the structural form of our frontier, the lower bound of the internal, network and overall net and gross efficiencies is of course 0. Rather than having an upper bound of 1 like conventional own efficiencies, the internal, network and overall net and gross efficiencies are not bounded from above. This is entirely intentional because we calculate absolute internal, network and overall net and gross efficiencies to ascertain whether the network efficiency spillovers are substantive. In contrast, network efficiencies that are relative to a benchmark may not be substantive because all the absolute values that are used to compute such efficiencies may be similarly small in magnitude. It should be emphasized though that the absence of an upper bound for the internal, network and overall net and gross efficiencies in no way impinges on their interpretation as they are all percentages. They are percentages because they are scaled conventional own net and gross efficiencies from the structural form of our network stochastic frontier. The magnitude of this scaling reflects the size of the linkage effect that partially / entirely makes up the internal, network or overall efficiency. By not placing an upper bound on the internal, network and overall net and gross efficiencies these efficiencies are relative to the relevant conventional own NIE, NVE or GVE benchmark. A particularly interesting situation is where the linkage effect is sufficiently large to yield an internal, network or overall efficiency score greater than 1. In this case, the linkage effect has pushed the firm beyond the best practice frontier for the relevant conventional own efficiency from the structural form of our frontier in Eq. 11.

For the technical details on how we calculate the internal, network and overall efficiencies in a spatial setting, which in our application we frame as a network, see Glass *et al.* (2016a; 2016b). Throughout our application we use Int, Nwk and Ov to denote the internal, network and overall net and gross efficiencies.

4.3 Data, Network Linkages and Competitive Regimes

In summary, the first stage involves estimating the network stochastic standard profit frontier and in the second stage we divide the first stage sample up into a number of pairs of subsamples, where each pair represents two competitive regimes. Each pair of competitive regimes represents a competition case and for each case PQR is used to test whether we can reject the null that competition is the same between the two subsamples. Our sample to estimate the frontier model comprises 183 large U.S. banks over the period $1994-15$. This sample is a balanced panel and thus comprises the core surviving large banks, so that in the second stage we can make like-for-like comparisons as the same banks (or least the same number of banks) are in a given pair of subsamples. Following Berger and Roman (2016), a bank is classified as large if it has total assets greater than \$3 billion in 2015.

Spatial / network analysis of bank efficiency using specifically designed spatial / network methods like we employ (i.e., those that are based on the network linkage matrix $\mathbf{L}_N$) is very much in its infancy. This is evident as there is just one other study that applies this type of method to banks to estimate a cost frontier (Glass and Kenjegalieva, 2019), whereas we consider a profit frontier. Given the paucity of banking studies that use the type of methods we employ, we remain consistent with the early evolution of the banking production literature as we base our choice of inputs and outputs on the intermediation approach (Sealey and Lindley, 1977). This approach assumes that banks use the savings of customers as inputs to make investments which represent their outputs. Our study therefore provides a platform for further work to continue tracking the evolution of the bank production literature by, for example, using the more recent value-added approach (Berger and Humphrey, 1992) to choose the inputs and outputs.

Whilst the intermediation approach adopts a mutually exclusive classification of inputs and outputs, there is no such clear distinction with the value-added approach as it considers all asset and liability categories to have some output characteristics. With the value-added approach the classification of inputs and outputs is based on whether the magnitude of the value-added suggests that the category is more like an input or more like an output. Differences arise between the classification of inputs and outputs using the two approaches. For example, current deposits would be an input using the intermediation approach, but using the value-added approach Berger and Humphrey (1992) find that current deposits is clearly an output, which is because there are bank fees associated with these deposits that do not involve further interest payments. Notwithstanding the emergence of the value-added approach, in line with Glass and Kenjegalieva (2019) our choice of inputs and outputs follows a U.S. banking paper by Koetter *et al.* (2012) that uses the intermediation approach.

Our network frontier and the PQR models are at the bank level. As a result, in the second stage we conduct a bank level Boone analysis of competition. Strictly speaking competition is defined in terms of the relevant market, which for U.S. banks is typically taken to be the MSA or non-MSA county. Although in our network frontier we specify the network linkages in $\mathbf{L}_N$ using branch location data, the other data we need for this model and the data for the RPDs in the PQR models is not available at the branch level. In densely populated metropolitan areas such as New York competition is likely to be fierce between very large banks with little opportunity for high margins. In terms of the Boone theory in figure 1, for a given RED, a bank's RPD in this type of market will be much lower than the RPD for a much smaller bank in a rural non-MSA county where there is far less competition leading to higher lending rates. Our bank level Boone analysis can therefore be viewed as collectively accounting for all the different RED-RPD relationships across the different markets that banks operate in.[15]

We obtain the majority of our data from the Reports of Condition and Income (i.e., the Call Reports), which were sourced from the Federal Deposit Insurance Corporation (FDIC). In table 1 we provide summary statistics for the data for our network stochastic standard profit frontier and, for the purposes of comparison, the second stage data for the same period. From table 1 we can see for the frontier model that the three input prices relate to the cost of fixed assets (w_1), labor (w_2) and deposits (w_3), and the output prices relate to the three lending and non-lending activities of banks: loans (p_1), securities (p_2) and non-interest income (p_3). To avoid endogenous RED measures in the second stage we estimate the frontier model and construct the dependent RPD variable for each of the PQR models using different measures of profit. For our frontier model the measure of profit is net operating income (π), which is deflated to 2005 prices using the CPI. The input and output prices are not deflated because, as is evident from table 1, they are ratios. All the deflated profit and input and output price data is first logged, then mean adjusted and, finally, to be consistent with the estimation of a translog cost function, we use one of the

[15] We thank an anonymous reviewer for prompting us to explain the relationship between the relevant markets and our bank level analysis of competition.

Table 1: Description of the variables and summary statistics

Variable description	Mean	Std. dev.
Network Stochastic Standard Profit Frontier (first stage)		
Real net operating income (thousands of U.S. dollars) (π)	$27,164,232$	$126,001,222$
Price of loans: Interest income from loans divided by loans and leases (p_1)	0.065	0.020
Price of securities: Interest income from securities divided by total securities (p_2)	0.325	17.793
Price of other activities: Approximated by total non-interest income divided by total assets (p_3)	0.016	0.040
Cost of fixed assets: Expenditure on fixed assets divided by the sum of the value of premises and fixed assets (w_1)	1.161	21.043
Cost of labor: Salaries divided by number of full-time equivalent total employees (w_2)	59.955	20.402
Cost of deposits: Interest expenses on deposits divided by total deposits (w_3)	0.020	0.013
Number of bank branches used to specify $\mathbf{L}_N$	181	590
Intensity of a bank's branch network in $\mathbf{L}_N$	7.972	9.412
Polynomial Quantile Regressions (second stage)		
Return on assets (ROA)	1.145	1.222

input prices, which in our analysis is w_1, as the normalizing factor for π and the other input prices. By mean adjusting the data all the first order internal, network and overall parameters from the reduced form of our frontier model can be interpreted as elasticities at the sample mean. This is because at the sample mean the terms in the partial derivatives of a translog function that relate to the higher order and interaction terms are zero. For each of the PQR models the dependent RPD variable is constructed using the return on assets (ROA) as the measure of profitability.

Using FDIC data from the Summary of Deposits on the state locations of each bank's branches, we specify $\mathbf{L}_N$ using the following four steps.

(i) Begin by specifying an $\mathbf{L}_N$ for each year by setting all the cells on the main diagonals to zero because a bank cannot be linked to itself.

(ii) (a) For each state where the ith bank operates, calculate the ratio of the number of jth bank branches to the number of ith bank branches.[16] (b) Where the ith bank operates, sum these ratios to obtain the non-zero off-diagonal elements. (c) Set all the other off-diagonal elements to zero to signify that the ith and jth banks' branch networks do not overlap.

(iii) Average the annual specifications of $\mathbf{L}_N$ from (ii). Also, recall that exogenous linkage weights is an underlying assumption of our network frontier model. As the off-diagonal elements of this average $\mathbf{L}_N$ are calculated at the micro level of branch networks across states and the variables are at the level of the banking firm, based on parallels with firm level studies in the spatial literature that draw on more disaggregated plant level information, it is reasonable to take the linkage weights to be exogenous. See table 1 for summary statistics on the number of bank branches and the off-diagonal elements. Each of these off-diagonal elements is interpreted as a bank's branch network intensity vis-à-vis the network of another bank.

(iv) We obtain the $\mathbf{L}_N$ we use in the estimation by normalizing the elements of the average $\mathbf{L}_N$

[16] We calculate the ratios at the state level as the number of calculations needed to obtain the ratios at the county or city level exceeded the limit in the Stata software.

from (iii) by dividing throughout by the largest cell. This is referred to in the spatial literature as normalizing by the largest eigenvalue. The benefit of this normalization is that it retains the information on the absolute intensities of the banks' branch networks as it does not change the proportional relationship between the linkage weights.[17]

Figure 2 presents the average state level branch distribution for GDSI banks over our study period, where the larger the area of a pie chart, the greater the total number of branches in the state.[18] It is clear from this figure that California has the largest number of branches and, although less obvious, Florida has the second largest number. Compared to these two states we can also see from this figure that there are a number of states with far fewer branches. Despite the removal of state branching restrictions, U.S. bank branching remains highly geographically concentrated. This is of course less marked for larger banks, but even for our sample there is quite a lot of evidence of geographically concentrated branch networks. To illustrate, the median bank in our sample in terms of total assets in 2015 only operates in two states. As figure 2 is for GDSI banks, we can see that some of the largest banks in our sample also have regionally oriented branch networks, although the geographical concentration of these networks varies. For example, we can see from this figure that the Bank of New York Mellon has a highly geographically concentrated branch network. In contrast, we can also see that some of the largest banks (Wells Fargo, Bank of America and JPMorgan Chase) have geographically dispersed branch networks because they are truly multi-market banks. Hannan and Prager (2004; 2009) and Rosen (2007) note the importance of accounting for this multi-market presence when examining changes in competition and how we account for this presence is what we turn to next.[19]

Figure 2 only presents the geographical branch distributions of GDSI banks because in a figure like this it is only possible to illustrate such distributions for a limited number of banks. Our specification of $\mathbf{L}_N$ though measures the strength of the geographical links between the branch networks of each pair of banks in the sample. $\mathbf{L}_N$ therefore accounts for the different geographical markets that banks operate in and the differences between their branch presence in these markets. We illustrate how differences in the geographical concentrations of banks' branch networks are accounted for by $\mathbf{L}_N$ by considering the geographical links between the GDSI bank Comerica and other banks in the sample. We therefore consider in the $\mathbf{L}_N$ matrix the off-diagonal $ith - jth$ cells in the ith row that relates to Comerica. In this row an off-diagonal cell is set to zero if the jth bank's branch network does not overlap with Comerica's (in other words, when the jth bank operates in states other than Arizona, California, Florida, Michigan and Texas). All the other off-diagonal cells in this row are non-zero and reflect the degree of overlap between the branch networks of the jth bank and Comerica. In particular, the non-zero off-diagonal cells

[17] It is common in the spatial literature to normalize the weights matrix by its row sums. This is intended for binary spatial weights that reflect, for example, contiguous neighboring regions, which is a different setting to the one we consider. We do not therefore pursue this any further because if we row-normalized our non-binary weights we would lose the information on the absolute intensities. This information would be replaced with relative intensities which would be difficult to interpret.

[18] GDSI banks in this paper consist of the globally systemically important banks, as defined by the Financial Stability Board (2015), and the domestically systemically important ones identified by the Board of Governors of the Federal Reserve System (2014).

[19] We thank an anonymous reviewer for prompting us to highlight how we account for the effect of this multi-market presence in the modeling.

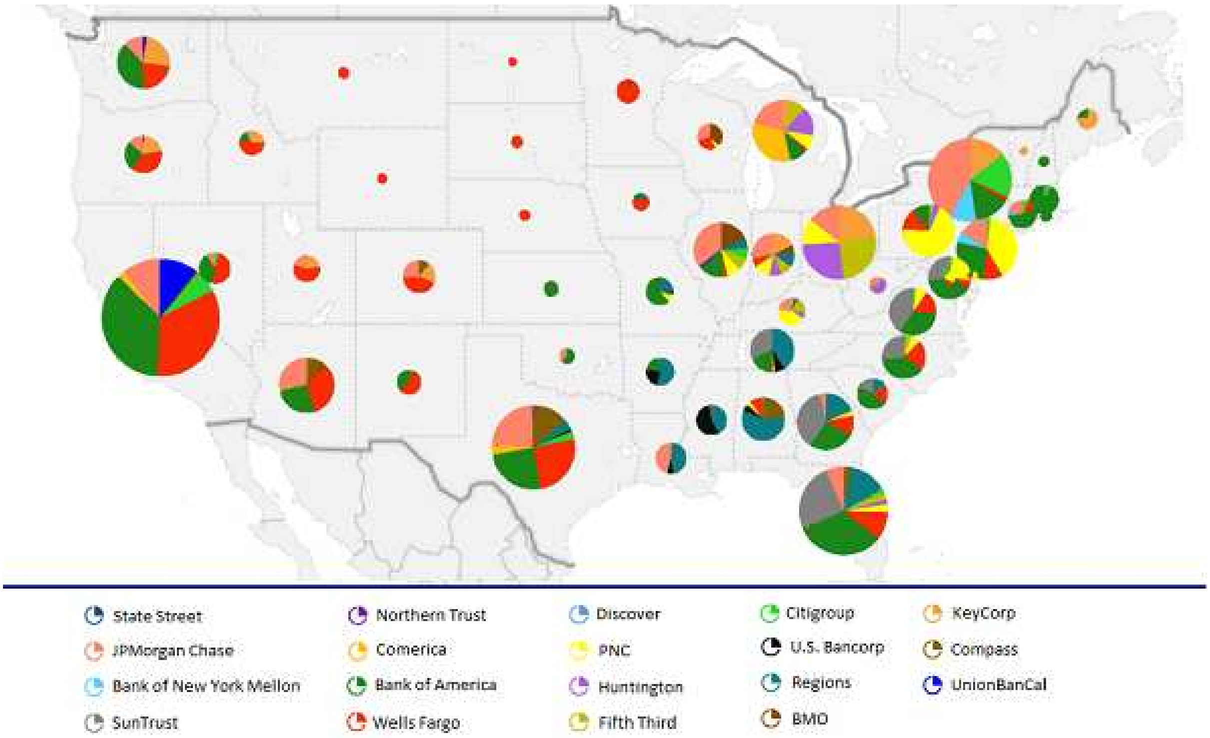


Figure 2: Average state level branch distribution for systemically important U.S. banks

represent the sum across the states of the ratio of the number of jth bank branches in a state to the number of ith bank branches. For example, we can see from figure 2 that Comerica has more branches in Michigan than the Bank of America, while the Bank of America has more branches in Texas than Comerica. These state differences in the number of branches are accounted for in the sum of the state ratios of the number of Bank of America (Comerica) branches to the number of Comerica (Bank of America) branches. This is because the ratio of the number of Bank of America (Comerica) branches to the number of Comerica (Bank of America) branches in Michigan is lower (higher) than the ratio for Texas.

For each of the competition cases we consider in our second stage analysis we test for sources of a change in competition between two regimes. Based on the nature of the competition cases we split our second stage into three parts. The competition cases are described below, where the two regimes for each case relate to different time periods, although this is not a requirement for empirical Boone testing. Different competitive regimes can instead be contemporaneous because they relate to, for example, firms in different industries, different sizes of firms within an industry, or firms located in different geographical areas.

Part I of the Second Stage: Pre-Crisis Period versus the Crisis Period and Beyond

- **Case 1** is for all 183 banks and the two competitive regimes are (A) the pre-crisis period (1994 – 2007) and (B) the crisis period and beyond (2008 – 2015). According to Berger and Bouwman (2013) the timeline of the U.S. subprime lending crisis spanned the period 2007:Q3 – 2009:Q4. As they regard the majority of 2007 as being pre-crisis and given we use annual data, we take 2007 as being the last year of the first competitive regime for this case.

- **Cases 2 & 3** are for GDSI banks and non-GDSI banks, respectively, and the two competitive regimes, (A) and (B), are as in case 1. GDSI banks for both competitive regimes are as defined in figure 2 and the non-GDSI banks are the remaining banks in the sample.

Part II of the Second Stage: Phases of the Business Cycle

- We find for cases $1-3$ evidence of more intense competition in regime (B). Related to this, we would expect an increase in competition during the recession phase of a business cycle and a decrease in the expansion phase. Over the course of a business cycle, it is possible therefore that the decrease in competition we associate with the expansion cancels out the increase that we associate with the recession. If this is indeed the case for the period $2008-15$, since we know for cases $1-3$ that the increase in competition is exogenous due to crisis induced phenomena, we can infer that this increase does not originate from the recession triggered by the crisis. This would instead suggest that the increase in competition was due to the combined effect of two other crisis induced phenomena- the collateral damage to banks from monetary policy to moderate the recession that the crisis initiated, and the post-crisis tightening of bank regulation to reduce risk. For the subperiods in part I, we examine if the increase in competition we associate with the recession (i.e., the first portion of each subperiod) is cancelled out by the decrease that we associate with the expansion (i.e., the second portion of each subperiod). **Case 4** is for all 183 banks and the two competitive regimes are (A) the 2001 recession and (B) the $2002-07$ expansion.[20]

- **Case 5** is for non-GDSI banks and the two competitive regimes, (A) and (B), are as in case 4.[21]

- **Case 6** is for all 183 banks and is for the recession and expansion phases of another business cycle, albeit an incomplete one. The two competitive regimes are (A) the complete $2008-09$ recession and (B) the incomplete $2010-15$ expansion.[22]

- **Cases 7 & 8** are for GDSI and non-GDSI banks, respectively, and the two competitive regimes, (A) and (B), are as in case 6.[23]

Part III of the Second Stage: Further Subperiods and Bank Classifications

[20] The years for the phases of the business cycles are based on dates from the National Bureau of Economic Research (NBER). The NBER dates the phases using quarters which we transform to years using our judgement. For the complete business cycle in our pre-crisis period the NBER dates for the recession and expansion are 2001:Q1 – 2001:Q4 and 2002:Q1 – 2007:Q4. In years this equates nicely to a recession in 2001 and a $2002-07$ expansion.

[21] Recall from the introductory section that we do not analyze GDSI banks over the business cycle in the pre-crisis period. This is because there are insufficient observations, as the recession in this cycle lasts for only one year and there are just 19 GDSI banks.

[22] For this incomplete business cycle, the NBER finds that 2007:Q4 – 2009:Q2 is the complete recession phase. In years this is approximated to be $2008-09$. According to the NBER the peak of the subsequent expansion is not reached in our study period. We therefore consider an incomplete expansion phase from $2010-15$.

[23] Although, as will become apparent, the findings for case 7 are entirely reasonable, the recession regime PQR model for GDSI banks is based on a limited number of observations as the recession spans two years.

- As cases $1-3$ consider two long subperiods in **cases 9-17** the two competitive regimes, (A) and (B), are successive shorter subperiods. The four distinct subperiods we consider are $1994-00$ (before the build-up to the crisis began); $2001-06$ (includes the build-up to the crisis); $2007-09$ (crisis period); and $2010-15$ (post-crisis period). Also, since case 2 is for GDSI banks, which are limited in number and fixed over our study period according to a 2014 / 15 classification, three of cases $9-17$ are based on a further classification that leads to a bigger group of the largest banks in the sample that changes annually. In particular, this bigger subsample comprises banks with total assets in the top third of the sample in each year. A further three cases are for the remaining two-thirds in each year and three cases are for the entire sample of banks.

4.4 Estimated Network Stochastic Standard Profit Frontier

The estimated structural form of our network stochastic standard profit frontier model is presented in table 2.[24] How well-specified this model is depends on, among other things, whether there is an endogeneity issue. To examine this we use a Hausman-Wu test, which has been used to test for endogeneity by Adams *et al.* (1999) and Glass *et al.* (2013) in the non-network stochastic frontier literature. This involves comparing our maximum likelihood (ML) model when the NAR variable is omitted with the corresponding model when the Hausman-Taylor estimator is used to account for the possibility that any of the remaining variables are endogenous by using instrumental variables. We need not be concerned about dropping the NAR variable from the models we use in this test because this is the only variable where our ML estimator accounts for the endogeneity. Whether the other variables in our model are endogenous is what we are interested in and test for. We cannot reject the null hypothesis that there is no difference between the sets of coefficients from the two models at the 10% level, which suggests that our ML model is robust to endogeneity concerns.

The goodness of fit is a further indicator of how well-specified our model is, which we measure using the adjusted R^2. The adjusted R^2 for our fitted model is 0.87, which is sufficiently high to suggest that our model is well-specified. Bolt and Humphrey (2015b) note though that if the goodness of fit is high, it is more likely that a change in competition is not economically substantive because it is based on a small absolute inefficiency. We suggest this is not the case with our fitted model because when we test for the presence in the error structure of the net time-invariant and net time-varying inefficiencies, we find that both inefficiencies are statistically significantly different from zero. We discuss these test results in more detail in the next subsection when we turn our attention to the efficiencies.

None of the reported structural coefficients in table 2 can be interpreted because the variable elasticities are also a function of λ, which is the coefficient on $\mathbf{L}\pi$. For models such as the one we estimate it is well-established in the spatial literature that it is the reduced form of the structural model that yields interpretable parameters, which we refer to as internal, network and overall

[24] As is standard in non-frontier and frontier random effects models, the $\log_{10} \varkappa$ parameter we report in table 2 is the weight that is attached to the cross-sectional component of the panel data. From table 2 we can see that the fitted $\log_{10} \varkappa$ parameter is highly significant at the 0.1% level.

Table 2: Estimated network stochastic standard profit frontier model

$\mathbf{p}_1$	-0.387^{***}	$\mathbf{p}_2\mathbf{p}_3$	0.023^{***}	$t\iota$	-0.037^{***}	$\mathbf{L}p_3$	-0.057
$\mathbf{p}_2$	-0.127^{***}	$\mathbf{w}_2^2$	-0.017^{**}	$t^2\iota$	0.001^{***}	$\mathbf{L}w_2$	-0.066
$\mathbf{p}_3$	-0.025^{***}	$\mathbf{w}_3^2$	0.020^{***}	$\mathbf{p}_1\,(t\iota)$	0.006^{*}	$\mathbf{L}w_3$	0.026
$\mathbf{w}_2$	0.692^{***}	$\mathbf{w}_2\mathbf{w}_3$	-0.023^{**}	$\mathbf{p}_2\,(t\iota)$	0.004^{*}	$\mathbf{L}\pi$	0.253^{***}
$\mathbf{w}_3$	0.196^{***}	$\mathbf{p}_1\mathbf{w}_2$	0.024	$\mathbf{p}_3\,(t\iota)$	-0.001^{*}	$\log_{10}\varkappa$	-2.153^{***}
$\mathbf{p}_1^2$	-0.080^{***}	$\mathbf{p}_1\mathbf{w}_3$	-0.043^{***}	$\mathbf{w}_2\,(t\iota)$	0.002	σ_v	0.101 (0.003)
$\mathbf{p}_2^2$	0.012^{***}	$\mathbf{p}_2\mathbf{w}_2$	-0.023^{*}	$\mathbf{w}_3\,(t\iota)$	-0.004^{***}	σ_u	0.069 (0.009)
$\mathbf{p}_3^2$	-0.012^{***}	$\mathbf{p}_2\mathbf{w}_3$	-0.035^{***}	$\zeta\iota$	-0.041	σ_δ	0.143 (0.006)
$\mathbf{p}_1\mathbf{p}_2$	0.125^{***}	$\mathbf{p}_3\mathbf{w}_2$	0.010	$\mathbf{L}\mathbf{p}_1$	-0.345^{***}	σ_ξ	0.292 (0.010)
$\mathbf{p}_1\mathbf{p}_3$	0.039^{***}	$\mathbf{p}_3\mathbf{w}_3$	-0.030^{***}	$\mathbf{L}\mathbf{p}_2$	0.087	LL	2711.22

Notes: *, ** and *** denote statistical significance at the 5%, 1% and 0.1% levels, respectively; LL denotes log-likelihood; and standard errors are in parentheses.

parameters. We present and comment on these internal, network and overall parameter estimates further in this discussion of our fitted network frontier. Notwithstanding that the structural coefficients from our fitted model cannot be interpreted, it should be noted that although it may appear that the signs of the first order input and output price coefficients in table 2 are inconsistent with production theory, this is not the case for reasons we will explain when we present the internal, network and overall parameters.

For some time in a spatial context the SAR parameter, which is the NAR coefficient λ in our setting, was interpreted as the elasticity of the spatial lag of the dependent variable, or, in other words, as a spillover elasticity. It is now well-established that λ cannot be interpreted as a spillover elasticity and that the spillover parameters in our setting are the network parameters, which, as we have seen in Eq. 13, are a function of, among other things, λ. The estimate of λ, however, does provide useful information on the degree of NAR dependence across the banks. The estimate of λ in table 2 is 0.253, which is significant at the 0.1% level. In the context of the magnitude of the corresponding estimates in the large empirical spatial literature, a significant λ estimate of 0.253 is interpreted as substantial positive NAR dependence between the banks. This estimate therefore provides support for our network approach to profit modelling for large U.S. banks.

Turning to the coefficients on the local network variables (i.e., the network lags of the first order output and input price variables), table 2 indicates that only one is significant, which is the $\mathbf{L}p_1$ parameter at the 0.1% level. This $\mathbf{L}p_1$ parameter, however, is large, which provides some support for our network Durbin model specification over the NAR model, as the latter omits local network variables. Moreover, this $\mathbf{L}p_1$ parameter is negative, which suggests that, on average, there is a negative correlation between the *ith* bank's profit and our weighted average of the p_1 observations of the *jth* banks that have a branch network that overlaps with that of the *ith* bank. Since p_1 is the price of loans (i.e., interest income as a share of loans) and a negative coefficient on a network lag of a dependent / independent variable is attributed to the effects of competition (Kao and Bera, 2013), we interpret the negative coefficient on $\mathbf{L}p_1$ as evidence of loan competition between banks with overlapping branch networks. This is the type of strategic

Table 3: Estimated internal, network and overall parameters

	Internal	Network	Overall		Internal	Network	Overall
$\mathbf{p}_1$	−0.387***	−0.0413***	−0.428***	$\mathbf{p}_1\mathbf{w}_2$	0.026	0.0006	0.026
$\mathbf{p}_2$	−0.127***	0.0059	−0.121***	$\mathbf{p}_1\mathbf{w}_3$	−0.045***	−0.0011**	−0.046***
$\mathbf{p}_3$	−0.025***	−0.0056	−0.030***	$\mathbf{p}_2\mathbf{w}_2$	−0.025**	−0.0006	−0.025**
$\mathbf{w}_2$	0.692***	0.0104**	0.702***	$\mathbf{p}_2\mathbf{w}_3$	−0.034***	−0.0008**	−0.035***
$\mathbf{w}_3$	0.196***	0.0068***	0.203***	$\mathbf{p}_3\mathbf{w}_2$	0.010	0.0002	0.010
$\mathbf{p}_1^2$	−0.077***	−0.0018***	−0.079***	$\mathbf{p}_3\mathbf{w}_3$	−0.030***	−0.0007***	−0.031***
$\mathbf{p}_2^2$	0.012***	0.0003**	0.012***	$t\iota$	−0.036***	−0.0009***	−0.037***
$\mathbf{p}_3^2$	−0.012***	−0.0003***	−0.012***	$t^2\iota$	0.001***	0.0000***	0.001***
$\mathbf{p}_1\mathbf{p}_2$	0.123***	0.0029***	0.125***	$\mathbf{p}_1\,(t\iota)$	0.006*	0.0001	0.006*
$\mathbf{p}_1\mathbf{p}_3$	0.039***	0.0009***	0.040***	$\mathbf{p}_2\,(t\iota)$	0.004*	0.0001	0.004*
$\mathbf{p}_2\mathbf{p}_3$	0.023***	0.0006**	0.024***	$\mathbf{p}_3\,(t\iota)$	−0.001*	0.0000	−0.002*
$\mathbf{w}_2^2$	−0.017**	−0.0004*	−0.017**	$\mathbf{w}_2\,(t\iota)$	0.002	0.0001	0.002
$\mathbf{w}_3^2$	0.020***	0.0005***	0.020***	$\mathbf{w}_3\,(t\iota)$	−0.004***	−0.0001**	−0.004***
$\mathbf{w}_2\mathbf{w}_3$	−0.023**	−0.0006*	−0.024**				

Notes: *, ** and *** denote statistical significance at the 5%, 1% and 0.1% levels, respectively.

behavior between banks that we would expect to observe, which supports our weighting of the observations of the jth banks that have a branch network that overlaps with that of the ith bank (i.e., our specification of $\mathbf{L}$).

Consider for the moment a non-spatial / network translog standard profit frontier. If mean adjusted data is used to estimate such a model, the coefficients on the first order output and input prices can be interpreted as elasticities at the sample mean. Production theory posits that profit is monotonically increasing (decreasing) in output (input) prices. By normalizing by an input price, the signs of the coefficients on the first order output (input) prices that production theory predicts for a fitted non-network standard profit frontier switch from positive (negative) to negative (positive).

In table 3 we present the internal, network and overall parameters from the reduced form of our structural network standard profit frontier. The interpretation of the internal parameters from the reduced form of our model is the same as the interpretation of the parameters from a non-network standard profit stochastic frontier. The monotonicity properties of a non-network standard profit function also therefore apply to the internal parameters from our model. From table 3 we can see that all the first order internal output and input price parameters are negative and positive, respectively. Thus, in line with production theory, we conclude at the sample mean that our model satisfies the monotonicity property of the translog standard profit function. Additionally, all the reported first order internal output and input price parameters are significant at the 0.1% level.

A network elasticity measures the magnitude of the spillover of a variable. In our case the direction of the spillover is to an individual bank from all the other banks in the sample, and an overall elasticity for a variable is the sum of the internal and network elasticities. Whereas production theory provides expected signs for the internal output and input price elasticities from our network standard profit frontier, this theory does not suggest expected signs for the network and overall output and input price elasticities. From table 3 we can see that a number of the network parameters are significant at the 5% level or lower (e.g., the parameters for the $\mathbf{p}_1$, $\mathbf{w}_2$ and $\mathbf{w}_3$ variables), which constitutes further support for our network approach to profit modeling for large U.S. banks.

Interestingly, despite the NAR coefficient in table 2 being significant, large and positive, the absolute magnitude of the largest network parameter for a first order output / input price in table 3 is quite small. The network parameters for $\mathbf{p}_2$, $\mathbf{p}_3$, $\mathbf{w}_2$ and $\mathbf{w}_3$ in table 3 are small because in the calculation of these parameters, the small coefficients on their network lags in table 2 (i.e., $\mathbf{Lp}_2$, $\mathbf{Lp}_3$, $\mathbf{Lw}_2$ and $\mathbf{Lw}_3$) dominate the large NAR parameter. To illustrate, as we noted above, we interpret the large negative coefficient on the network lag of the price of loans ($\mathbf{Lp}_1$) in table 2 as evidence of loan competition between banks with overlapping branch networks. To obtain the quite small negative network parameter for $\mathbf{p}_1$ in table 3, we can infer in the calculation of this parameter that the coefficient on the network lag of $\mathbf{p}_1$ is slightly more dominant than the large positive NAR coefficient. The latter reflects the positive profit dependence between banks with overlapping branch networks, as, among other things, they face the same local and regional market conditions.

4.5 Discussion of the Efficiency Estimates

In our fitted structural model we find evidence of own net time-invariant and own net time-varying inefficiencies (δ and $\mathbf{u}_t$, respectively) that are significantly different from zero, as we reject $H_0 : \widehat{\sigma}^2_\delta = 0$ and $H_0 : \widehat{\sigma}^2_{\mathbf{u}_t} = 0$ at the 1% level. For the entire sample, GDSI banks and the remaining banks we summarize in table 4 the own NVE, NIE and GVE scores from this model. Recall that even though the structural form of our model accounts for global NAR dependence (i.e., spillovers from a bank's 1st order network, 2nd order network and so on and so forth) and also some local network dependencies (i.e., spillovers from a bank's 1st order network only), the own efficiencies we report contain no form of efficiency spillover and can therefore be interpreted in the same way as efficiencies from a standard non-network frontier. In contrast, the internal, network and overall NVE, NIE and GVE estimates from the reduced form of our model are partially / entirely made up of an efficiency spillover. As GVE represents a more complete picture of economic performance than NVE and NIE, we also summarize in table 4 the internal, network and overall GVE scores from the reduced form of our model.

Recall also that any GVE measure (i.e., own, internal, network or overall GVE) is obtained by multiplying the corresponding NVE and NIE scores. We can see therefore from the mean own efficiencies in table 4 for the entire sample and the two subsamples that NVE is a much bigger contributor to GVE than NIE. This indicates that the vast majority of mean own gross time-varying inefficiency is a persistent phenomenon. Various conjectures can be made to try and explain this persistent underperformance, including the possibility of a continual quiet life culture across large banks (Koetter *et al.*, 2012). The mean own NIE for GDSI banks and, as a result, the mean own GVE for this subsample are always less than the other corresponding reported estimates, although they are not out of line with these other estimates.

Since the mean GVE^{Int} scores for the entire sample and two subsamples in table 4 are essen-

Table 4: Various efficiency estimates for the entire sample and two subsamples

	Own NVE	Own NIE	Own GVE	GVE^{Int}	GVE^{Nwk}	GVE^{Ov}
			All banks (183 banks)			
Mean	0.952	0.757	0.721	0.722	0.075	0.797
Std. dev.	0.014	0.135	0.129	0.130	0.009	0.134
			GDSI banks (19 banks)			
Mean	0.952	0.702	0.669	0.671	0.073	0.744
Std. dev.	0.012	0.122	0.117	0.118	0.008	0.122
			Non-GDSI banks (164 banks)			
Mean	0.952	0.763	0.727	0.728	0.075	0.803
Std. dev.	0.014	0.134	0.129	0.130	0.009	0.134

Notes: Own signifies an efficiency from the structural form of our frontier model; NVE denotes net time-varying efficiency; NIE denotes net time-invariant efficiency; GVE denotes gross time-varying efficiency; Int denotes internal; Nwk denotes network; Ov denotes overall; and GDSI denotes globally and domestically systemically important banks.

tially the same as the corresponding own mean GVE, we can conclude for these three samples that there is basically no efficiency feedback. Drawing parallels with our findings for the mean own efficiencies, the mean GVE^{Int} for GDSI banks and thus the mean GVE^{Ov} for this subsample are always below the other corresponding reported scores, although they are not out of line with these other scores. Finally on the efficiencies, we note that the mean GVE^{Nwk} scores for the entire sample and the two subsamples are around 7.5%. Although this is not a very large efficiency spillover it is certainly non-negligible as the mean GVE^{Nwk} for each of these three samples is more than 10% of the corresponding own GVE.

4.6 Boone Test Results for the Pre-Crisis Period versus the Crisis Period and Beyond

In line with our finding in the previous subsection of no efficiency feedback in the mean GVE^{Int} score, we also observe no efficiency feedback in the mean NIE^{Int} and NVE^{Int} estimates. As a result, on average, the GVE^{Int}, NIE^{Int} and NVE^{Int} scores are of the same order of magnitude as the own GVE, NIE and NVE measures from the structural form of our network frontier. Own GVE, NIE and NVE could also be obtained from the corresponding non-network specification of our model, which would involve omitting from Eq. 11 all the variables that are pre-multiplied by $\mathbf{L}_N$. Moreover, own NIE or own NVE could be obtained from a standard non-network frontier model where inefficiency is either time-invariant or time-varying. We would not of course want to estimate such non-network models for our application because based on the significant network variables in table 2 these non-network models would be misspecified, which is a source of biased efficiencies. We instead ascertain the impact of our extended Boone test for multiple REDs that represent different ways of departmentalizing the GVE^{Ov} RED, by comparing our extended test results with standard Boone test results for an individual RED which we calculate using in turn GVE^{Int}, NIE^{Int} and NVE^{Int} from our more appropriately specified network frontier model.

For each initial and new competitive regime (denoted A and B) in cases $1-3$, which represents part I of our second stage Boone analysis, we present in table 5 estimates of the four specifications of the PQR model. We apply the standard Boone test to the GVE^{Int} RED using models A(iii) and B(iii) in this table, and apply this test separately to the NIE^{Int} and NVE^{Int} REDs using models A(iv) and B(iv) in the same table. Before we focus on the Boone test results for cases $1-3$, we highlight some of the findings from table 5. In line with the Boone theory where a single all-encompassing RED is the only determinant of the RPD, in models A(i) and B(i) for cases 1 and 3 the GVE^{Ov} RED regressor is significant. In the same models for case 2 (GDSI banks) the GVE^{Ov} RED is not significant. Various reasons for this finding could be put forward, such as the possibility of the 'too-big-to-fail' status of GDSI banks distorting the competitive landscape leading to the absence of a relationship between the RPD and the GVE^{Ov} RED. For any PQR model it does not follow that there will be no significant difference between competition in two regimes if the RED measures are not significant. This is because the difference between insignificant RED measures can be significant.

It is clear that the models in table 5 support departmentalizing the GVE^{Ov} RED as models (ii)-(iv) provide additional information. This is evident as there are a number of instances where the (in)significance of the GVE^{Ov} RED in model (i) is at odds with that for a departmentalized RED in models (ii)-(iv). Comparing models (i)-(iv) for cases $1-3$ indicates a similarity between the (in)significance of corresponding RED regressors for cases 1 and 3. Despite the size of the GDSI banks (case 2), we can conclude therefore that it is the large number of non-GDSI banks (case 3) which are driving the results in table 5 for the entire sample (case 1). Moreover, although we report a couple of significant network RED coefficients (NVE^{Nwk} RED in model A(iv) for cases 1 and 3), in line with what we expected, the evidence points to internal RED measures being more important determinants of the RPD.

In table 6 we present for cases $1-3$ the empirical Boone test results for the difference in the integrals from the entire models for the new and initial competitive regimes. For the two regimes we apply the standard empirical Boone test to model specification (i) as it has a single RED regressor. For both regimes for model specifications (ii)-(iv), however, we apply our extended empirical Boone test to collectively account for the impact of the multiple REDs. We can see from this table that the difference in the integrals is always negative and significant. This suggests for the entire sample, GDSI banks and non-GDSI banks that there is robust evidence of competition being more intense in the new regime covering the crisis and beyond, vis-à-vis the initial pre-crisis regime. That is, for each of these three groups of banks the relative profits increased at banks who were relatively more efficient in the new competitive regime compared to the initial one.

Consistent with the increase in competition we observe for cases $1-3$ is the decline in the profits of GDSI and non-GDSI banks during $2008-15$. This decline was more marked for GDSI banks, which is in line with their average efficiency, as measured by own GVE, being 6% lower than that for non-GDSI banks during this period. Using ROA, as this is the variable we use to calculate the RPD, we illustrate the declines in profit by noting that, on average, the ROA for GDSI banks declined by 54% in $2008-15$ vis-à-vis $1994-07$, while for non-GDSI banks there was a 37% decline between these periods. We suggest four reasons for the relative sizes of

Table 5: Subsample quantile regressions at the third quartile for competition cases 1-3

	Case 1: All banks (183 banks)	Case 2: GDSI banks (19 banks)	Case 3: Non-GDSI banks (164 banks)
		1994 − 07	
Model A(i)			
RED: GVE^{Ov}	0.094***	0.122	0.094***
Constant	0.325***	0.318***	0.323***
Model A(ii)			
RED: NIE^{Ov}	0.102***	0.109	0.093***
RED: NVE^{Ov}	0.164***	0.419***	0.132***
Constant	0.197***	0.035	0.226***
Model A(iii)			
RED: GVE^{Int}	0.092***	0.209*	0.093***
RED: GVE^{Nwk}	0.021	−0.139	0.027
Constant	0.313***	0.355***	0.307***
Model A(iv)			
RED: NIE^{Int}	0.102***	0.206*	0.096***
RED: NIE^{Nwk}	0.011	−0.105	0.018
RED: NVE^{Int}	0.169***	0.441***	0.120***
RED: NVE^{Nwk}	0.073**	0.121	0.071**
Constant	0.129**	−0.075	0.166***
		2008 − 15	
Model B(i)			
RED: GVE^{Ov}	0.050**	0.061	0.045**
Constant	0.502***	0.489***	0.506***
Model B(ii)			
RED: NIE^{Ov}	0.042*	0.086	0.038
RED: NVE^{Ov}	0.103***	0.221	0.104**
Constant	0.434***	0.312*	0.437***
Model B(iii)			
RED: GVE^{Int}	0.051**	0.170	0.045*
RED: GVE^{Nwk}	0.014	−0.152	0.021
Constant	0.493***	0.509***	0.494***
Model B(iv)			
RED: NIE^{Int}	0.038	0.220	0.024
RED: NIE^{Nwk}	0.016	−0.242	0.026
RED: NVE^{Int}	0.106**	0.312*	0.106**
RED: NVE^{Nwk}	0.033	0.017	0.03
Constant	0.402***	0.311*	0.408***

Notes: RED: GVE^{Ov}, GVE^{Int} and GVE^{Nwk} are the relative overall, internal and network gross time-varying efficiency differences; RED: NIE^{Ov}, NIE^{Int} and NIE^{Nwk} are the relative net time-invariant efficiency differences; RED: NVE^{Ov}, NVE^{Int} and NVE^{Nwk} are the relative net time-varying efficiency differences; and A and B denote the initial and new competitive regimes.

these declines in profitability.[25] First, GDSI banks were singled out to substantially increase their capital positions. This raised their cost of funding loans since capital is much more expensive than demand deposits, savings deposits or purchased funds. By itself, this lowered loan growth and returns. Second, via the so-called Volcker rule in the 2010 Dodd-Frank legislation, GDSI banks, who were responsible for most of the trading activities in the industry, were singled out to substantially reduce these activities. This markedly reduced trading revenue at GDSI banks, which was a core element of their profitability during the initial competitive regime. Third, for

[25] We thank an anonymous reviewer for providing these explanations of our findings for cases $1 - 3$.

Table 6: Empirical Boone test results for a change in competition for cases 1-3

Model specification	RPD integral 1994 − 07	RPD integral 2008 − 15	Difference in the integrals	Wald test (F−test) of H_0: $\Delta = 0$	p−value
			Case 1: All banks (183 banks)		
Model (i)	0.047 (0.009)	0.025 (0.008)	−0.022	329.33 [4023]	0.000
Model (ii)	0.133 (0.022)	0.072 (0.016)	−0.061	331.16 [4021]	0.000
Model (iii)	0.057 (0.010)	0.033 (0.009)	−0.024	432.62 [4021]	0.000
Model (iv)	0.178 (0.029)	0.097 (0.027)	−0.081	333.93 [4017]	0.000
			Case 2: GDSI banks (19 banks)		
Model (i)	0.061 (0.032)	0.030 (0.067)	−0.031	15.52 [415]	0.000
Model (ii)	0.264 (0.051)	0.153 (0.097)	−0.111	13.14 [413]	0.000
Model (iii)	0.035 (0.038)	0.009 (0.059)	−0.026	27.08 [413]	0.000
Model (iv)	0.331 (0.089)	0.154 (0.111)	−0.177	15.86 [409]	0.000
			Case 3: Non-GDSI banks (164 banks)		
Model (i)	0.047 (0.009)	0.023 (0.008)	−0.024	335.50 [3603]	0.000
Model (ii)	0.113 (0.022)	0.071 (0.018)	−0.042	368.20 [3603]	0.000
Model (iii)	0.060 (0.010)	0.033 (0.009)	−0.027	418.32 [3603]	0.000
Model (iv)	0.152 (0.029)	0.093 (0.032)	−0.059	304.44 [3599]	0.000

Notes: standard errors are in (.) and degrees of freedom are in [.].

all the banks in our sample the Durbin amendment to the Dodd-Frank legislation cut debit card interchange fee revenue by one-half. Fourth, the Great Recession affected all banks in the form of (a) the recession itself, which increased loan losses and negatively impacted on two profit centers as all trading and loan growth declined; and (b) the associated monetary policy, which represented the collateral damage to banks from the recession because it had a negative impact on their profitability as it lowered the loan-deposit rate spread, i.e., it lowered loan returns more than it reduced deposit rates (which did not fall below zero).

Since the quantification of competition using standard empirical measures is well-understood, for Boone testing to offer new insights it is important to understand how it can also be used to quantify a change in competition. We therefore propose how to proceed with this and first note from figure 1 that when competition intensifies, the RPD falls for a given RED. For an individual bank this downward move can be interpreted as a percentage because the RPD is an index (see figure 1 where $0 \leq$RPD≤ 1). As a result, the difference in the integrals in table 6 can be interpreted as the average percentage change in competition across the industry. The difference in the integrals therefore measures the average percentage change in the RPD in the industry that can be attributed to the change in competition if there was no improvement in the banks' REDs. We report in table 6 some small changes in competition (e.g., −2.2% for the entire sample from model (i)) and some large ones (e.g., −17.7% for GDSI banks from model (iv)). Interestingly, we can see for cases $1-3$ from table 6 that the increase in competition is always larger from models that have net RED regressors (models (ii) and (iv)) than from models with at least one gross RED regressor (models (i) and (iii)). This suggests that some information may be being lost by using gross REDs as regressors rather than net REDs. This is probably because a gross RED accounts for multiple net REDs, which highlights the usefulness of our extended Boone test for multiple REDs.

For competition cases $1-3$ we present in figure 3 our extended empirical Boone test results for the multiple REDs in models (ii)-(iv). These multiple REDs represent different ways of departmentalizing the GVE^{Ov} RED, which is why our extended test is of the sources of the increase in competition we observed above. For completeness and to enable comparisons we also present in figure 3 the standard empirical Boone test results for the single all-encompassing GVE^{Ov} RED in model (i). The four key findings from figure 3, all of which support our decision to departmentalize the GVE^{Ov} RED as they highlight the additional learnings, are as follows. First, by comparing the standard Boone test results for the individual GVE^{Int}, NIE^{Int} and NVE^{Int} REDs in figure 3 with the results for the multiple REDs for the corresponding entire departmentalization of GVE^{Ov}, we can see that our extended Boone test ensures that we do not overlook significant sources of the increase in competition.[26] To illustrate, for the entire sample and non-GDSI banks we can see from figure 3 that by focusing solely on the GVE^{Int} RED we account for the big significant source of the increase in competition, but also overlook the GVE^{Nwk} RED, which for this departmentalization is the smaller significant other source. Second, we can see from figure 3 that the significant increase in competition for GDSI banks from model (i) in table 6 is not due to any individual significant sources from models (ii)-(iv), but is instead due to their combined effect. Third, all the significant sources of the increase in competition for the entire sample and non-GDSI banks relate to time-varying departmentalized REDs (i.e., the REDs relating to NVE^{Ov}; GVE^{Nwk}; GVE^{Int}; NVE^{Nwk}; and NVE^{Int}). Fourth, we can see that our extended Boone test results in figure 3 decompose the difference in the integrals in table 6 for models (ii)-(iv) into the contributions from individual REDs. For example, for model (ii) using the entire sample, the contributions of the NIE^{Ov} and NVE^{Ov} REDs to the -0.061 difference in the integrals are -0.030 and -0.031.

As a final point in this discussion of the first set of competition cases, we note that for a change in competition from a Boone analysis to be informative it needs to be thought of in the wider context of other social benefits and costs that occurred simultaneously. We therefore qualify the increase in competition we observe for cases $1-3$.[27] Although typically we associate a rise in social welfare with a more competitive environment, we associate cases $1-3$ with a substantial decrease in welfare in $2008-15$. This is due to the negative effects of the Great Recession far outweighing any social benefits from the increase in competition.[28]

4.7 Boone Test Results for Phases of the Business Cycle

As we noted in the introductory section, a change in competition can be endogenous and is hence actionable by the antitrust authorities, or exogenous, which the antitrust authorities can do little, if anything, about. For cases $1-3$ we know that the increase in competition we observe is

[26] For example, the entire departmentalization of the GVE^{Ov} RED that relates to the GVE^{Int} RED also includes the GVE^{Nwk} RED.

[27] We thank an anonymous reviewer for suggesting this.

[28] There are other competitive regimes in banking which have also had a perverse effect on social welfare. Typically we would associate the lifting of bank branching restrictions to allow banks to move into new markets with an increase in welfare. The opposite happened in Spain because savings banks that were previously restricted to certain regions over-branched. This wasted resources and many banks were forced to merge, while others made losses. Overall, the rise in competition was temporary and many would conclude that the accompanying short-lived social benefits were more than offset by the decline in social welfare from what followed.

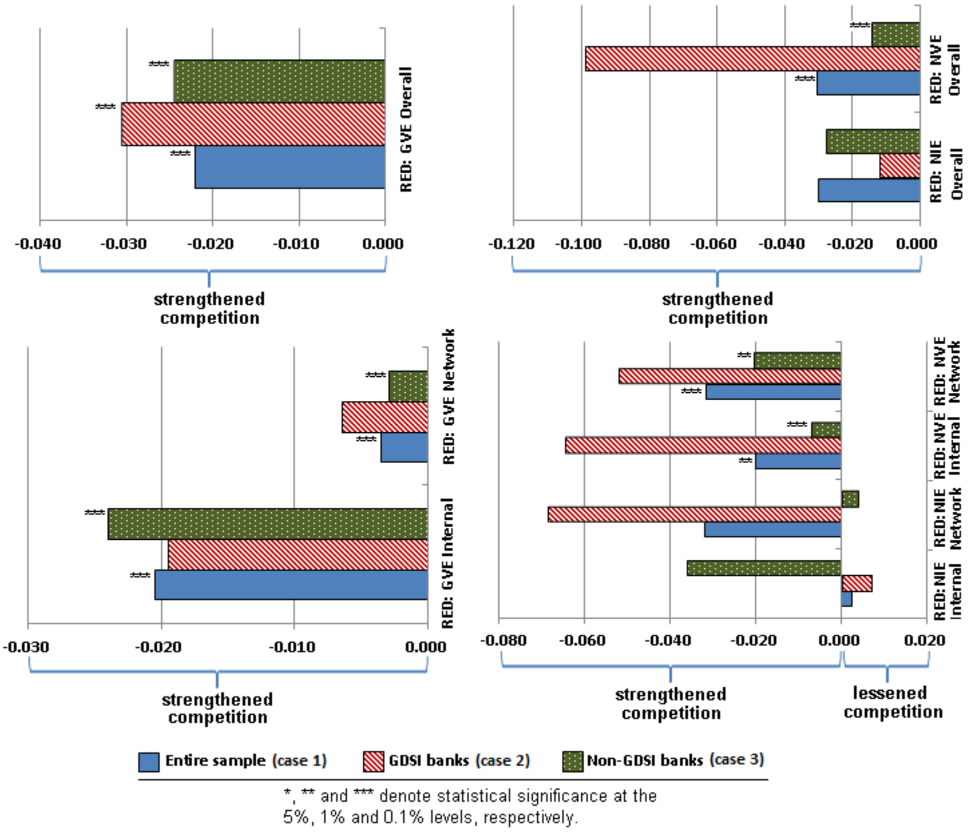


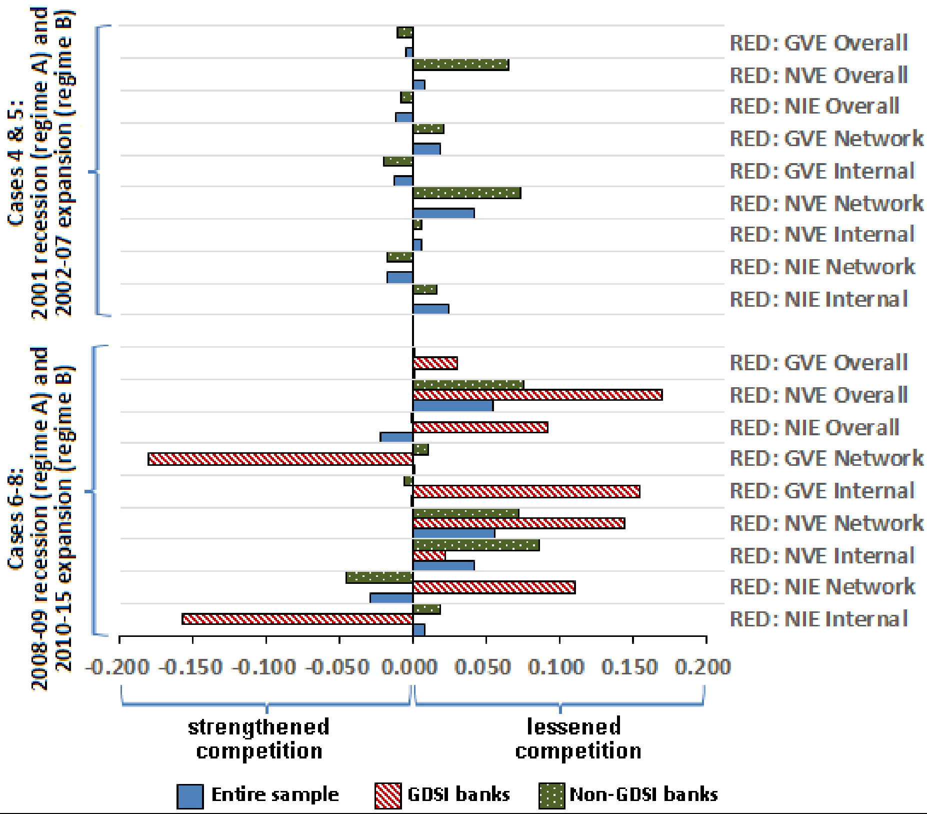
Figure 3: Empirical Boone test results for sources of a change in competition for cases 1-3

exogenous. This is because it is the result of crisis induced phenomena that forced a decline in profits upon the banking industry. Recall that there are three such phenomena- the U.S. economy going into recession; the decrease in the loan-deposit rate spread, which represented the collateral damage to banks from monetary policy to moderate the recession; and the post-crisis tightening of bank regulation to reduce risk.

We now turn our attention to the origins of the exogenous increase in competition for cases $1-3$. We do so by introducing an approach to disentangle the effect of the business cycle on bank competition from the combined effect of the other two crisis induced phenomena. This involves testing within the initial and new regimes whether the decrease in competition that we expect in the expansion phase of the business cycle cancels out the increase that we expect in the recession. We apply our approach to: (i) the complete recession and expansion phases in the pre-crisis period $(1994-07)$ for the entire sample (case 4) and non-GDSI banks (case 5);[29] and (ii) the complete recession and incomplete expansion in the period covering the crisis and beyond $(2008-15)$ for the entire sample (case 6), GDSI banks (case 7) and non-GDSI banks (case 8).

For cases $4-8$, we present in figure 4 the results of the standard Boone test using the single all-encompassing GVE^{Ov} RED and the results of our extended test using various departmentalizations of this RED. From this figure we can see for each case that the standard Boone test indicates that the decrease in competition we associate with the expansion phase is nullified by the increase we associate with the recession. This cancelling out of changes in competition over the business cycle suggests that the exogenous increase in competition for cases $1-3$ has two

[29] For reasons previously given we do not consider these phases for GDSI banks.



There are no REDs for cases 4-8 associated with a statistically significant change in the intensity of competition at the 5% level. Only the RED: NVE Network for non-GDSI banks (case 8) is associated with a marginally statistically significant change in the intensity of competition (i.e., at the 10% level).

Figure 4: Empirical Boone test results for sources of a change in competition for cases 4-8

origins- the collateral damage to banks from monetary policy to moderate the recession that the crisis initiated, and the post-crisis tightening of bank regulation to reduce risk. Consistent with this cancelling out, all the results for cases $4-8$ in figure 4 for our extended test (with one possible exception) suggest that each source of the increase in competition we associate with the recession is negated in the ensuing expansion.[30]

4.8 Boone Test Results for Further Subperiods and Bank Classifications

As cases $1-3$ consider two long subperiods in cases $9-17$ we consider whether competition differs in successive shorter subperiods. We consider four shorter subperiods: $1994-00$ (before the build-up to the crisis began); $2001-06$ (build-up to the crisis); $2007-09$ (crisis period); and $2010-15$ (post-crisis period). Three cases are for the entire sample (9, 12 and 15) and since case 2 relates to a small fixed group of GDSI banks, three cases consider a bigger group of the largest banks that changes annually (10, 13 and 16). This subsample consists of banks with total assets in the top third of the sample in each year and a further three cases are for the remaining two-thirds (11, 14 and 17).

[30] The NVE^{Nwk} RED for non-GDSI banks, which is part of case 8, is the possible exception. This is because it is the source of a marginally significant (i.e., at the 10% level) negative competitive effect over the incomplete business cycle during $2008-15$.

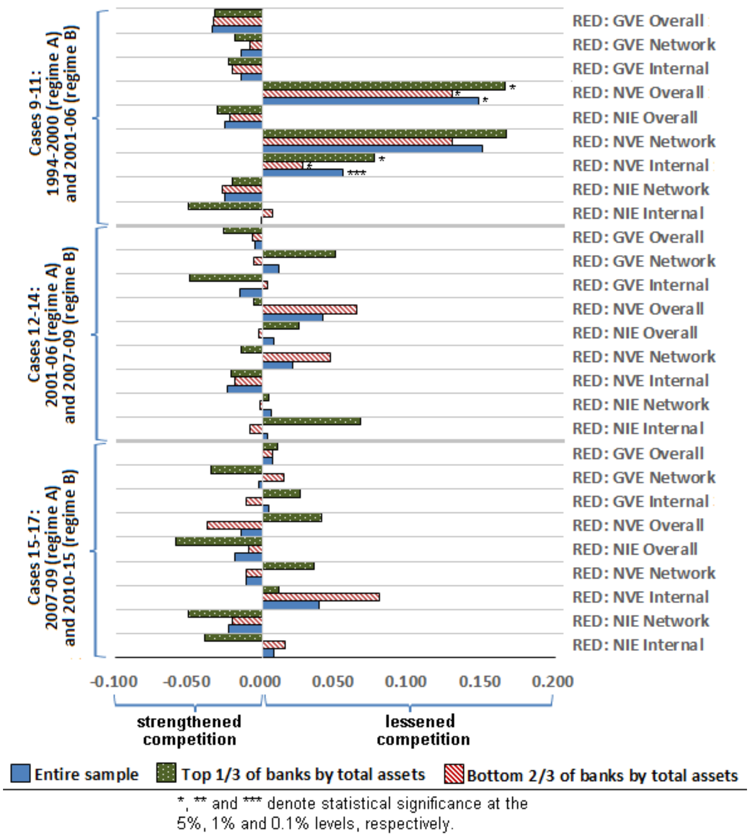


Figure 5: Empirical Boone test results for sources of a change in competition for cases 9-17

In figure 5, we present for cases $9-17$ the results of the standard Boone test using the single all-encompassing GVE^{Ov} RED and the results of our extended test using different departmentalizations of this RED. From this figure we can see for each case that the standard Boone test indicates that competition is not statistically significantly different in the successive subperiod. This is interesting because comparing these results with the results for the standard Boone test using the GVE^{Ov} RED for cases $1-3$ suggests that longer subperiods are needed to observe a change in competition.

From the extended test results for the departmentalized REDs in figure 5, we can see that there are just six instances where a source of competition is significantly different in the successive subperiod. (These instances are the NVE^{Int} and NVE^{Ov} REDs when the initial and new regimes are $1994-00$ and $2001-06$ for the entire sample, the top third of banks and the bottom two-thirds, where these results are part of cases $9-11$, respectively). This relative lack of evidence for cases $9-17$ of sources of competition that are significantly different in the new regime is broadly consistent with the standard Boone test results using the GVE^{Ov} RED which indicate that competition did not change for these cases. Moreover, as the NVE^{Ov} RED can be departmentalized into NVE^{Int} and NVE^{Nwk} REDs, we can conclude for cases $9-11$ that it is the significant NVE^{Int} RED that is driving the significant results for the NVE^{Ov} RED.

5 Concluding Remarks

The two contributions this study makes are varied in nature as the first is an applied methodological one, while the second is empirical for banks. Our first contribution extends the method for the empirical implementation of the Boone competition test to test for different sources of a change in competition. This involves moving from a test that uses a single all-compassing efficiency to one that uses multiple efficiencies. In other words, we identify the different sources of a change in competition using multiple efficiencies with different features, which involves drawing on recent advances in efficiency measurement. The multiple efficiencies we incorporate into our extended Boone test have the following different features: time-variance; time-invariance; internalization to the firm; and network effects as firms compete via their location choices for, among other things, network linkages.

Our second contribution is a comprehensive empirical application of our extended Boone test to large U.S. banks. In this application we provide deeper insights into how a change in competition from a Boone analysis may be quantified. We also emphasize that one can only carry out a true test of a change in endogenous competition (i.e., competition that is actionable by the antitrust authorities) when: (a) it is possible to distinguish between the exogenous and endogenous changes in competition; or (b) the analysis is performed on data where the exogenous effects on competition are considered to be minimal, or have remained constant over the study period. Of course for our study period neither (a) or (b) is the case, but because we know that crisis induced phenomena forced a decline in profits upon the banking industry, we instead analyze the exogenous origins of the decline in banking competition we observe for the period $2008-15$. We do so by suggesting an approach to investigate if one can rule out the exogenous phase of the business cycle as the origin of a change in competition. When we apply this approach we find no change in competition over the complete and incomplete business cycles we consider (i.e., it appears that the increase in competition in the recession is cancelled out by the decrease in the expansion phase). On the basis of this finding, the exogenous increase in competition we observe for the period $2008-15$ has two origins- the decrease in the loan-deposit rate spread, which represented the collateral damage to banks from monetary policy to moderate the Great Recession; and the post-crisis tightening of bank regulation to reduce risk.

References

Aigner, D., C. A. K. Lovell and P. Schmidt (1977): 'Formulation and estimation of stochastic frontier production function models'. *Journal of Econometrics*, vol. 6, pp. 21-37.

Adams, R. M., A. N. Berger and R. C. Sickles (1999): 'Semiparametric approaches to stochastic panel frontiers with applications in the banking industry'. *Journal of Business and Economic Statistics*, Vol. 17, pp. 349-358.

Anselin, L. (2003): 'Spatial externalities, spatial multipliers and spatial econometrics'. *International Regional Science Review*, vol. 26, pp. 153-166.

Bank for International Settlements (2018): 'Structural changes in banking after the crisis'. Committee on the Global Financial System (CGFS) papers, No. 60.

Battese, G. E. and T. J. Coelli (1988): 'Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data'. *Journal of Econometrics*, vol. 38, pp. 387-399.

Berger, A. N. (1993): '"Distribution-free" estimates of efficiency in the U.S. banking industry and tests of the standard distributional assumptions'. *Journal of Productivity Analysis*, vol. 4, pp. 261-292.

Berger, A. N. and C. H. S. Bouwman (2013): 'How does capital affect bank performance during financial crises?' *Journal of Financial Economics*, vol. 109, pp. 146-176.

Berger, A. N. and D. B. Humphrey (1992): *Measurement and efficiency issues in commercial banking.* In *Output Measurement in the Service Sectors,* Griliches, Z. (Ed). Chicago: NBER, University of Chicago Press.

Berger, A. N. and R. A. Roman (2017): 'Did saving Wall Street really save Main Street? The real effects of TARP on local economic conditions'. *Journal of Financial and Quantitative Analysis*, vol. 52, pp. 1827-1867.

Board of Governors of the Federal Reserve System (2014): 'Comprehensive capital analysis and review 2014: Assessment framework and results'.

Bolt, W. and D. Humphrey (2010): 'Bank competition efficiency in Europe'. *Journal of Banking and Finance*, vol. 34, pp. 1808-1817.

Bolt, W. and D. Humphrey (2015a): 'Assessing bank competition for consumer loans'. *Journal of Banking and Finance*, vol. 61, pp. 127-141.

Bolt, W. and D. Humphrey (2015b): 'A frontier measure of U.S. banking competition'. *European Journal of Operational Research*, vol. 246, pp. 450-461.

Boone, J. (2008): 'A new way to measure competition'. *Economic Journal*, vol. 118, pp. 1245-1261.

Casu, B., A. Ferrari, C. Girardone and J. O. S. Wilson (2016): 'Integration, productivity and technological spillovers: Evidence for Eurozone banking industries'. *European Journal of Operational Research*, vol. 255, pp. 971-983.

Claessens, S. and L. Laeven (2004): 'What drives bank competition? Some international evidence'. *Journal of Money, Credit and Banking*, vol. 36, pp. 563-583.

Coelli, T. J., S. Perelman and E. Romano (1999): 'Accounting for environmental influences in stochastic frontier models: With application to international airlines'. *Journal of Productivity Analysis*, vol. 11, pp. 251-273.

Degryse, H., M. Kim and S. Ongena (2009): *Microeconometrics of Banking- Methods, Applications and Results.* New York: Oxford University Press.

Duygun, M., M. Shaban and T. Weyman-Jones (2015): 'Measuring competition using the Boone relative profit difference indicator'. *Economics Letters*, vol. 132, pp. 117-120.

Elhorst, J. P. (2009): *Spatial panel data models.* In the *Handbook of Applied Spatial Analysis,* Fischer, M. M., and A. Getis (Eds). New York: Springer.

Fiordelisi, F. and P. Molyneux (2010): 'Total factor productivity and shareholder returns in banking'. *Omega*, vol. 38, pp. 241-253.

Financial Stability Board (2015): '2015 update of list of global systemically important banks (G-SIBs)'.

Glass, A. J. and K. Kenjegalieva (2019): 'A spatial productivity index in the presence of efficiency spillovers: Evidence for U.S. banks, 1992-2015'. *European Journal of Operational Research*, vol. 273, pp. 1165-1179.

Glass, A. J., K. Kenjegalieva and R. C. Sickles (2013): 'How efficiently do U.S. cities manage roadway congestion?' *Journal of Productivity Analysis*, vol. 40, pp.407-428.

Glass, A. J., K. Kenjegalieva and R. C. Sickles (2016a): 'A spatial autoregressive stochastic frontier model for panel data with asymmetric efficiency spillovers'. *Journal of Econometrics*, vol. 190, pp. 289-300.

Glass, A. J., K. Kenjegalieva, R. C. Sickles and T. Weyman-Jones (2016b): 'The spatial efficiency multiplier and random effects in spatial stochastic frontier models'. Available at
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2873520

Goddard, J. and J. O. S. Wilson (2009): 'Competition in banking: A disequilibrium approach'. *Journal of Banking and Finance*, vol. 33, pp. 2282-2292.

Gouriéroux, C., A. Holly and A. Monfort (1982): 'Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters'. *Econometrica*, vol. 50, pp. 63-80.

Greene, W. H. (1980): 'Maximum likelihood estimation of econometric frontier functions'. *Journal of Econometrics*, vol. 13, pp. 27-56.

Hannan, T. H. and R. A. Prager (2004): 'The competitive implications of multimarket bank branching'. *Journal of Banking and Finance*, vol. 28, pp. 1889-1914.

Hannan, T. H. and R. A. Prager (2009): 'The profitability of small single-market banks in an era of multi-market banking'. *Journal of Banking and Finance*, vol. 33, pp. 263-271.

Hirtle, B. (2007): 'The impact of network size on bank branch performance'. *Journal of Banking and Finance*, vol. 31, pp. 3782-3805.

Kao, Y-H. and A. K. Bera (2013): 'Spatial regression: The curious case of negative spatial dependence'. University of Illinois, Urbana-Champaign, Mimeo.

Koetter, M., J. W. Kolari and L. Spoerdijk (2012): 'Enjoying the quiet life under deregulation? Evidence from adjusted Lerner indices for U.S. banks'. *Review of Economics and Statistics*, vol. 94, pp. 462-480.

LeSage, J. and R. K. Pace (2009): *Introduction to Spatial Econometrics.* Boca Raton, Florida: CRC Press, Taylor and Francis Group.

Liu, H., P. Molyneux and J. O. S. Wilson (2013): 'Competition and stability in European banking: A regional analysis'. *Manchester School*, vol. 81, pp. 176–201.

Panzar, J. C. and J. N. Rosse (1987): 'Testing for "monopoly" equilibrium'. *Journal of Industrial Economics*, vol. 35, pp. 443-456.

Rosen, R. (2007): 'Banking market conditions and deposit interest rates'. *Journal of Banking and Finance*, vol. 31, pp. 3862-3884.

Sealey, C. and J. T. Lindley (1977): 'Inputs, outputs and a theory of production and cost at depository financial institutions'. *Journal of Finance*, vol. 32, pp. 1251-1266.

Tabak, B. M., R. B. Miranda and D. M. Fazio (2013): 'A geographically weighted approach to measuring efficiency in panel data: The case of US saving banks'. *Journal of Banking and Finance*, vol. 37, pp. 3747-3756.

Wheelock, D. C. and P. W. Wilson (2012): 'Do large banks have lower costs? New estimates of returns to scale for U.S. banks'. *Journal of Money, Credit and Banking*, vol. 44, pp. 171-199.

Wheelock, D. C. and P. W. Wilson (2018): 'The evolution of scale economies in US banking'. *Journal of Applied Econometrics*, vol. 33, pp. 16-28.

Yeyati, E. L. and A. Micco (2007): 'Concentration and foreign penetration in Latin American banking sectors: Impact on competition and risk'. *Journal of Banking and Finance*, vol. 31, pp. 1633-1647.