Article:

Adetutu, M.O., Glass, A.J. orcid.org/0000-0002-5984-7666, Kenjegalieva, K. orcid.org/0000-0001-7323-077X et al. (1 more author) (2025) Product level market power spillovers among U.S. banks. Papers in Regional Science, 104. 100093. ISSN 1056-8190

https://doi.org/10.1016/j.pirs.2025.100093

Product level market power spillovers among U.S. banks

Morakinyo O. Adetutu[*], Anthony J. Glass[†], Karligash Kenjegalieva[‡] and Kayode A. Odusanya[§]

February 2025

Abstract

Bank market power has far-reaching effects as, among other things, it affects the price of credit. Even though it is well-known that banks are spatially interdependent due to rival banks having branches in the same geographical areas, the literature on bank market power overlooks this. To measure market power spillovers, we set out an approach to calculate spill-in and spill-out Lerner indices for firms and their products. To account for the marked consolidation over the sample, we use unbalanced panel data comprising over 45,000 observations for large commercial U.S. banks. From spatial stochastic frontier models, we obtain estimates of these indices (with and without adjustment for inefficiency spill-ins and spill-outs). We observe high spill-in Lerner indices for some banks, which is consistent with consolidation in the industry leading to concerns about bank market power. In line with larger agglomeration effects being conducive to higher spillovers, banks with high spillover Lerner indices tend to have branches in major cities.

Key words: Multi-product banks; Branch geography; Spatial stochastic frontier analysis; Market power; Lerner index.

[*]Loughborough Business School, Loughborough University, Loughborough, Leics, LE11 3TU, UK (Email: m.o.adetutu@lboro.ac.uk).

[†]Corresponding author. Sheffield University Management School, Conduit Road, Sheffield, S10 1FL, UK (Email: a.j.glass@sheffield.ac.uk). The authors acknowledge the handing editor and two anonymous reviewers for their comments.

[‡]Loughborough Business School, Loughborough University, Loughborough, Leics, LE11 3TU, UK (Email: k.a.kenjegalieva@lboro.ac.uk). This paper acknowledges the use of the Lovelace High Performance Computing service at Loughborough University.

[§]Loughborough Business School, Loughborough University, Loughborough, Leics, LE11 3TU, UK (Email: k.odusanya@lboro.ac.uk).

1 Introduction

A firm has market power if it is able to raise its price to increase its profits. Other things unchanged, the higher price will reduce the output quantity the firm sells, which together will lower consumer benefit. Accordingly, the importance of measuring the market power of firms has long since been recognized. For banks, the implications of market power are far-reaching because, among other things, it affects the prices of credit to firms and individuals, which has a wider effect on the general business environment. In line with such key implications, there is a vast literature on measuring bank market power that covers a range of countries, e.g., European Monetary Union (EMU) countries (e.g., Maudos and De Guevara, 2007; Delis and Tsionas, 2009; Wang *et al.*, 2020; Coccorese *et al.*, 2021), the United States (U.S.) (e.g., Shaffer and Spierdijk, 2020; Wang *et al.*, 2022; Mi *et al.*, 2024) and emerging countries (Semih Yildirim and Philippatos, 2007; Efthyvoulou and Yildirim, 2014; Danisman and Demirel, 2019), to name only a small selection of studies in this large literature. In the U.S. banking industry, consolidation has substantially reduced the number of banks, which has led to the largest banks having a much larger share of the industry's total assets. This consolidation has led to concerns about the market power of these banks, thereby underlining the practical relevance of our analysis. Accordingly, rather than analyze market power across the U.S. banking industry or across a key subset of U.S. banks, we focus on market power at the micro levels of individual large U.S. banks and their products.

It is well-known that banks are interdependent, which is due to, among other things, rival banks having branches in the same geographical areas. For individual banks and their multiple outputs, and using approaches that overlook the spatial interconnectedness between the banks, the extant literature estimates the usual own market power. However, by overlooking this interconnectedness these own market power estimates may be biased. This is because these estimates may be conflated with the market power spillovers pertaining to the omitted interconnectedness: namely, a bank's asymmetric market power spill-in and spill-out from and to the other sampled banks. To address this potential bias, we introduce the first method to estimate bank and product level market power spillovers. Moreover, as our method is not specific to banks, subject to data availability, it can be applied to estimate these micro level market power spillovers in other industries.

We report bank and product level asymmetric market power spill-ins and spill-outs with and without adjustment for the corresponding asymmetric inefficiency spillovers. We obtain and compute the inefficiency and market power spillovers from a fixed effects spatial stochastic frontier analysis (SSFA). This involves extending the non-spatial fixed effects stochastic frontier analysis (SFA) in Chen *et al.* (2014) to the spatial setting by allowing the spatial lag of the dependent variable, i.e., the spatial autoregressive (SAR) variable, to impact the frontier (e.g., Glass *et al.*, 2016; Jin and Lee, 2020; Lai and Tran, 2022; Tran *et al.*, 2023; Tran and Tsionas, 2023). Consequently, we do not account for the spatial dependence using the spatial lag of (i) the disturbance (e.g., Druska and Horrace, 2004; Orea and Álvarez, 2019; Hou *et al.*, 2021; Skevas and Skevas, 2021); (ii) inefficiency (e.g., Orea and Álvarez, 2019; Skevas and Skevas, 2021; Hou *et al.* 2021; Fusco *et al.* 2024); or (iii) each x regressor (e.g., Adetutu *et al.*, 2015). This is because a model with the SAR variable yields the asymmetric indirect global (i.e., first, second, third, etc., order neighbor) spillovers of the x variables, inefficiencies and noise that we need to calculate the asymmetric market power spill-ins and spill-outs.[1] Our model and estimation procedure therefore add to the set of spatial frontier approaches in the regional literature (Ramajo *et al.*, 2017; Ramajo and Hewings, 2018; Kutlu and Nair-Reichert, 2019; Algeri *et al.*, 2022; Glass and Kenjegalieva, 2024). More generally, for a comprehensive coverage of the spatial frontier literature that documents its growth and extends beyond the regional literature, see Ayouba (2023).

In contrast to maximum likelihood estimation of a linear fixed effects model, the usual demeaning (i.e., within transformation) of the data to eliminate the fixed effects does not yield a tractable likelihood function for non-linear models, e.g., a stochastic frontier (Wang and Ho, 2010). This has led to three ways of estimating a fixed effects stochastic frontier. The first is to retain the fixed effects dummy variables by not demeaning (Greene, 2005), but this may lead to inconsistent estimates due to the well-known incidental parameters problem (Neyman and Scott, 1948). The second involves modeling time-varying inefficiency as the product of the following: a particular function that depends on time-varying exogenous variables, and a time-invariant inefficiency term that has a truncated normal distribution (see Wang and Ho, 2010, for details). By modeling inefficiency in this way, demeaning the data yields a tractable likelihood function. In the third way, as well as making the usual half-normal and normal distributional assumptions for the time-varying inefficiency and noise components of the error, the error term is also assumed to have a closed skew normal distribution (see Chen *et al.*, 2014, for details). This modeling of the error term is a further way of obtaining a tractable likelihood function with demeaned data. Here we extend the more recent Chen *et al.* approach to the case of SAR dependence. Moreover, a further feature of our analysis is the unbalanced panel data.[2] This is key for our analysis as the aforementioned consolidation in the U.S. banking industry relates to more than a 50% fall in the number of banks over our sample. Turning now to discuss in more detail the main focus of this paper: namely, the general methodology we introduce to quantify market power spillovers and the empirical analysis of these spillovers for large commercial U.S. banks.

We introduce bank and product level asymmetric spill-in and spill-out Lerner indices. To put this contribution into context we first briefly consider the standard non-spatial product level Lerner index for a bank (e.g., Shaffer and Spierdijk, 2020; Wang *et al.*, 2020; Mi *et al.*,

[1] From a model with a spatial lag of noise or inefficiency, we only obtain asymmetric indirect global noise or inefficiency spillovers. Moreover, a spatial model with only spatial lags of the x variables, i.e., the SLX model (Halleck Vega and Elhorst, 2015), yields only the impacts of the x variables of a unit's first order neighbors (i.e., unidirectional local spillovers in the inward direction). We also gave consideration to augmenting our SAR model with the spatial lags of the x regressors, i.e., the spatial Durbin model. Whilst this model would yield the required asymmetric indirect global spillovers of the x variables, inefficiencies and noise, we did not pursue this model. This is because our SAR model is a more parsimonious specification, as the spatial Durbin model would involve including spatial lags of all the squares and interactions in our translog functions (cost, revenue and output distance). In addition, as we are primarily interested in the asymmetric market power spill-ins and spill-outs, our model follows a number of non-spatial and spatial stochastic frontiers that focus on the determinants of the frontier. For spatial stochastic frontiers that also include determinants of the variance or mean of inefficiency, see Gude *et al.* (2018), Kutlu *et al.* (2020) and Galli (2023a; 2023b; 2024).

[2] More generally, for unbalanced spatial panel data analysis that does not involve inefficiency measurement using spatial frontier methods, see Egger *et al.* (2005) and Baltagi *et al.* (2007; 2015).

2024).

$$L_{kit} = \frac{P_{kit} - MC_{kit}}{P_{kit}} = \frac{\frac{R_{kit}}{Q_{kit}} - MC_{kit}}{\frac{R_{kit}}{Q_{kit}}} < 1, \quad (1)$$

where the banks, time periods and products which the banks offer are indexed $i \in 1, ..., N$, $t \in 1, ..., T$, and $k \in 1, ..., K$, respectively For product k of bank i in period t, P_{kit} is the output price, MC_{kit} is marginal cost, Q_{kit} is the output quantity and R_{kit} is revenue. For the single output case, when a firm has no market power L_{it} is 0, but when a firm has some market power $0 < L_{it} < 1$. In the multiple output case $P_{kit} - MC_{kit} \geq 0$ (and hence $L_{kit} \geq 0$) is not guaranteed. $P_{kit} - MC_{kit} < 0$ (and hence $L_{kit} < 0$) could be due to the optimal cross-subsidization of product k by bank i, e.g., a bank may cross-subsidize some off-balance sheet activities by pricing them below MC_{ik} to some existing customers that use other products (Shaffer and Spierdijk, 2020). Alternatively, $P_{kit} - MC_{kit} < 0$ may not represent profit maximizing behavior due to the bank having a sub-optimal business strategy for product k. Frontier analysis caters for both possibilities as it allows for profit maximization or sub-optimal profits by not explicitly imposing the former.

Both types of non-spatial Lerner index reported in the literature can be obtained from Eq. 1. One type involves an adjustment for inefficiency, while the other does not. To obtain the latter: (i) R_{kit} and Q_{kit} assume the bank lies on its revenue and output distance frontiers, respectively; and (ii) $MC_{kit} = \frac{C_{it}}{Q_{kit}} \frac{\partial \ln C_{it}}{\partial \ln Q_{kit}}$ (e.g., Shaffer and Spierdijk, 2020) is calculated using $\frac{\partial \ln C_{it}}{\partial \ln Q_{kit}}$ from a fitted cost function and the data points C_{it} and Q_{kit}, where C_{it} is based on the bank being on its cost frontier. The other type of L_{kit} is calculated in a similar way and accounts for inefficiencies. This means that the numerator and denominator of $\frac{R_{kit}}{Q_{kit}}$ are adjusted for revenue and output distance inefficiencies, and in the calculation of MC_{kit}, C_{it} is adjusted for cost inefficiency.[3] For further details on adjusting Lerner indices for inefficiencies, see subsection 2.2.

Our extension of the Lerner index to introduce asymmetric spill-in and spill-out indices is motivated by the index being a widely used measure of market power. In the banking literature, however, other measures of market power are also used. To further motivate our extension by, among other things, highlighting the practical reporting of banking Lerner indices by the World Bank, we briefly review in Appendix 1 the three main measures of bank market power that were candidates for extension to the case of spill-ins and spill-outs. In terms of more general motivation of our analysis, to the best of our knowledge, Dix and Orzach (2023) is the only other study to consider market power spillovers. Their approach, however, differs from ours as they analyze *within* airline market power spillovers when an airline operates connecting flights. Their approach is also specific to an airline's connecting flights, whereas our spatial approach can analyze the general case of market power spillovers *between* firms and/or different firms' products.

The calculation of the bank and product level asymmetric spill-in and spill-out Lerner indices is a two-step process. In the first step we estimate spatial cost, alternative revenue and output distance stochastic frontier models. In the second step, we use these fitted models to calculate

[3] In the calculation of the inefficiency adjusted estimate of MC_{kit}, Q_{kit} is not adjusted for output distance inefficiency. This is because Q_{kit} is given in the estimation of the cost function.

two types of asymmetric bidirectional spatial Lerner indices. The first type are overall indices and represent market power spill-ins and spill-outs from and to all other banks. The second type are partitioned indices and measure the corresponding spill-ins and spill-outs from and to 1st order, 2nd order, etc. neighbors. The partitioned indices are informative as they allow us to examine how the market power spill-ins and spill-outs die out across higher order neighborhood sets. As we discuss further in due course, computing the asymmetric spatial Lerner indices involves obtaining estimates of the unobserved price and marginal cost spill-ins and spill-outs for individual banks and their products. To obtain the estimates of the asymmetric bidirectional price spillovers, we obtain revenue and quantity spill-ins and spill-outs from the results for the spatial alternative revenue and output distance models. By applying the non-spatial approach in Shaffer and Spierdijk (2020) to our bank level spatial cost model, we obtain the marginal cost spill-in and spill-out for each output. We also report further overall and partitioned spatial Lerner indices that are adjusted for the overall and partitioned inefficiency spill-ins and spill-outs to and from a bank.

Bank interconnectedness is an important phenomenon as it underlies various systemic risks, e.g., bank run contagion. Accordingly, U.S. bank regulatory authorities dedicate a lot of resources to monitor different forms of this interconnectedness. Using over $45,000$ observations for large commercial U.S. banks (1994:Q1 – 2022:Q4), we obtain estimates of bank market power spillovers. The three key results for these spillovers, which we now preview, represent new information about the interconnectedness of U.S. banks. First, consistent with consolidation in the industry leading to concerns about the market power of the largest banks, a number of banks have relatively high spillover Lerner indices, e.g., two global systemically important banks (Bank of America and JPMorgan Chase). This finding suggests that overlooking bank market power spillovers may result in U.S. competition authorities understating the market power impact of a large bank merger. The implication being that overlooking these spillovers may lead to unexpectedly larger increases in the price of credit and, as a result, unexpectedly bigger negative impacts on the general business environment and household welfare. Therefore, from a policy perspective, we suggest that U.S. competition authorities should account for such spillovers when assessing future large bank mergers.

Second, we report a bank level spill-in Lerner index for quintile 5 of the bank size distribution that is well below that for the pool of quintile 5 banks that are in 95% of the study period. This emphasizes the importance of unbalanced panel data for our empirical case as the lower market power spill-ins for the banks outside this pool are intuitive as they may have contributed to some of these banks dropping out the sample. Third, banks with both spill-in and spill-out Lerner indices that are in the top thirds of the estimates tend to have branches in major cities. This stands to reason as the bigger agglomeration effects in major cities will promote market power spillovers.

The remainder of this paper is organized as follows. Section 2 sets out the two-step empirical methodology. The first step is presented in 2.1 and comprises three parts. (i) The SAR cost, alternative revenue and output distance stochastic frontiers. (ii) The approaches to spatially partition the asymmetric (spill-in and spill-out) indirect marginal effects and inefficiencies across 1st order, 2nd order, etc. neighborhood sets. The own coefficients and inefficiencies from the models represent the impacts on a single bank. However, a change in an explanatory variable and a particular bank's inefficiency can potentially affect the dependent variables of all the other banks. The indirect marginal effects and inefficiencies account for this. (iii) How we use the above overall and partitioned indirect marginal effects and inefficiencies to construct the corresponding indirect translog functions. The second step is presented in 2.2 and focuses how we use these functions to obtain the (un)partitioned spill-in and spill-out Lerner indices. Section 3 presents the empirical analysis and section 4 summarizes.

2 Modeling framework

2.1 Step 1: SAR stochastic frontiers, marginal effects and inefficiencies

We estimate SAR specifications of the cost, alternative revenue and output distance stochastic frontier models. In the standard revenue function, revenue is a function of input quantities and output prices, whereas in the alternative revenue function, revenue is a function of outputs and input prices. We use an alternative revenue function as it has well-established merits (e.g., Berger and Mester, 2003). Among other things: (i) it accounts for the constraints on a bank's ability to change its output quantities in the short-run, which is captured by the inclusion of these quantities in the function; and (ii) there is less measurement error with an alternative revenue function as input prices are more accurately measured than output prices.

The SAR cost, alternative revenue and output distance stochastic frontier models are set out in Eqs. $2-4$, respectively, where in below three models all the variables are logged.

$$c_{it} = \alpha + TL\left(t, q_{it}, m_{it}\right) + z_{it}\gamma' + \delta \sum_{j=1}^{N_t} w_{ijt} c_{jt} + b_t + d_i + v_{it} + u_{it}, \tag{2}$$

$$r_{it} = \alpha + TL\left(t, q_{it}, m_{it}\right) + z_{it}\gamma' + \delta \sum_{j=1}^{N_t} w_{ijt} r_{jt} + b_t + d_i + v_{it} - u_{it}, \tag{3}$$

$$-q_{kit} = \alpha + TL\left(t, \widetilde{q}_{it}, s_{it}\right) + z_{it}\gamma' + \delta \sum_{j=1}^{N_t} w_{ijt}\left(-q_{kjt}\right) + b_t + d_i + v_{it} + u_{it}. \tag{4}$$

As we use the estimation results for Eqs. $2-4$ to calculate market power spillovers, we use matched unbalanced panel datasets for these models that comprise the same banks for the same time periods. The banks in period t are indexed $i, j \in 1, ..., N_t$ where $i \neq j$. For the ith bank in period t, c_{it}, r_{it} and q_{kit} are the cost, revenue and quantity of the kth output, respectively. TL denotes the translog functional form, where m_{it} is the vector of input prices, $\widetilde{q}_{it}$ is the vector of $K-1$ outputs and s_{it} is the vector of input quantities. Some of the parameters to be estimated include the common intercept α, the coefficients in TL, and the vector of coefficients (γ') on the non-spatial environment variables (z). The relationships between the dependent and independent variables collectively represent the frontier. Of the independent variables, only those in the TL function influence the monotonicity and curvature of the frontier. We, therefore, follow the efficiency and productivity literature and refer to the covariates outside TL as environmental variables as they shift the frontier up or down.

$\sum_{j=1}^{N_t} w_{ijt} c_{jt}$, $\sum_{j=1}^{N_t} w_{ijt} r_{jt}$ and $\sum_{j=1}^{N_t} w_{ijt}\left(-q_{kjt}\right)$ are observations of the SAR environmental variables. To construct these variables we use the exogenous, a priori specified, non-negative spatial weights (w_{ijt}'s), which are collected in the $N_t \times N_t$ matrix $\mathbf{W}_t$. $\mathbf{W}_t$ represents which banks neighbor one another and the strength of the linkages between the banks, where all the elements on the main diagonal are set to zero to rule out self-influence. The feasible range of values of each SAR coefficient is $\delta \in \left(\frac{1}{\min(h_1^{\min},...,h_T^{\min})}, \frac{1}{\max(h_1^{\max},...,h_T^{\max})}\right)$, where $h_t^{\min}$ and $h_t^{\max}$ are the most negative and positive real characteristic roots of $\mathbf{W}_t$. Note also that $\mathbf{W}_t$ is normalized. See 3.1 for details of this normalization and the specification of $\mathbf{W}_t$ for the empirical analysis, where the normalizing factor we use yields $h_t^{\max} = 1$.

We account for the effects of time in Eqs. $2-4$ using time period effects (b_t) and a non-linear time trend (as t, t^2 and interactions with t are part of TL). By including time period effects to account for common shocks across the banks, due to, for example, systemic factors, we do not conflate these shocks with the effect of the SAR variable. Given the non-linear time trend, which is a proxy for technical change, is estimated using data over the whole sample, b_t captures a common departure from the time trend in a particular period.

In a stochastic frontier model the composed error is of particular interest, where here this error is $\varepsilon_{it} = v_{it} \pm u_{it}$. This comprises noise, $v_{it}\tilde{}\mathrm{N}\big(0, \sigma_v^2\big)$, and inefficiency, which, as is common, is assumed to be half-normally distributed, $u_{it}\tilde{}\mathrm{N}^+\left(0, \sigma_u^2\right)$. Note that u_{it} has a positive (negative) sign in the cost (revenue) frontier model. This is because it measures how much a bank is above (below) its best practice cost (revenue) frontier. The output distance function (ODF) assumes that a bank is seeking to maximize multiple outputs using a given quantity of inputs. A bank's shortfall, however, from its best practice ODF in Eq. 4 (u_{it}) has a positive sign, whereas in the single output stochastic production frontier model this sign is negative. This positive sign is because the negative *kth* output is on the left-hand side of Eq. 4.

To account for unobserved heterogeneity we use fixed effects (d_i). Related to this, and as noted in the introductory section, we estimate Eqs. $2-4$ by adapting the estimation procedure in Chen *et al.* (2014) to the case of SAR dependence. This involves drawing on their assumption that the composed error (ε_{it}) has a closed skew normal distribution. For further discussion of this, see Appendix 2 for the maximum likelihood procedure to estimate Eqs. $2-4$. To simplify the notation in the presentation of this estimation procedure, we set out this procedure for the SAR cost frontier in Eq. 5.

$$c_{it} = \alpha + x_{it}\beta' + \delta \sum_{j=1}^{N_t} w_{ijt} c_{jt} + b_t + d_i + v_{it} + u_{it}. \tag{5}$$

In Eq. 5, the vector of observations z_{it} and the observations for the variables in TL in Eq. 2 are collected in x_{it}. x_{it} relates to the $1 \times A$ vector of non-spatial regressors (indexed $a \in 1, ..., A$) and β' is the associated $A \times 1$ vector of coefficients.

Whilst β_a represents the own effect of a change in x_{ait} on the *ith* bank, this is not the marginal effect of this change. This is because this change will also affect other banks via the SAR variable. The direct, indirect and total effects account for how this spatial interaction influences the effect of x_{ait}. Briefly turning to discuss these spatial effects as the asymmetric overall indirect spill-in and spill-out effects $-$ and the partitioning of these effects into impacts pertaining to immediate neighbors, neighbors' neighbors, etc.$-$ are central to the method to obtain the spill-in and spill-out Lerner indices. The partitioned indirect Lerner indices therefore indicate the spatial extent of the propagation of the market power spillovers from a bank to other banks. The spatial effects relate to the data generating process (DGP) in Eq. 6, which we obtain by stacking Eq. 5 across successive cross-sections and rearranging.

$$\mathbf{c}_t = (\mathbf{I}_t - \delta \mathbf{W}_t)^{-1} \left(\alpha \iota_t + \mathbf{X}_t \beta' + \mathbf{b}_t + \mathbf{d} + \mathbf{v}_t + \mathbf{u}_t \right), \tag{6}$$

where ι_t is the $N_t \times 1$ vector of ones; bold lower case letters are $N_t \times 1$ vectors; $\mathbf{X}_t$ denotes the $N_t \times A$ matrix; and $\mathbf{I}_t$ is the $N_t \times N_t$ identity matrix.

We obtain the partitioned indirect effects from Eq. 7. This involves expanding the spatial multiplier matrix, $(\mathbf{I}_t - \delta \mathbf{W}_t)^{-1}$, in Eq. 6 to give the series that premultiplies on the right-hand side of Eq. 7. We then differentiate the new form of Eq. 6 with respect to $\mathbf{x}_{at}$ (i.e., the *ath* column of $\mathbf{X}_t$) to obtain Eq. 7.

$$\begin{bmatrix} \frac{\partial c_1}{\partial x_{a,1}} & \cdots & \frac{\partial c_1}{\partial x_{a,N}} \\ \vdots & \ddots & \vdots \\ \frac{\partial c_N}{\partial x_{a,1}} & \cdots & \frac{\partial c_N}{\partial x_{a,N}} \end{bmatrix}_t = \left(\mathbf{I}_t + \delta \mathbf{W}_t + \delta^2 \mathbf{W}_t^2 + \delta^3 \mathbf{W}_t^3 + ... \right) \begin{bmatrix} \beta_a & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \beta_a \end{bmatrix}. \tag{7}$$

Multiplying each matrix in $\mathbf{I}_t + \delta \mathbf{W}_t + ...$ by the $N_t \times N_t$ matrix $\mathbf{D}_t = \text{diag}(\beta_a)$ yields the partitioned effects of $\mathbf{x}_a$ for the following orders of $\mathbf{W}_t$. $\mathbf{I}_t \mathbf{W}_t^0 \mathbf{D}_t = \mathbf{I}_t \mathbf{D}_t$: the main diagonal of $\mathbf{I}_t \mathbf{D}_t$ comprises direct effects for the *ith* bank that are net of the impacts of the spatial interaction (i.e., the usual own effects), where the off-diagonal elements are zero. $\delta \mathbf{W}_t^1 \mathbf{D}_t$: the off-diagonal elements of $\delta \mathbf{W}_t \mathbf{D}_t$ represent indirect effects that relate to the banks' 1st order (i.e., immediate) neighborhood sets, where the main diagonal elements are zero. $\delta^2 \mathbf{W}_t^2 \mathbf{D}_t$, $\delta^3 \mathbf{W}_t^3 \mathbf{D}_t$, etc: the main diagonal elements of these matrices represent further components of the overall direct effects that rebound to the *ith* bank from its 1st order, 2nd order, etc. neighbors (i.e., feedback, which in practice is typically small). The off-diagonal elements of these matrices represent indirect effects that relate to the banks' 2nd order, 3rd order, etc. neighborhood sets.

If we do not expand $(\mathbf{I}_t - \delta \mathbf{W}_t)^{-1}$, the right-hand side of Eq. 7 yields a matrix product comprising overall direct and indirect effects on the main and off diagonals. For ease we report means of the direct and (un)partitioned indirect effects across the banks and time periods. Summing the (un)partitioned direct and indirect effects yields the corresponding total effects. For period t, these indirect effects are the mean column and row sums of the off-diagonal elements of the relevant matrix. When there is a change in x_a, the former sum quantifies the mean spill-out from a bank to all the other relevant banks, while the latter sum quantifies the mean spill-in to a bank from these other banks. Whereas the sample mean spill-out and spill-in effects are symmetric, using the column and row sums of the off-diagonal elements in the relevant matrix for an individual bank (or the mean of these sums for any subset of the banks in the sample), we obtain asymmetric indirect spill-out and spill-in effects. The statistical inference for the direct, indirect and total coefficients is via simulation. For this, from the variance-covariance matrix we draw 200 Halton sequences of parameter values, with each value having a random component drawn from $\text{N}(0, 1)$.

The decomposition on the right-hand side of Eq. 8a yields the unpartitioned direct (Dir) and pairwise indirect (Ind) inefficiencies (e.g., Kutlu, 2018). By summing the pairwise Ind elements horizontally and vertically, we obtain the asymmetric unpartitioned indirect inefficiency spill-in and spill-out to and from a bank. By expanding $(\mathbf{I}_t - \delta \mathbf{W}_t)^{-1}$ in Eq. 8a, we obtain the partitioned decomposition of the direct and pairwise indirect inefficiencies in Eq. 8b for orders of $\mathbf{W}_t$, where for simplicity we drop the Dir and Ind superscripts. For $\mathbf{W}_t^1$, $\mathbf{W}_t^2$, etc. the pairwise indirect inefficiencies are summed horizontally and vertically. These partitioned Ind inefficiencies are also central to the method for the partitioned indirect Lerner indices to measure the spatial extent of the propagation of market power spillovers.

$$(\mathbf{I}_t - \delta \mathbf{W}_t)^{-1} \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix}_t = \begin{pmatrix} u_{11}^{Dir} + & \cdots & + u_{1N}^{Ind} \\ \vdots & \ddots & \vdots \\ u_{N1}^{Ind} + & \cdots & + u_{NN}^{Dir} \end{pmatrix}_t \tag{8a}$$

$$= \underbrace{\begin{pmatrix} u_{11} + & \cdots & +0 \\ \vdots & \ddots & \vdots \\ 0+ & \cdots & + u_{NN} \end{pmatrix}_t}_{\mathbf{W}_t^0} + \underbrace{\begin{pmatrix} 0+ & \cdots & + u_{1N} \\ \vdots & \ddots & \vdots \\ u_{N1} + & \cdots & +0 \end{pmatrix}_t}_{\mathbf{W}_t^1} + \underbrace{\begin{pmatrix} u_{11} + & \cdots & + u_{1N} \\ \vdots & \ddots & \vdots \\ u_{1N} + & \cdots & + u_{NN} \end{pmatrix}_t}_{\mathbf{W}_t^2} + ... \tag{8b}$$

Using the direct and (un)partitioned asymmetric indirect (spill-in and spill-out) measures of the inefficiencies, time period effects, errors and coefficients, and the non-spatial independent variables these coefficients pre-multiply, it follows from the DGP in Eq. 6 that we can construct direct and (un)partitioned asymmetric indirect functions. Note that these functions are constructed, and are not regressions, as the dependent variables are not observed. To obtain the direct and (un)partitioned asymmetric indirect time period effects and errors, we follow the method to obtain the corresponding inefficiencies.

To illustrate, in Eqs. $9 - 11$ we present the forms of the functions for direct cost $\left(c_{it}^{Dir}\right)$ and the two asymmetric (un)partitioned indirect costs $\left(c_{In,it}^{Ind} \text{ and } c_{Out,it}^{Ind}\right)$. We emphasize that Eqs. $9 - 11$ are constructed, and are not regressions, as we do not observe the dependent variables. We, therefore, use the right-hand sides of these equations to obtain predictions of c_{it}^{Dir}, $c_{In,it}^{Ind}$ and $c_{Out,it}^{Ind}$. Note that for simplicity in Eqs. $9-11$, we drop the notation used in Appendix 2 for the within transformation of the variables. Additionally, in the indirect functions, the partitioned and unpartitioned indirect spill-in and spill-out coefficients pre-multiply the same non-spatial variables, so for simplicity in Eqs. 10 and 11 we do not distinguish between unpartitioned and partitioned parameters, inefficiencies and errors. Unlike a SAR stochastic frontier model (e.g., Eq. 2), the direct and (un)partitioned indirect functions do not contain any spatial variables. This is because these spatial impacts are accounted for in the computation of the direct and (un)partitioned indirect parameters, inefficiencies and errors.

$$
\begin{aligned}
c_{it}^{Dir} &= \theta_i^{Dir}t + \frac{1}{2}\xi_i^{Dir}t^2 + \kappa_i^{Dir}q_{it}' + \psi_i^{Dir}m_{it}' + \frac{1}{2}q_{it}'\Gamma_i^{Dir}q_{it} + \frac{1}{2}m_{it}'\Upsilon_i^{Dir}m_{it} + \\
&\quad q_{it}'\Psi_i^{Dir}m_{it} + \varsigma_i^{Dir}q_{it}'t + \rho_i^{Dir}m_{it}'t + \zeta_i^{Dir}z_{it}' + b_t^{Dir} + v_{it}^{Dir} + u_{it}^{Dir}, \qquad (9)\\
c_{In,it}^{Ind} &= \theta_{In,i}^{Ind}t + \frac{1}{2}\xi_{In,i}^{Ind}t^2 + \kappa_{In,i}^{Ind}q_{it}' + \psi_{In,i}^{Ind}m_{it}' + \frac{1}{2}q_{it}'\Gamma_{In,i}^{Ind}q_{it} + \frac{1}{2}m_{it}'\Upsilon_{In,i}^{Ind}m_{it} + \\
&\quad q_{it}'\Psi_{In,i}^{Ind}m_{it} + \varsigma_{In,i}^{Ind}q_{it}'t + \rho_{In,i}^{Ind}m_{it}'t + \zeta_{In,i}^{Ind}z_{it}' + b_{In,t}^{Ind} + v_{In,it}^{Ind} + u_{In,it}^{Ind}, \qquad (10)\\
c_{Out,it}^{Ind} &= \theta_{Out,i}^{Ind}t + \frac{1}{2}\xi_{Out,i}^{Ind}t^2 + \kappa_{Out,i}^{Ind}q_{it}' + \psi_{Out,i}^{Ind}m_{it}' + \frac{1}{2}q_{it}'\Gamma_{Out,i}^{Ind}q_{it} + \frac{1}{2}m_{it}'\Upsilon_{Out,i}^{Ind}m + \\
&\quad q_{it}'\Psi_{Out,i}^{Ind}m_{it} + \varsigma_{Out,i}^{Ind}q_{it}'t + \rho_{Out,i}^{Ind}m_{it}'t + \zeta_{Out,i}^{Ind}z_{it}' + b_{Out,t}^{Ind} + v_{Out,it}^{Ind} + u_{Out,it}^{Ind}. \qquad (11)
\end{aligned}
$$

In Eqs. $9-11$, $\theta_i^{Dir}t + ... + \rho_i^{Dir}m_{it}'t$, $\theta_{In,i}^{Ind}t + ... + \rho_{In,i}^{Ind}m_{it}'t$ and $\theta_{Out,i}^{Ind}t + ... + \rho_{Out,i}^{Ind}m_{it}'t$ are the direct and indirect spill-in and spill-out translog functions that correspond to the own $TL\left(t, q_{it}, m_{it}\right)$ in Eq. 2. Moreover, in Eqs. $9-11$ a direct parameter is denoted by the superscript Dir and a indirect spill-in (spill-out) parameter is denoted by the superscript Ind and subscript In (Out). Vectors of direct and indirect spill-in and spill-out parameters pre-multiply vectors denoted by $'$. Γ, Υ and Ψ are used to denote matrices of direct and indirect spill-in and spill-out coefficients on the interactions and squared terms that relate to the outputs and input prices. We also distinguish the vector of own parameters γ' on z_{it} in Eq. 2 from the vectors of direct and indirect spill-in and spill-out parameters on this variable by using ζ to denote these parameters in Eqs. $9-11$. We next in 2.2 turn to discuss how we use Eqs. $9-11$ to obtain the (direct-own and (un)partitioned asymmetric indirect) Lerner indices.

2.2 Step 2: Measuring market power spill-ins and spill-outs

Computing Lerner indices for banks has involved using average revenues at different levels of disaggregation as measures of the prices of a bank's aggregated and disaggregated outputs. Shaffer and Spierdijk (2020) classify the large non-spatial Lerner index banking literature into three groups. The first group reports an aggregate Lerner index for each bank, where in many studies in this group this index is based on an aggregate average revenue measure that uses a bank's total assets as a proxy for its aggregate output. The second group reports a disaggregated Lerner index for each bank output and the third reports a weighted average of these disaggregated indices for each bank. Our study extends the second and third groups, where we first set out how we calculate the asymmetric (un)partitioned spill-in and spill-out Lerner indices for individual outputs. At the end of this subsection when we pull together the different parts of the exposition, we discuss how we obtain a weighted average of the spill-in (spill-out) indices across a bank's outputs. In short, this involves applying the approach for consistent aggregation of non-spatial Lerner indices (Shaffer and Spierdijk, 2020).

In the empirical analysis, to assess the impact on the results, we report (un)partitioned spill-in and spill-out Lerner indices with and without inefficiency adjustments. As we discuss in detail further in this subsection, our extension to compute market power spill-ins and spill-outs requires a number of (un)partitioned spill-in and spill-out measures, which are unobserved and must, therefore, be predicted (with and without adjustment for inefficiency). The approach to obtain the Lerner index for partitioned market power spill-ins and spill-outs is the same as for the corresponding unpartitioned index. We do not therefore distinguish between an unpartitioned/partitioned index in the below method. We present this method in terms of the spill-in Lerner index, $L^{Ind}_{In,kit}$, which we calculate using Eq. 12. Note that whilst we focus on market power spill-ins and spill-outs, our modeling framework is entirely consistent with the non-spatial Lerner index. This is because our approach also yields the direct Lerner index (L^{Dir}_{kit}) measure of a bank's own market power, where L^{Dir}_{kit} is calculated in the same type of way as $L^{Ind}_{In,kit}$. Accordingly, L^{Dir}_{kit} is < 1 as all the above standard coverage in the introductory section of the non-spatial Lerner index (Eq. 1) applies.

$$L^{Ind}_{In,kit} = \frac{\frac{R^{Ind}_{In,kit}}{Q^{Ind}_{In,kit}} - MC^{Ind}_{In,kit}}{\frac{R^{Ind}_{In,kit}}{Q^{Ind}_{In,kit}}}, \tag{12}$$

$$MC^{Ind}_{In,kit} = \frac{C^{Ind}_{In,it}}{Q^{Ind}_{In,kit}} \frac{\partial c^{Ind}_{In,it}}{\partial q_{kit}}. \tag{13}$$

Compared to the simple standard own case of $\{R_{kit}, Q_{kit}, MC_{kit}\} > 0$ and $L_{kit} < 1$ (and using the corresponding terminology from our modeling framework $\left\{R^{Dir}_{kit}, Q^{Dir}_{kit}, MC^{Dir}_{kit}\right\} > 0$ and $L^{Dir}_{kit} < 1$), consideration of spillovers leads to a larger set of possible values of $L^{Ind}_{In,kit} \lesseqgtr 0$. Note that due to our interest in spillovers, we overlook the case of no market power spill-ins ($L^{Ind}_{In,kit} = 0$). $L^{Ind}_{In,kit} \lesseqgtr 0$ is due to the larger set of possible values of $\left\{R^{Ind}_{In,kit}, Q^{Ind}_{In,kit}, MC^{Ind}_{In,kit}\right\} \lesseqgtr 0$, where here we again overlook a zero value for any of these three spill-ins. When such a spill-in is positive, this means that the variable for the *ith* bank tends to move in the same direction as the corresponding variable of other spatial interdependent banks. Applying reasoning from the spatial literature, this positive spatial correlation is consistent with banks being impacted by common economic phenomena at different spatial scales, such as industrywide regulation, the Federal Open Market Committee's (FOMC's) setting of the federal funds rate, market conditions, and headline changes in economies at the city, state, regional and national levels. Conversely, when any of the aforementioned three spill-ins are negative, this means that the variable for the *ith* bank tends to move in the opposite direction to the corresponding variable of other spatially interdependent banks. In the spatial literature, this negative spatial correlation is attributed to the effects of spatial competition. Note though that whilst $L^{Ind}_{In,kit} \lesseqgtr 0$ represents a larger set of possible values, we will see in the empirical results that there are some clear and plausible patterns in the market power spillovers.

As we do not observe the spill-in measures on the right-hand side of Eq. 12 for $L^{Ind}_{In,kit}$, we predict these measures. In the same type of way, we obtain L^{Dir}_{kit} and $L^{Ind}_{Out,kit}$ by predicting the corresponding measures. With regard to how we compute $MC^{Ind}_{In,kit}$ in Eq. 13, having estimated the SAR stochastic cost frontier in Eq. 2, we use the corresponding DGP to obtain the (un)partitioned indirect spill-in coefficients, inefficiencies and errors (see subsection 2.1). We then use these coefficients, inefficiencies and errors to construct the (un)partitioned spill-in logged cost function in Eq. 10. From Eq. 10, we get $MC^{Ind}_{In,kit} = \partial c^{Ind}_{In,it}/\partial q_{kit}$ for each product.

To predict the (un)partitioned variable $C^{Ind}_{In,it}$ in levels for the *ith* bank in Eq. 13, we avoid the complication of obtaining this using the prediction of the (un)partitioned $c^{Ind}_{In,it}$ from Eq. 10, as this would involve reversing a number of data transformations. Instead, we follow the approach to obtain the direct and indirect spill-in and spill-out inefficiencies, which involves using the SAR coefficient from the estimated Eq. 2 and the spatial multiplier matrix, $(\mathbf{I}_t - \delta \mathbf{W}_t)^{-1}$. We obtain predictions in levels of C^{Dir}_{it}, $C^{Ind}_{In,it}$ and $C^{Ind}_{Out,it}$ by decomposing $(\mathbf{I}_t - \delta \mathbf{W}_t)^{-1} C_{it}$. The relevant element on the main diagonal of this decomposition is the prediction of C^{Dir}_{it} and the relevant off-diagonal elements are summed horizontally and vertically to obtain the predictions of $C^{Ind}_{In,it}$ and $C^{Ind}_{Out,it}$, respectively. These predictions are then adjusted upwards using the bank level direct and indirect spill-in and spill-out cost inefficiencies.

Turning to how we obtain the predictions of the variables $Q^{Ind}_{In,kit}$ and $R^{Ind}_{In,kit}$ in levels for the kth product in Eq. 12. To obtain the former and corresponding predictions of Q^{Dir}_{kit} and $Q^{Ind}_{Out,kit}$ for all the K products of the ith bank, we first estimate K specifications of the ODF in Eq. 4. Using the SAR coefficient from the fitted ODF with the kth output as the dependent variable and the above approach that yields the predictions of C^{Dir}_{it}, $C^{Ind}_{In,it}$ and $C^{Ind}_{Out,it}$, we obtain the predictions of Q^{Dir}_{kit}, $Q^{Ind}_{In,kit}$ and $Q^{Ind}_{Out,kit}$ by decomposing $(\mathbf{I}_t - \delta \mathbf{W}_t)^{-1} Q_{kit}$. The predictions of Q^{Dir}_{kit}, $Q^{Ind}_{In,kit}$ and $Q^{Ind}_{Out,kit}$ for all the K products of the ith bank are then adjusted downwards using the bank level direct and indirect spill-in and spill-out output distance inefficiencies. As these inefficiencies are based on a radial contraction (expansion) of all of a bank's K outputs, these inefficiencies are used to adjust the predictions of Q^{Dir}_{kit}, $Q^{Ind}_{In,kit}$ and $Q^{Ind}_{Out,kit}$ for all K products of the ith bank.[4]

Along the same lines, using the SAR coefficient from the fitted alternative revenue function in Eq. 3, we decompose $(\mathbf{I}_t - \delta \mathbf{W}_t)^{-1} R_{it}$ to obtain bank level predictions of R^{Dir}_{it}, $R^{Ind}_{In,it}$ and $R^{Ind}_{Out,it}$. Next, we obtain the direct and indirect spill-in and spill-out revenue shares for the ith bank's kth product. These shares are the first order derivatives of the bank level direct and indirect spill-in and spill-out alternative revenue functions with respect to the kth output (see Eqs. $9 - 11$ for the corresponding product level cost functions). Multiplying the bank level predictions of R^{Dir}_{it}, $R^{Ind}_{In,it}$ and $R^{Ind}_{Out,it}$ by these direct and indirect spill-in and spill-out revenue shares yields the predictions of the product level direct and asymmetric indirect revenues (R^{Dir}_{kit}, $R^{Ind}_{In,kit}$ and $R^{Ind}_{Out,kit}$). These product level revenue predictions are then adjusted using the bank level direct and indirect spill-in and spill-out altenative revenue inefficiencies. As these inefficiencies are based on a radial expansion (contraction) of the real monetary volumes of all of a bank's K outputs, these inefficiencies are used to adjust the predictions of R^{Dir}_{kit}, $R^{Ind}_{In,kit}$ and $R^{Ind}_{Out,kit}$ for all K products of the ith bank.

For the non-spatial case, the average of the Lerner indices for a bank's products weighted by their revenue shares is the theoretically consistent aggregated bank level index (Shaffer and Spierdijk, 2020). As our approach to calculate the spill-in and spill-out indirect product level Lerner indices follows the non-spatial case, we apply this theoretically consistent aggregation to calculate indirect bank level Lerner indices.

[4] In line with a well-known result in production theory, in the empirical analysis these output distance inefficiencies are insensitive to which of the K outputs is the dependent variable. It does not matter, therefore, which of the K specifications of the ODF we obtain the inefficiencies from to adjust Q^{Dir}_{kit}, $Q^{Ind}_{In,kit}$ and $Q^{Ind}_{Out,kit}$.

3 Empirical analysis

3.1 Data and the spatial weights matrix

To obtain the market power spill-ins and spill-outs, we first estimate the spatial specifications of the cost, alternative revenue and output distance stochastic frontiers (Eqs. $2-4$). There is an exact correspondence between these models as we draw on a rich data source (Federal Deposit Insurance Corporation (FDIC) data from the Call Reports) to construct three datasets comprising corresponding bank-period observations. Each dataset is large as it comprises $45,759$ quarterly observations for large commercial U.S. banks for the period 1994:Q1 $-$ 2022:Q4. In contrast to the balanced panel data used in previous spatial banking studies (Glass and Kenjegalieva, 2019; 2023; Glass *et al*, 2020a; 2020b), our panel data is unbalanced. This is important as it accounts for the marked consolidation among U.S. banks due to bank failures and mergers and acquisitions, which in our sample involves a fall from a high of 591 banks (1994:Q2) to a low of 272 in the aftermath of the financial crisis (2010:Q4).

A key factor that affects the spatial dependence between banks is their branch networks as they indicate which banks operate in the same geographical markets. Accordingly, and as we discuss in more detail below, the a priori specification of the spatial weights are based on the geographical overlap of their branch networks. Unlike the extant balanced panel data applications of the spatial weights matrix to banking (Glass and Kenjegalieva, 2019; 2023; Glass *et al*, 2020a; 2020b), which, as is common in applied spatial econometrics, use a fixed matrix across the study period (e.g., average of the per period matrices), here this matrix is time-varying due to the panel data being unbalanced. Hence, in contrast to the above spatial banking studies, our matrix reflects the evolution of each bank's branch network over the study period.

Capturing this evolution is important as there have been big changes to these networks. This is because from June 1997, the 1994 Riegle-Neal Interstate Banking and Branching Efficiency Act allowed banks to expand their branch networks outside their state of origin. Due to some deregulations in the 1980s and early 1990s, and some states implementing the Riegle-Neal Act in advance of the deadline, there were banks with branches outside of their state of origin prior to June 1997. Our study period, therefore, starts in 1994:Q1 as (i) this is the first quarter when the geographical information on branch networks is available; and (ii) this allows us to capture the evolution of branch networks in response to the Riegle-Neal Act, as this quarter represents the first early implementation of the Act by a state (Alaska) (Dick, 2006). Of the commercial U.S. banks, we consider the large ones as they have the largest branch networks, so there will be a greater overlap between their networks, which is consistent with greater spatial dependence. This also allows us to focus on how the branch networks of large banks responded to the Riegle-Neal Act, as these banks were best equipped for branch network expansion. We define a large U.S. bank using the same real total assets threshold as Berger and Roman (2017). Having used the GDP deflator to convert the total assets of the banks into 2012:Q4 U.S. dollars, we define a bank as large if it has real total assets greater than \$3 billion at any point in the study period. No smaller banks are included as they will likely have a different best practice frontier to large banks.

Table 1: Description of the variables and summary statistics

Variable description	Variable notation	Mean	Std. Dev.
Total operating cost (000s of 2012:Q4 U.S. dollars): Sum of salaries, interest expenses on deposits, and expenditure on fixed assets and premises	c	$155,540.7$	$698,792.6$
Total revenue (000s of 2012:Q4 U.S. dollars): Sum of non-interest income minus service fees on deposits and income from loans and leases, and securities	r	$395,332.1$	$1,700,855.1$
Input prices			
Cost of deposits: Interest expenses on deposits divided by deposits	m_1	0.005	0.004
Cost of labor: Salaries divided by the number of full-time equivalent employees	m_2	19.526	11.480
Cost of fixed assets and premises: Expenditure on fixed assets and premises divided by their value	m_3	0.830	57.384
Inputs			
Total deposits (000s of 2012:Q4 U.S. dollars)	s_1	$17,026,147.0$	$92,684,257.7$
Number of full-time equivalent employees	s_2	$3,497.7$	$15,817.9$
Fixed assets and premises (000s of 2012:Q4 U.S. dollars)	s_3	$206,004.8$	$884,389.5$
Outputs			
Total loans and leases (000s of 2012:Q4 U.S. dollars)	q_1	$12,860,747.8$	$59,134,004.2$
Total securities (000s of 2012:Q4 U.S. dollars)	q_2	$4,618,489.9$	$26,473,630.6$
Off-balance sheet (OBS) items (000s of 2012:Q4 U.S. dollars): Measured as non-interest income capitalization credit equivalents of OBS items, where we calculate this equivalence measure using the approach in Boyd and Gertler (1994)	q_3	$21,203,926.2$	$164,187,005.1$
Non-spatial environmental variables (z's)			
Loan loss allowance as a share of loans and leases	LLA	0.017	0.021
Tier 1 capital ratio: Tier 1 capital divided by total assets	$Tier1CR$	0.091	0.048
Tier 2 capital ratio: Tier 2 capital divided by total assets	$Tier2CR$	0.011	0.008
Equity ratio: Total equity capital divided by total assets	$Equity$	0.104	0.052
Asset quality: Ratio of non-performing loans to total loans	NPL	0.012	0.023
Scope of the bank loan portfolio: Hirschman-Herfindahl Index (HHI) across a bank's real estate loans, farm loans, commercial and industrial loans, loans to individuals and other loans as ratios of total loans	HHI	0.543	0.178
Age: Number of years the institution has been established	Age	71.13	50.53
Security share: Securities as a share of of total assets	$Security$	0.215	0.138

Throughout we use FDIC Call Report data for the variables, where we draw on the widely-used intermediation approach (Sealey and Lindley, 1977) to settle on which variables are outputs, inputs and input prices. Apart from the input prices as they are ratios, the GDP deflator is used to deflate monetary variables to 2012:Q4 U.S. dollars. We then obtain the data for all the flow (income and expenditure) variables by first differencing the observations for quarters $2-4$ (Wheelock and Wilson, 2018). For the summary statistics, descriptions of the variables and notation, see table 1.

We take logs of all the continuous variables in table 1 and then mean adjust. We mean adjust so we can interpret the own coefficients on the first order variables in the TL functions in Eqs. $2-4$ (and the direct, (un)partitioned indirect and total coefficients on these variables) as elasticities at the sample mean. We are then in a position to estimate the alternative revenue function. This is because production theory does not require this function to be homogenous of degree one in input prices, so none of the variables are normalized (Berger *et al.*, 1996;

Wheelock and Wilson, 2018). In contrast, this homogeneity is a property of the cost function. We, therefore, use m_3 as the normalizing factor for c, m_1 and m_2. Moreover, as the ODF function assumes that a bank aims maximize its three outputs using a given quantity of inputs, we use the left-hand side output as the normalizing factor for the two right-hand side outputs.

We use the same a priori normalized specification of $\mathbf{W}_t$ to estimate each spatial stochastic frontier. Ahead of the presentation of our specification of $\mathbf{W}_t$, we note that it was influenced by the following two factors that informed the spatial weights in the analysis of U.S. banks by Glass and Kenjegalieva (2023). First, following the vast majority of the spatial econometrics literature, the spatial weights in Eqs. $2-4$ are exogenous. In line with this, we specify $\mathbf{W}_t$ using a measure of the geographical links between banks' branch operations. Second, as we present an economic application, we recognize that the spatial weights matrix should have some economic foundation (Corrado and Fingleton, 2012).

At the outset we ruled out specifying $\mathbf{W}_t$ using the distances between banks' headquarters, as the locations of their headquarters would not reflect the geographical evolution of the banks' branch networks following the Riegle-Neal Act. Another possibility, which we ultimately did not pursue, is to use data on branch deposit levels to construct off-diagonal weights that represent the economic connectivity of banks' branch networks. A possible approach to reflect the economic connectivity of the ij-th banks' branch networks is to use the ratio of the jth and ith banks' branch deposits across the latter's branch network as the ij-th weight. While a measure of the economic connectivity between banks' branch networks is informative, we do not use this ratio (or any other branch deposit based measure) to populate $\mathbf{W}_t$. This is because these economic distance based weights would likely be endogenous and accounting for this by incorporating an appropriate method from the general non-frontier spatial literature to our new spatial stochastic frontier framework would be no small development. This development is thus outside the scope of this paper and an area for further work. Rather than use economic distance based weights without accounting for their possible endogeneity, we exercise caution using the following approach, which represents a halfway house between geographical and economic weights.

As is standard, the elements on the main diagonal of the pre-normalized spatial weights matrix $\widetilde{\mathbf{W}}_t$ are set to zero to rule out self-influence. To calculate the off-diagonal elements of this matrix, we use the available annual mid-year FDIC information from the 'Summary of Deposits' on the locations of the banks' branches. Annual data is also available for the variables in table 1, but we instead favor a richer, higher frequency quarterly analysis. To combine the quarterly data for the variables with the annual branch locations, we apply these locations to each quarter in a year. To calculate each off-diagonal element in $\widetilde{\mathbf{W}}_t$, we sum across 51 territories (50 states and the District of Columbia) the ratio of the number of jth bank branches in a territory to the number of ith bank branches. This sum, therefore, represents the jth bank's relative branch intensity. Based on this, we view $\widetilde{\mathbf{W}}_t$ as being a halfway house between exclusively geographical and exclusively economic weights matrices. This is based on $\widetilde{\mathbf{W}}_t$ being geographical in nature, which is consistent with the weights being exogenous; and the branch geography on which $\widetilde{\mathbf{W}}_t$ is based on underpinning the economic links between banks in the form of their branch deposits in the same markets.

Summarizing, we formally calculate the elements of $\widetilde{\mathbf{W}}_t$ as follows, where the territories are indexed $s \in 1, ..., 51$.

$$\widetilde{w}_{ijt} = \begin{cases} \sum_{s=1}^{51} \frac{\text{Number of } jth \text{ bank branches in } s \text{ in period } t}{\text{Number of } ith \text{ bank branches in } s \text{ in period } t} & \text{for } i \neq j \\ 0 & \text{for } i = j \end{cases} . \tag{14}$$

We then normalize $\widetilde{\mathbf{W}}_t$ by its largest eigenvalue to obtain the $\mathbf{W}_t$ we use in the estimations. This normalizing factor has the advantage of preserving the original information on the relative branch network intensities, as it leaves the proportional relationships between the elements of $\widetilde{\mathbf{W}}_t$ unchanged. In contrast, the common alternative of row-normalizing would not preserve this information.

3.2 Estimated models and (in)efficiencies

In table 2, we present estimates of the models in Eqs. $2-4$: namely, SAR stochastic cost, alternative revenue and output distance frontiers, where for the latter the left-hand side output is securities ($-q_2$). See table A in Appendix 3 for the estimates of the other two SAR output distance frontiers when the left-hand side output is loans ($-q_1$) and off-balance sheet items ($-q_3$). We can see from the fitted cost and alternative revenue models that the coefficients on the first order outputs and input prices have the expected positive signs. From the reported distance functions, the coefficients on the first order outputs are, as expected, positive. As the left-hand side outputs in the distance functions are negative, with one exception (the small and insignificant coefficient on fixed assets and premises (s_3) in the distance function when the left-hand side output is $-q_3$), the fitted coefficients on the first order inputs in these models have the expected negative signs.

It is not surprising that the coefficient on s_3 is not significant in the distance function when the left-hand side output is $-q_3$ (and $-q_1$). This is because this is consistent with the rise in online banking leading to a decline in the role of brick and mortar branches, which has resulted in a wave of branch closures. In the cost and alternative revenue models, all the coefficients on the first order outputs and input prices are significant at the 1% level. From the reported ODFs, and with the exception of the two aforementioned coefficients on s_3, we observe significant coefficients at the 1% level on the first order inputs and outputs. Furthermore, based on the magnitudes of the returns to scale for U.S. banks in the extant literature, the estimates of these returns at the sample mean from our cost, alternative revenue and ODFs are reasonable (0.91, 0.95 and $1.03-1.07$, respectively).

Table 2: Selected SAR stochastic frontier models

SAR stochastic cost frontier				**SAR stochastic alternative revenue frontier**				**SAR stochastic output distance frontier: Left-hand side output $-q_2$**			
	Model coeff		*Model coeff*		*Model coeff*		*Model coeff*		*Model coeff*		*Model coeff*
q_1	0.681^{***}	q_3t	$0.043 \times 10^{-2***}$	q_1	0.704^{***}	q_3m_1	-0.009^{***}	s_1	-0.846^{***}	s_3t	$0.028 \times 10^{-2***}$
q_2	0.132^{***}	m_1t	-0.001^{***}	q_2	0.121^{***}	q_3m_2	0.021^{***}	s_2	-0.172^{***}	q_1t	-0.001^{***}
q_3	0.100^{***}	m_2t	-0.007×10^{-2}	q_3	0.121^{***}	q_3m_3	0.012^{***}	s_3	-0.009^{***}	q_3t	0.001^{***}
m_1	0.395^{***}	LLA	0.881^{***}	m_1	0.124^{***}	t	-0.043^{***}	q_1	0.743^{***}	LLA	0.614^{***}
m_2	0.531^{***}	$Tier1CR$	-0.464^{***}	m_2	0.151^{***}	t^2	0.001^{***}	q_3	0.079^{***}	$Tier1CR$	-1.083^{***}
q_1^2	0.031^{***}	$Tier2CR$	-1.885^{***}	m_3	0.031^{***}	q_1t	-0.005^{***}	s_1^2	-0.054^{***}	$Tier2CR$	-0.368^{**}
q_2^2	0.015^{***}	$Equity$	0.395^{***}	q_1^2	0.044^{***}	q_2t	-0.001^{***}	s_2^2	-0.028^{***}	$Equity$	0.048
q_3^2	0.007^{***}	NPL	0.989^{***}	q_2^2	0.012^{***}	q_3t	0.002^{***}	s_3^2	-0.002^{***}	NPL	0.804^{***}
q_1q_2	-0.008^{***}	HHI	-0.075^{***}	q_3^2	0.010^{***}	m_1t	0.001^{***}	s_1s_2	0.106^{***}	HHI	-0.153^{***}
q_1q_3	-0.028^{***}	Agw	-0.055^{***}	q_1q_2	-0.023^{***}	m_2t	-0.004^{***}	s_1s_3	-0.010^{***}	Age	-0.002
q_2q_3	0.004^{***}	$Security$	0.303^{***}	q_1q_3	-0.037^{***}	m_3t	-0.002^{***}	s_2s_3	-0.004^{*}	$Security$	0.024
m_1^2	0.030^{***}	W_tc	-0.160^{***}	q_2q_3	0.002^{***}	LLA	1.089^{***}	q_1^2	0.027^{***}	$W_t\,(-q_2)$	-0.264^{***}
m_2^2	0.052^{***}	σ	0.053^{***}	m_1^2	0.021^{***}	$Tier1CR$	-0.022	q_3^2	0.009^{***}	σ	0.050^{***}
m_1m_2	-0.097^{***}	λ	0.890^{***}	m_2^2	0.044^{***}	$Tier2CR$	-0.222	q_1q_3	-0.021^{***}	λ	0.675^{***}
q_1m_1	0.011^{***}			m_3^2	-0.004^{***}	$Equity$	0.825^{***}	s_1q_1	0.002		
q_1m_2	-0.036×10^{-2}	LL	$31,660.7$	m_1m_2	0.059^{***}	NPL	-0.055	s_1q_3	0.012^{***}	LL	$31,577.3$
q_2m_1	0.013^{***}			m_1m_3	-0.010^{***}	HHI	-0.038^{***}	s_2q_1	0.006^{***}		
q_2m_2	-0.006^{***}			m_2m_3	-0.014^{***}	Age	0.073^{***}	s_2q_3	-0.014^{***}		
q_3m_1	-0.003^{***}			q_1m_1	-0.001	$Security$	0.102^{***}	s_3q_1	0.013^{***}		
q_3m_2	-0.003^{**}			q_1m_2	-0.005^{*}	W_tr	0.136^{***}	s_3q_3	-0.007^{***}		
t	$-0.089 \times 10^{-2***}$			q_1m_3	0.037×10^{-2}	σ	0.027^{***}	t	-0.002^{***}		
t^2	$-0.001 \times 10^{-2***}$			q_2m_1	0.007^{***}	λ	0.878^{***}	t^2	$0.001 \times 10^{-2***}$		
q_1t	$-0.018 \times 10^{-2**}$			q_2m_2	0.003			s_1t	$-0.025 \times 10^{-2***}$		
q_2t	$0.021 \times 10^{-2***}$			q_2m_3	-0.011^{***}	LL	$31,581.8$	s_2t	-0.002^{***}		

Note: *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Turning to the results for the environmental variables, where we are particularly interested in the estimates of the SAR coefficients. From table 2 and table A, we observe that the SAR coefficient in each model is significant at the 1% level, which supports our modeling approach to account for the spatial dependence in the datasets. We are the first to report a spatial stochastic distance function for banks, which is an interesting line of inquiry as we find that the SAR variables in the three fitted ODFs have marked negative effects. This is consistent with output competition between banks with overlapping branch networks (i.e., between first order neighboring banks), where the entry and exit of banks to and from the sample is an aspect of this competition which our unbalanced panel data captures. While the sample comprises only large commercial U.S. banks, the negative spatial correlation between neighboring banks' corresponding outputs that the SAR parameters in the distance functions are picking up may also reflect the marked variation in bank sizes in our sample, as measured by their outputs. Such marked variation is evident as in table 1 the standard deviation of each output is much larger than its mean.

An interesting picture emerges when we relate the above discussion of spatial output competition to the SAR coefficients from the cost and alternative revenue models. The SAR coefficient is also negative in the cost model and while non-negligible it is smaller in absolute magnitude than the corresponding estimates from the ODFs. This suggests that while there is spatial cost competition between banks with overlapping branch networks, which will be impacted by the aforementioned output competition, this cost competition is not as strong. The implication is that the SAR coefficient in the cost model is capturing some partial offsetting of the effect of the output competition on the cost rivalry between first order neighboring banks, which is consistent with costs being affected by factors other than outputs, e.g., input prices. This partial offsetting points to a positive spatially correlated component of the banks' costs that relates to these other factors. This is consistent with these other factors across the banks being similarly impacted due to the banks' exposure to common phenomena, such as the FOMC's setting of the federal funds rate and changes in labor and real estate markets at different spatial scales, i.e., at the city, state, regional and national levels. Further behavior that would contribute to a positive spatially correlated component of banks' costs is banks setting deposit rates that mimic their rivals' rates.

The SAR coefficient in the fitted alternative revenue model is positive. As this parameter will be influenced by the above spatial output competition, this positive coefficient suggests that the effect of this competition is dominated by a positive spatially correlated component of the banks' revenues that relates to factors other than output quantities, e.g., prices of outputs and inputs. Again this spatially correlated component would be consistent with these other factors being similarly impacted due to the banks' exposure to common phenomena (e.g., FOMC federal funds rate setting and market conditions), as well as banks mimicking certain behavior of rivals, such as their loan rates.

Turning to the findings for the other environmental variables. (i) In the cost, output distance and alternative revenue models, the LLA parameters are positive and significant. This reflects that there is a higher cost associated with higher LLA, and recalling that the left-hand side outputs are negative, to cover these higher costs banks reduce their outputs, but at the same time more risky loans are a source of higher revenue. (ii) The significant $Tier1CR$ and $Tier2CR$ parameters are negative, which we suggest is because banks with higher capital ratios are more stable and associated with relatively lower costs and higher outputs. (iii) With the exception of the ODF where the left-hand side output is securities ($-q_2$), the *Equity* parameter is positive and significant. This is consistent with relatively more equity finance at higher revenue banks, and equity being a more costly source of finance, leading to banks with relatively high equity ratios covering this additional cost by reducing two outputs (loans and off-balance sheet items). (iv) In all the models apart from the alternative revenue function, the NPL parameter is positive and significant, which reflects the higher costs associated with more bad loans and suggests that banks reduce their outputs to cover these additional costs. (v) The coefficients on the HHI are all negative and significant. This indicates that the specialization associated with a less diversified loan portfolio reduces costs and also revenues. Our results suggest that the latter relates to lower output prices (e.g., lower loan rates), as the ODFs indicate that a more specialized loan portfolio enables banks to channel their resources to increase all three outputs. (vi) The *Age* coefficient is negative and significant in the ODF when the left-hand side output is off-balance sheet items ($-q_3$), indicating that older banks are more engaged in this activity. The significant *Age* coefficients in the cost and alternative revenue models are negative and positive, where the lower costs of older banks may be due to their established systems and customer networks, and their higher revenues may be the result of greater trust in older banks because of the reputation they have built up over a longer period. (vii) Finally, we find that the coefficient on *Security* is positive and significant in the cost and alternative revenue models, indicating that, on average, if a bank increases its security ratio, rather than, for instance, increasing its traditional lending, its costs and revenue will both rise. The *Security* parameters are also positive and significant in the ODFs when the left-hand side outputs are $-q_1$ and $-q_3$. This indicates that, on average, an increase in a bank's security ratio is associated with falls in its other outputs (loans and off-balance sheet items).

We next provide a snapshot of the direct, indirect and total effects. In table 3, we present these mean effects for (a) the first order outputs and input prices from the cost and alternative revenue models; and (b) the first order outputs and inputs from the ODFs. With regard to the interpretation of these reported effects, they are all elasticities at the sample mean. As we report means of the partitioned and overall indirect effects across the banks, these effects represent symmetric spill-ins and spill-outs from and to a bank's 1st−3rd order neighbors and the other $N-1$ banks in the sample, respectively. To obtain the asymmetric spill-in and spill-out Lerner indices in 3.3, we use the asymmetric partitioned and overall indirect spill-in and spill-out effects for individual banks. We can see that the significance and magnitudes of the direct effects in table 3 are the same as we observe for the corresponding coefficients in table 2 and table A. This indicates that there is no spatial feedback in the direct effects, which is not unexpected as when feedback is observed (i.e., the above corresponding parameters differ) it is usually negligible. All the overall indirect effects are significant in table 3 with the exception of those for fixed assets and premises (s_3) from the ODFs when the left-hand side outputs are $-q_1$ and $-q_3$. The reason we give for these insignificant indirect results for s_3 is the same as we gave above for the same type of findings for s_3 in table A in Appendix 3. For the models where

Table 3: Mean direct, indirect and total coefficients

Variable	Direct coeff	Indirect spillover coeff				Total coeff
		1st order	2nd order	3rd order	Overall	
			SAR stochastic cost frontier			
q_1	0.681***	−0.043***	0.005***	−0.001***	−0.039***	0.642***
q_2	0.132***	−0.008***	0.001***	-0.011×10^{-2}***	−0.007***	0.125***
q_3	0.100***	−0.006***	0.001***	−0.009***	−0.006***	0.095***
m_1	0.395***	−0.025***	0.003***	−0.034***	−0.022***	0.373***
m_2	0.531***	−0.033***	0.004***	−0.045***	−0.030***	0.501***
			SAR stochastic alternative revenue frontier			
q_1	0.704***	0.038***	0.004***	0.037×10^{-2}***	0.042***	0.746***
q_2	0.121***	0.006***	0.001***	0.006×10^{-2}***	0.007***	0.128***
q_3	0.121***	0.007***	0.001***	0.006×10^{-2}***	0.007***	0.128***
m_1	0.124***	0.007***	0.001***	0.007×10^{-2}***	0.007***	0.132***
m_2	0.151***	0.008***	0.001***	0.008×10^{-2}***	0.009***	0.160***
m_3	0.031***	0.002***	0.016×10^{-2}***	0.002×10^{-2}***	0.002***	0.033***
			SAR stochastic output distance frontier: Left-hand side output $-q_1$			
s_1	−0.850***	0.077***	−0.013***	0.002***	0.066***	−0.784***
s_2	−0.183***	0.017***	−0.003***	0.047×10^{-2}***	0.014***	−0.169***
s_3	−0.003	0.030×10^{-2}	-0.005×10^{-2}	0.001×10^{-2}	0.026×10^{-2}	−0.003***
q_2	0.170***	−0.015***	0.003***	-0.043×10^{-2}***	−0.013***	0.157***
q_3	0.079***	−0.007***	0.001***	-0.020×10^{-2}***	−0.006***	0.073***
			SAR stochastic output distance frontier: Left-hand side output $-q_2$			
s_1	−0.846***	0.088***	−0.017***	0.003***	0.074***	−0.772***
s_2	−0.172***	0.018***	−0.003***	0.001***	0.015***	−0.157***
s_3	−0.009***	0.001***	-0.018×10^{-2}***	0.004×10^{-2}***	0.001***	−0.009***
q_1	0.743***	−0.077***	0.015***	−0.003***	−0.065***	0.678***
q_3	0.079***	−0.008***	0.002***	-0.031×10^{-2}***	−0.007***	0.072***
			SAR stochastic output distance frontier: Left-hand side output $-q_3$			
s_1	−0.860***	0.100***	−0.021***	0.005***	0.083***	−0.777***
s_2	−0.209***	0.024***	−0.005***	0.001***	0.020***	−0.189***
s_3	0.002	-0.024×10^{-2}*	0.005×10^{-2}*	-0.001×10^{-2}*	-0.020×10^{-2}	0.002
q_1	0.774***	−0.090***	0.019***	−0.004***	−0.075***	0.700***
q_2	0.153***	−0.018***	0.004***	−0.001***	−0.015***	0.138***

Note: *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

the SAR coefficient is negative (cost and ODFs, see table 2 and table A), we observe from table 3 that the signs of the corresponding direct and overall indirect effects differ.

As expected, we can see from the reported partitioned indirect coefficients that the spillovers die out across higher order neighbors (i.e., immediate neighbors, neighbors' neighbors, etc.). For the four aforementioned models where the SAR coefficient is negative, the signs of the partitioned indirect effects of a variable differ for successive orders of $\mathbf{W}_t$, which we suggest is due to the following. First, we attribute a negative partitioned indirect effect on a particular bank of interest to its spatial competition with other banks in a certain neighborhood set (or put another way, other banks pertaining to a certain order of $\mathbf{W}_t$, e.g., a bank's immediate neighbors). Second, these other banks may focus on this competitive rivalry, which may detract their attention away from competition with banks in the next spatial neighborhood set (e.g., neighbors' neighbors), leading to a positive partitioned indirect effect on the particular bank of interest for the next order of $\mathbf{W}_t$.

Finally on table 3, we can see that there are significant indirect coefficients that are non-

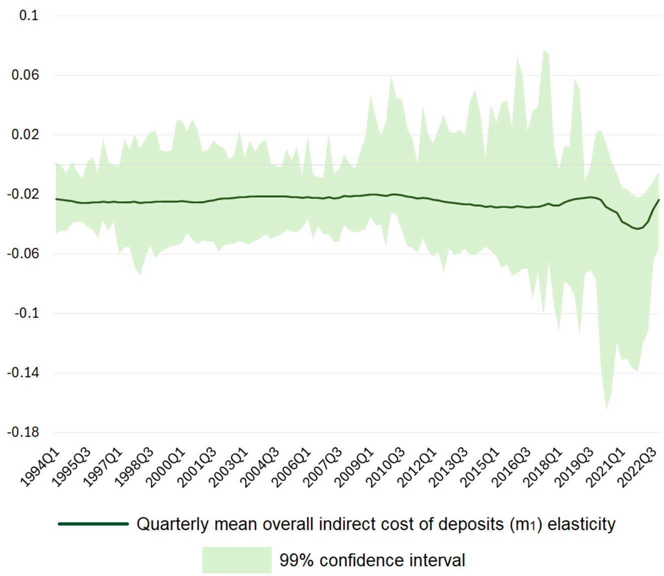

Figure 1: Quarterly mean overall indirect cost of deposits elasticity and the 99 percent confidence interval

negligible, such as the partitioned $\mathbf{W}_t^1$ and overall indirect coefficients for s_1 (deposits). Other indirect coefficients are smaller and significant, but it does not follow that the corresponding impacts will be small. This is because we analyze large commercial U.S. banks, so these coefficients will pre-multiply many large observations in the sample. Nevertheless, the indirect coefficients from all five models are typically smaller than those in spatial banking studies that use balanced panel data (Glass and Kenjegalieva, 2019; 2023; Glass *et al*, 2020a; 2020b). We suggest that a contributing factor to this is that when compared to the spatial weights matrices in these balanced panel data studies, there are a number of additional relatively small weights in the $\mathbf{W}_t$ we use here. These additional weights are relatively small as they represent the weaker links with banks that are not in the dataset for the whole study period, where including these weaker links in the sample pushes down the magnitudes of the indirect coefficients.

It is important to highlight though that the indirect coefficients in table 3 are for the sample average bank, so based on the large standard deviations (vis-à-vis the means) in table 1, there will be some notably larger indirect elasticities outside the sample mean. To illustrate, for the sample average bank, the overall indirect cost of deposits (m_1) elasticity from the cost model is on the smaller side (table 3), whereas the lower and upper ends of the 99% confidence interval in figure 1 indicates that, outside the sample mean there are cases where this indirect elasticity is non-negligible.

Before we turn to the estimates of the market power spillovers, we discuss the (in)efficiency results. In figure 2, we present the quarterly mean own efficiencies from the cost and alternative revenue models and ODF model when the left-hand side output is securities ($-q_2$). We do not present the own efficiencies from the two other ODFs because, as is the case theoretically, the results are the same as for the ODF in figure 2. We make two remarks about the efficiencies in this figure. First, we can see that over the study period the efficiencies are towards the higher end of the spectrum, which we attribute to the inclusion of time period effects. To account for the role of time in the evolution of the dependent variable, often only a time trend is included

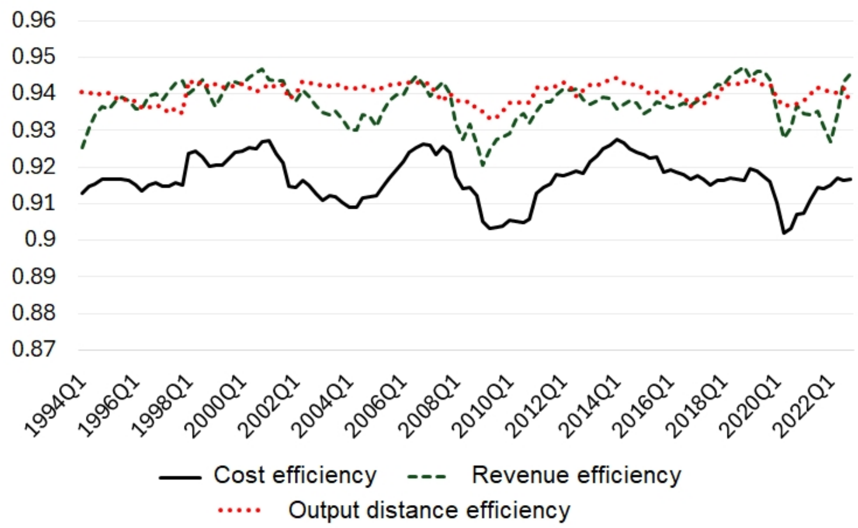

Figure 2: Quarterly mean own efficiencies

in non-spatial banking efficiency studies, previous spatial efficiency studies of banks (Glass and Kenjegalieva, 2019; Glass *et al.*, 2020b), and SFA studies more generally. Whilst we follow this literature by including a time trend, we also follow the common approach in the spatial econometrics literature of using time period effects to capture common shocks, and thus do not conflate these shocks with the effect of the SAR variable. We also suggest that by including time period effects, the common negative shocks (e.g., due to systemic factors) that banks face are not conflated with inefficiency, leading to mean own efficiencies that are towards the higher end of the spectrum.

Second, as expected though, we observe periods from the beginning of the financial crisis (2007:Q3) and during the COVID pandemic where the own cost and alternative revenue efficiencies decrease. These declines are on the smaller side (a couple of percentage points), which we suggest is due to our above first remark about figure 2: namely, as the financial crisis and the COVID pandemic were common negative shocks across the banks, the impacts of these shocks are primarily captured by the time period effects, rather than the efficiencies. However, as we analyze large U.S. banks, there are many large revenue and cost observations in the sample, so the financial implications for the banks of efficiency decreases by a couple of percentage points will not be small.

In table 4, we present quarterly mean spill-in and spill-out inefficiencies from the cost and alternative revenue models and one of the ODFs. We do not present the corresponding results from the other two distance functions because, as expected, the results are the same as for the ODF in table 4. Among the reported results are the mean overall and 1st−3rd order partitioned spill-in and spill-out indirect inefficiencies for each size (real total assets) quintile. To put the indirect inefficiencies into context, we also report mean direct inefficiencies. Note also to aid the discussion, all the results in table 4 are the relevant inefficiency $\times$ 100. Whilst the own efficiencies are the exponential of a bank's distance above or below the frontier, and not one minus inefficiency, we can see that applying the latter to the sample average direct inefficiencies would yield efficiencies that are not too different from the sample average own efficiencies (see figure 2). The mean direct inefficiencies for the quintiles indicate that the smallest banks in the sample are the most inefficient, followed by the largest banks, where both findings are reasonable.

Table 4: Mean direct and indirect inefficiencies

	Direct inefficiency	Indirect inefficiency							
		1st order		2nd order		3rd order		Overall	
		Spill-in	Spill-out	Spill-in	Spill-out	Spill-in	Spill-out	Spill-in	Spill-out
SAR stochastic cost frontier									
1st quintile	9.214	-0.043	-0.553	0.003	0.044	-0.000	-0.004	-0.040	-0.513
2nd quintile	8.389	-0.107	-0.317	0.011	0.034	-0.001	-0.004	-0.098	-0.286
3rd quintile	8.473	-0.175	-0.417	0.021	0.051	-0.003	-0.006	-0.156	-0.372
4th quintile	8.673	-0.316	-0.701	0.038	0.088	-0.005	-0.011	-0.283	-0.623
5th quintile	8.954	-2.284	-0.930	0.262	0.117	-0.032	-0.015	-2.050	-0.827
Sample	8.741	-0.584	-0.584	0.067	0.067	-0.008	-0.008	-0.524	-0.524
SAR stochastic alternative revenue frontier									
1st quintile	6.694	0.027	0.347	0.002	0.025	0.000	0.002	0.029	0.374
2nd quintile	6.320	0.068	0.205	0.006	0.019	0.001	0.002	0.075	0.226
3rd quintile	6.397	0.110	0.267	0.011	0.027	0.001	0.003	0.123	0.297
4th quintile	6.353	0.197	0.438	0.020	0.046	0.002	0.005	0.220	0.490
5th quintile	6.474	1.425	0.567	0.139	0.061	0.014	0.007	1.580	0.636
Sample	6.447	0.365	0.365	0.035	0.035	0.004	0.004	0.404	0.404
SAR stochastic output distance frontier: Left-hand side output $-q_2$									
1st quintile	6.272	-0.051	-0.682	0.006	0.091	-0.001	-0.013	-0.046	-0.603
2nd quintile	6.098	-0.126	-0.413	0.020	0.072	-0.004	-0.013	-0.109	-0.352
3rd quintile	6.068	-0.202	-0.483	0.040	0.094	-0.008	-0.019	-0.168	-0.405
4th quintile	6.142	-0.363	-0.773	0.070	0.156	-0.014	-0.032	-0.304	-0.643
5th quintile	6.157	-2.643	-1.028	0.487	0.210	-0.097	-0.045	-2.235	-0.854
Sample	6.148	-0.676	-0.676	0.125	0.125	-0.024	-0.024	-0.676	-0.676

We make four remarks about the indirect inefficiencies in table 4. First, the overall and partitioned indirect inefficiencies are symmetric for the whole sample and asymmetric for a subsample, where, as expected, each first order indirect inefficiency is the largest component of the corresponding overall measure. Second, where the SAR coefficient is negative (cost model and ODFs), the signs of the partitioned indirect inefficiencies differ for successive orders of $\mathbf{W}_t$. The reason we give for this is the same as we gave above for the same type of results for the partitioned indirect coefficients (table 3). Third, the partitioned indirect inefficiencies die out across higher order neighborhood sets, where this tends to effectively stop at the 3rd order set (i.e., when the inefficiency spillovers approach zero). Fourth, for quintiles $1-4$ (but interestingly not quintile 5), the overall spill-out inefficiencies are greater than the corresponding spill-in.

3.3 Estimates of the market power spill-ins and spill-outs

In table 5, for the sample and quintiles of the bank size distribution, we report two sets of direct and indirect (overall and partitioned) bank and product level Lerner indices. One set excludes the relevant inefficiency adjustment and the other accounts for inefficiency. While we know that the values used to obtain the direct-own Lerner indices are substantive for the banks, the types of values used to obtain the indirect Lerner indices are smaller, but non-negligible. This is evident as the mean bank level ratio of the overall indirect and direct average revenues is

5.22%. Turning now to the results in this table.

Table 5: Lerner indices

	Not adjusted for inefficiency						Adjusted for inefficiency					
	Bank size quintile						Bank size quintile					
	1st	2nd	3rd	4th	5th	Sample	1st	2nd	3rd	4th	5th	Sample
L^{Dir}						**Direct Lerner indices**						
Loans	47.8	56.4	56.4	55.8	49.4	53.2	42.4	52.5	52.4	51.8	44.5	48.7
Securities	64.4	54.9	53.7	56.3	59.3	57.7	60.9	51.1	49.6	52.5	55.4	53.9
OBS	77.5	59.9	56.2	60.9	71.1	65.1	75.6	56.4	52.1	57.1	68.5	61.9
Bank level	60.7	57.4	56.6	57.4	58.0	58.0	56.8	53.6	52.6	53.6	54.0	54.1
L_{In}^{Ind}						**Overall indirect spillover Lerner indices**						
Loans	98.9	97.8	96.8	93.9	32.9	84.1	98.9	97.8	96.8	93.9	28.0	83.1
Securities	99.2	97.7	96.5	93.6	65.4	90.5	99.2	97.7	96.5	93.5	63.0	90.0
OBS	99.5	98.0	96.8	94.5	80.1	93.8	99.5	98.0	96.8	94.4	78.8	93.5
Bank level	99.1	97.8	96.8	94.1	61.0	89.8	99.1	97.8	96.8	94.0	58.5	89.3
L_{Out}^{Ind}												
Loans	97.2	98.5	98.2	97.3	95.3	97.3	97.1	98.5	98.1	97.2	95.2	97.2
Securities	98.3	98.5	98.0	97.4	96.8	97.8	98.3	98.5	98.0	97.3	96.7	97.8
OBS	99.2	98.7	98.0	97.6	97.6	98.2	99.2	98.6	98.0	97.5	97.6	98.2
Bank level	98.0	98.6	98.1	97.4	96.4	97.7	98.0	98.6	98.1	97.3	96.4	97.7
L_{In}^{Ind}						**1st order indirect spillover Lerner indices**						
Loans	98.7	97.1	95.6	91.9	33.0*	78.5	98.7	97.1	95.6	91.8	29.6*	77.5
Securities	99.0	97.0	95.2	91.5	53.0	87.2	99.0	97.0	95.2	91.4	50.0	86.6
OBS	99.4	97.3	95.6	92.7	73.9	91.8	99.4	97.3	95.5	92.7	72.3	91.4
Bank level	98.9	97.2	95.6	92.1	47.4	86.3	98.9	97.2	95.6	92.0	44.5	85.7
L_{Out}^{Ind}												
Loans	81.8	90.2	87.4	81.0	67.0	81.5	81.5	90.0	87.2	80.6	66.4	81.1
Securities	88.9	89.8	86.3	81.9	77.8	85.0	88.9	89.6	86.0	81.4	77.3	84.7
OBS	94.2	90.8	86.2	82.9	83.4	87.5	94.1	90.6	85.9	82.4	82.9	87.2
Bank level	87.2	90.5	87.2	81.8	75.1	84.4	87.1	90.3	86.9	81.4	74.6	84.1
L_{In}^{Ind}						**2nd order indirect spillover Lerner indices**						
Loans	99.8	99.4	98.8	97.8	75.5	94.3	99.8	99.4	98.8	97.8	75.5	94.3
Securities	99.8	99.3	98.6	97.8	87.5	96.6	99.8	99.3	98.6	97.8	87.5	96.6
OBS	99.9	99.3	98.7	98.2	93.7	98.0	99.9	99.3	98.7	98.2	93.8	98.0
Bank level	99.8	99.4	98.8	97.9	86.3	96.4	99.8	99.4	98.8	97.9	86.3	96.5
L_{Out}^{Ind}												
Loans	95.9	97.5	96.5	94.5	90.1	94.9	95.9	97.5	96.5	94.4	90.1	94.9
Securities	97.5	97.3	96.1	94.8	93.8	95.9	97.5	97.3	96.1	94.8	93.8	95.9
OBS	98.3	97.5	95.9	95.0	95.2	96.4	98.3	97.5	95.9	95.0	95.2	96.4
Bank level	97.3	97.6	96.4	94.7	92.7	95.7	97.3	97.6	96.3	94.7	92.7	95.7

Notes: OBS denotes off-balance sheet items. * denotes a mean based on 95% of the banks in the 5th quintile due to outliers.

We can see that when the direct Lerner indices include the inefficiency adjustment the reported indices are lower than when inefficiency is overlooked. All the reported direct-own Lerner indices are also some way above 0 and below 1 and thus not out of line with the non-spatial Lerner indices for large U.S. banks in the literature. With regard to the magnitudes of the direct Lerner indices for the quintiles, we can see that for loans these indices are the lowest for quintiles 1 and 5, where for the latter this could be because the larger banks may not focus on specialist bespoke loans with high profit margins, but on standard types of loans to their large customer bases. For the former this could be because these banks have smaller loan market shares and may therefore place a greater emphasis on other activities, which is in line with quintile 1 having the highest direct Lerner indices for securities and off-balance sheet (OBS) items. The next highest direct Lerner indices for securities and OBS items are for quintile 5, where the greater resources of quintile 5 banks is likely to be a contributing factor to these results.

Of the indirect Lerner indices in table 5, the most notable evidence of a lower index when inefficiency is accounted for is the overall and 1st order spill-in indices for quintile 5. In addition, we note that the partitioned indirect spill-in Lerner indices have an intuitive pattern. As the order of the neighborhood set increases ($\mathbf{W}_t$ to $\mathbf{W}_t^2$), these partitioned Lerner indices, and particularly those for quintile 5, increase, where there are a number of other cases where these indices are approaching 1 (i.e., quintiles $1-4$). To aid the interpretation of these indices recall that as the order of $\mathbf{W}_t$ increases, the magnitudes of the partitioned indirect spill-in coefficients die out. This means that the partitioned marginal cost spill-ins will also die out, giving a partitioned indirect spill-in Lerner index approaching 1, which points to a high market power spill-in. The interpretation of the overall indirect spill-in Lerner index is the same, as it is the collective representation of the corresponding partitioned indices.

For the partitioned indirect spill-in Lerner indices, the picture for quintile 5 differs from what we observe for the other quintiles. For quintile 5 these indices are smaller than for quintiles $1-4$. This points to relatively higher marginal cost spill-ins for quintile 5 banks and thus lower market power spill-ins. We find that this contrasts though with results we present below for a number of individual quintile 5 banks (see table 6). Given the spill-outs from a bank lead to spill-ins for other banks, it is no surprise to find that the partitioned indirect spill-out Lerner indices exhibit the same type of pattern as the above corresponding spill-in indices. The only difference is that the partitioned indirect spill-out Lerner indices for quintile 5 are larger than the corresponding spill-in index. This indicates that the quintile 5 banks have relatively lower marginal cost spill-outs, which leads to higher market power spill-outs. To sum up, to different degrees, we can see from the overall indirect Lerner indices for the sample that the market power spillovers tend to be high, which underlines the importance of the interconnectedness in the banking industry.

Next, as there is marked variation in the number of banks over the sample period, we consider a core subset of banks that are in the data sample for at least 95% of the time periods. From this pool, in table 6 we present the top and bottom 5 inefficiency adjusted Lerner indices for banks in the 5th and 1st−4th quintiles of the bank size distribution.[5] Looking at the top and bottom 5 direct Lerner indices indicates that there is a wide difference in the market power of some of the banks. As we would expect, there are global systemically important banks (Bank of America and JPMorgan Chase) with a top 5 direct Lerner index. There are, however, banks with a top 5 direct Lerner index that are much smaller − the smallest being the First Financial Bank in quintile 2. Interestingly, the collapsed Silicon Valley Bank (March 2023) has a top 5 direct Lerner index. This finding is likely because it specialized in providing products to venture capital-backed technology startups, which would involve the bank having a higher markup to reflect the greater risks associated with this type of business.

We make three remarks about the indirect Lerner indices in table 6. First, there are a number of banks with top 5 overall indirect Lerner indices that are in the mid-to-high 90s,

[5] We present bank level results in table 6 due to space constraints. The corresponding product level results are available on request.

Table 6: Lerner indices for selected banks

	5th quintile banks			1st-4th quintile banks	
		Direct Lerner index (inefficiency adjusted)			
Top 5	Bank of America	64.1	*Top 5*	Westamerica	69.3
	First Hawaiian Bank	62.5		Farmers and Merchants	65.0
	U.S. Bank National	62.3		First Financial Bank	64.6
	JPMorgan Chase	61.0		FirstBank of Colorado	64.1
	Silicon Valley	60.2		NBT Bank	62.3
Bottom 5	State Street Bank	52.0	*Bottom 5*	Equity Bank	42.8
	Union Bank of California	49.6		Central Bank	42.5
	First-Citizens Bank	49.0		PlainsCapital	42.5
	BMO Bank	48.1		The First	41.4
	BancorpSouth	46.1		RCB Bank	40.8
		Overall indirect spill-in Lerner index (inefficiency adjusted)			
Top 5	Bank of America	96.5	*Top 5*	Enterprise Bank	97.7
	JPMorgan Chase	96.2		TIB	97.4
	U.S. Bank National	95.7		Pinnacle Bank	97.2
	Frost Bank	95.3		Washington Trust	97.2
	First Hawaiian Bank	94.3		NBT Bank	97.2
Bottom 5	State Street Bank	87.3	*Bottom 5*	CNB Bank	73.9
	First-Citizens Bank	86.7		Bremer Bank	73.8
	Commerce Bank	84.9		Fremont Bank	72.1
	Valley National Bank	81.9		First State Community Bank	71.4
	Hibernia Bank	79.6		Merchants Bank of Indiana	71.4
		Overall indirect spill-out Lerner index (inefficiency adjusted)			
Top 5	Zions	99.5	*Top 5*	First Mid	99.5
	Comerica	99.5		Bank of Stockton	99.5
	State Street Bank	99.4		First Interstate Bank	99.5
	BB&T	99.3		Tompkins	99.4
	PNC Bank	99.2		First National Bank Texas	99.4
Bottom 5	Bank of Hawaii	97.5	*Bottom 5*	Amarillo National Bank	96.2
	BMO Bank	97.3		The First	95.8
	Northern Trust	97.1		Stockman Bank	95.7
	Valley National	97.1		Central Bank	95.3
	BancorpSouth	96.5		RCB Bank	95.2

Note: Each bank in this table is in the data sample for a least 95% of the study period

e.g., Bank of America and JPMorgan Chase. This indicates that these banks have high market power spill-ins and spill-outs, which is consistent with concerns about the market power of the very large U.S. banks. Second, the reported bottom 5 overall spill-in Lerner indices range from $71.4 - 87.3$, which, to different degrees, indicates that these banks have lower market power spill-ins. Third, the bottom 5 overall spill-in indices for the quintile 5 banks are above the corresponding mean in table 5 (61.0). This indicates that, on average, the quintile 5 banks that do not survive for 95% of the study period have lower market power spill-ins. This highlights the value of unbalanced panel data for our empirical case as the lower market power spill-ins for these banks may have contributed to some of them dropping out of the sample.

The final set of bank level results we present is the geographical distribution of the overall indirect (spill-in and spill-out) Lerner indices. For 2022:Q4, figure 3 overlays the bivariate heat map of these two indices onto the banks' branch locations. In this map, the top, middle and bottom thirds of the distributions are used to group the bank level pairs of these indices.[6] We can see that only 1.9% of pairs comprise values that are in the middle thirds of the distributions. It is also evident that banks with pairs of values in the top thirds (13.7%) tend to have branches in

[6] The three corresponding product level heat maps are similar to figure 3 and so are not presented for brevity.

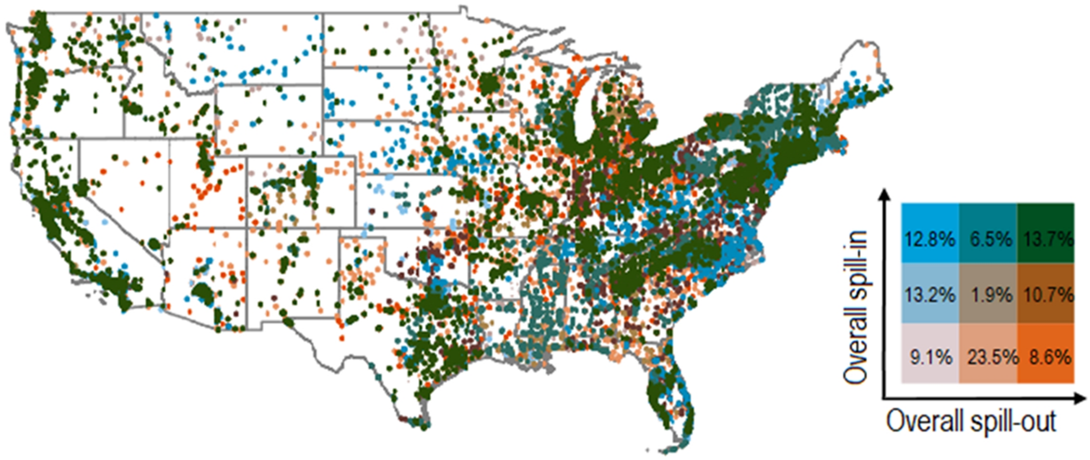

Figure 3: Bank-branch geographical distribution of the overall indirect Lerner indices for 2022:Q4

densely populated metropolitan statistical areas (MSAs) (e.g., New York; Boston; Minneapolis-St. Paul; Chicago; Detroit; Atlanta; Houston; Dallas-Forth Worth; Los Angeles; San Diego; San Francisco; Seattle; Portland). This is in line with our expectations as the larger agglomeration effects in densely populated areas is conducive to larger market power spillovers.

4 Summary and policy relevance

It is well-known that U.S. banks are interconnected. There are a number of reasons for this, a key one being rival banks with branches in the same geographical areas. This points to spatial dependence between neighboring banks, which represents the net effect of the negative bank correlation due to their competitive rivalry and positive correlation because banks face common economic phenomena. Key examples of these phenomena are industrywide regulation, the FOMC's setting of the federal funds rate, market conditions, and headline changes in economies at the city, state, regional and national levels. Studies have shown that significant SAR dependence between U.S. banks leads to spillovers of total factor productivity (TFP) growth (Glass and Kenjegalieva, 2019; Glass *et al.* 2020b) and returns to scale spillovers (Glass *et al.*, 2020a; Glass and Kenjegalieva, 2023). It is reasonable to therefore consider market power spillovers between U.S. banks and their corresponding products. This is an important issue because bank market power affects, among other things, the price and availability of credit, which has a wider effect on the general business environment.

There are large bodies of non-spatial banking studies on productivity and efficiency, scale economies and market power, where the methods to analyze all but the latter have been extended to spillovers at the bank level. Accordingly, at the micro levels of banks and their products, we introduce a method to obtain asymmetric bidirectional spillover Lerner indices (with and without adjustment for inefficiency spill-ins and spill-outs). Rather than follow the extant banking spatial SFA studies, which all use relatively small balanced panel datasets, we estimate these indices for large commercial U.S. banks using a large unbalanced panel dataset. Whilst this means the estimation is more computationally demanding, unbalanced panel data is a better empirical representation as it reflects the marked consolidation in the industry.

Bank interconnectedness is very important as it contributes to a number of risks, e.g., bank run contagion. Accordingly, U.S. bank regulatory authorities extensively monitor different forms of bank interconnectedness. The following key empirical findings use the market power spillovers to provide new information about bank interconnectedness. First, consistent with consolidation in the industry leading to concerns about the market power of the largest U.S. banks, we find that a number of banks have high indirect spillover Lerner indices, e.g., two global systemically important banks (Bank of America and JPMorgan Chase). This finding suggests that overlooking bank market power spillovers may result in U.S. competition authorities understating the market power impact of a large bank merger. The implication being that overlooking these spillovers may lead to unexpectedly larger increases in the price of credit and, as a result, unexpectedly bigger negative impacts on the general business environment and household welfare. From a policy perspective, we therefore suggest that U.S. competition authorities should account for such spillovers when assessing future large bank mergers.

Second, the mean bank level overall spill-in Lerner index for quintile 5 of the size distribution is well below the bottom 5 corresponding indices from the pool of quintile 5 banks that are in the sample for the vast majority of the study period. This underlines the benefit of unbalanced panel data for our empirical case, as the lower market power spill-ins for the banks outside this pool are intuitive as they may have contributed to some of these banks dropping out of the sample. Third and finally, we find that the banks with bidirectional overall indirect Lerner indices in the top thirds of the distributions tend to have branches in densely populated cities. This is also intuitive as the larger agglomeration effects in major cities will facilitate higher market power spillovers.

References

Algeri, C., L. Anselin, A. F. Forgione & C. Migliardo (2022): 'Spatial dependence in the technical efficiency of local banks'. *Papers in Regional Science*, vol. 101, pp. 685-716.

Adetutu, M., A. J. Glass, K. Kenjegalieva and R. C. Sickles (2015): 'The effects of efficiency and TFP growth on pollution in Europe: A multistage spatial analysis'. *Journal of Productivity Analysis*, vol. 43, pp. 307-326.

Ayouba, K. (2023): 'Spatial dependence in production frontier models'. *Journal of Productivity Analysis*, vol. 60, pp. 21-36.

Baltagi, B. H., G. Bresson and J. M. Etienne (2015): 'Hedonic housing prices in Paris: An unbalanced spatial lag pseudo-panel model with nested random effects'. *Journal of Applied Econometrics*, vol. 30, pp. 509-528.

Baltagi, B. H., P. Egger and M. Pfaffermayr (2007): 'Estimating models of complex FDI: Are there third-country effects?'. *Journal of Econometrics*, vol. 140, pp. 260-281.

Berger, A. N., D. B. Humphrey and L. B. Pulley (1996): 'Do consumers pay for one-stop banking? Evidence from an alternative revenue function'. *Journal of Banking and Finance*, vol. 20, pp. 1601–1621.

Berger, A. N. and L. J. Mester (2003): 'Explaining the dramatic changes in performance of US banks: Technological change, deregulation, and dynamic changes in competition'. *Journal of Financial Intermediation*, vol. 12,pp. 57–95.

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.

Bikker, J. A., S. Shaffer and L. Spierdijk (2012): 'Assessing competition with the Panzar-Rosse model: The role of scale, costs, and equilibrium'. *Review of Economics and Statistics*, vol. 94, pp. 1025-1044.

Boyd, J. H. and M. Gertler (1994): 'Are banks dead? Or are the reports greatly exaggerated?' *Federal Reserve Bank of Minneapolis Quarterly Review*, vol. 18, pp. 2-23.

Chen, Y. Y., P. Schmidt and H-J. Wang (2014): 'Consistent estimation of the fixed effects stochastic frontier model'. *Journal of Econometrics*, vol. 181, pp. 65-76.

Coccorese, P. (2009): 'Market power in local banking monopolies'. *Journal of Banking and Finance*, vol. 33, pp. 1196-1210.

Coccorese, P., C. Girardone and S. Shaffer (2021): 'What affects bank market power in the euro area? A country-level structural model approach'. *Journal of International Money and Finance*, vol. 117, p. 102443.

Corrado, L. and B. Fingleton (2012): 'Where is the economics in spatial econometrics?' *Journal of Regional Science*, vol. 52, pp. 210-239.

Danisman, G. O. and P. Demirel (2019): 'Bank risk-taking in developed countries: The influence of market power and bank regulations'. *Journal of International Financial Markets, Institutions and Money*, vol. 59, pp. 202-217.

Degryse, H., P. Morales-Acevedo and S. Ongena (2019): *Competition in the banking sector.* In *The Oxford Handbook of Banking,* Berger, A. N., P. Molyneux and J. O. S. Wilson (Eds). Oxford: Oxford University Press.

Delis, M. D. and E. G. Tsionas (2009): 'The joint estimation of bank-level market power and efficiency'. *Journal of Banking and Finance*, vol. 33, pp.1842-1850.

Dick, A. A. (2006): 'Nationwide branching and its impact on market structure, quality, and bank performance'. *Journal of Business*, vol. 79, pp. 567-592.

Dix, R. and R. Orzach (2023): 'Market power spillovers across airline routes'. MIT working paper. Available at https://economics.mit.edu/sites/default/files/inline-files/airlines_0.pdf

Druska, V. and W. C. Horrace (2004): 'Generalized moments estimation for spatial panel data: Indonesian rice farming'. *American Journal of Agricultural Economics*, vol. 86, pp.185-198.

Efthyvoulou, G. and C. Yildirim (2014): 'Market power in CEE banking sectors and the impact of the global financial crisis'. *Journal of Banking and Finance*, vol. 40, pp. 11-27.

Egger, P., M. Pfaffermayr and H. Winner (2005): 'An unbalanced spatial panel data approach to US state tax competition'. *Economics Letters*, vol. 88, pp. 329-335.

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.

Fusco, E., G. Arbia, F. Vidoli and V. Nardelli (2024): 'On spatio-temporal stochastic frontier models'. Available at https://www.arxiv.org/pdf/2410.20915

Galli, F. (2024): 'Accounting for unobserved individual heterogeneity in spatial stochastic frontier models: The case of Italian innovative start-ups'. *Spatial Economic Analysis*, pp. 1-26.

Galli, F. (2023a): 'A spatial stochastic frontier model introducing inefficiency spillovers'. *Journal of the Royal Statistical Society Series C: Applied Statistics*, vol. 72, pp. 346-367.

Galli, F. (2023): 'A spatial stochastic frontier model including both frontier and error-based spatial cross-sectional dependence'. *Spatial Economic Analysis*, vol. 18, pp. 239-258.

Glass, A. J. and K. Kenjegalieva (2024): 'The role of inefficiency in a productivity puzzle: Regional evidence for Great Britain'. *Journal of Regional Science*, vol. 64, pp. 1287-1322

Glass, A. J. and K. Kenjegalieva (2023): 'Dynamic returns to scale and geography in US banking'. *Papers in Regional Science*, vol. 102, pp. 53-86.

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

Glass, A. J., A. Kenjegaliev and K. Kenjegalieva (2020a): 'Spatial scale and product mix economies in US banking with simultaneous spillover regimes'. *European Journal of Operational Research*, vol. 284, pp. 693-711.

Glass, A. J., K. Kenjegalieva and M. Douch (2020b): 'Uncovering spatial productivity centers using asymmetric bidirectional spillovers'. *European Journal of Operational Research*, vol. 285, pp. 767-788.

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

Greene, W. H. (2005): 'Reconsidering heterogeneity in panel data estimators of the stochastic frontier model'. *Journal of Econometrics*, vol. 126, pp. 269-303.

Gude, A., I. Álvarez and L. Orea (2018): 'Heterogeneous spillovers among Spanish provinces: a generalized spatial stochastic frontier model'. *Journal of Productivity Analysis*, vol. 50, pp. 155-173.

Halleck Vega, S. and J. P. Elhorst (2015): 'The SLX model'. *Journal of Regional Science*, vol. 55, pp. 339-363.

Hou, Z., S. Zhao and S. C. Kumbhakar (2023): ‘The GMM estimation of semiparametric spatial stochastic frontier models’. *European Journal of Operational Research*, vol. 305, pp. 1450-1464.

Jin, F. and L-F. Lee (2020): ‘Asymptotic properties of a spatial autoregressive stochastic frontier model’. *Journal of Spatial Econometrics*, vol. 1.

Jondrow, J., C. A. K. Lovell, I. S. Materov and P. Schmidt (1982): ‘On the estimation of technical inefficiency in the stochastic frontier production function model’. *Journal of Econometrics*, vol. 19, pp. 233-238.

Kutlu, L. (2018): ‘Estimating efficiency in a spatial autoregressive stochastic frontier model’. *Economics Letters*, vol. 163, pp. 155-157.

Kutlu, L. and U. Nair-Reichert (2019): ‘Agglomeration effects and spatial spillovers in efficiency analysis: A distribution-free methodology’. *Regional Studies*, vol. 53, pp. 1565-1574.

Kutlu, L., K. C. Tran and M. G. Tsionas (2020): ‘A spatial stochastic frontier model with endogenous frontier and environmental variables’. *European Journal of Operational Research*, vol. 286, pp. 389-399.

Lai, H. P. and K. C. Tran (2022): ‘Persistent and transient inefficiency in a spatial autoregressive panel stochastic frontier model’. *Journal of Productivity Analysis*, vol. 58, pp. 1-13.

Maudos, J. and J. F. De Guevara (2007): ‘The cost of market power in banking: Social welfare loss vs. cost inefficiency.’ *Journal of Banking and Finance*, vol. 31, pp. 2103-2125.

Mi, B., L. Zhang, L. Han and Y. Shen (2024): ‘Bank market power and financial reporting quality’. *Journal of Corporate Finance*, vol. 84, p. 102530.

Neyman, J. and E. L. Scott (1948): ‘Consistent estimates based on partially consistent observations. *Econometrica*, vol. 16, pp.1-32.

Orea, L. and I. C. Álvarez (2019): ‘A new stochastic frontier model with cross sectional effects in both noise and inefficiency terms’. *Journal of Econometrics*, vol. 213, pp. 556-577.

Ramajo, J., J. M. Cordero and M. Á. Márquez (2017): ‘European regional efficiency and geographical externalities: A spatial nonparametric frontier analysis. *Journal of Geographical Systems*, vol. 19, pp. 319-348.

Ramajo, J. and G. J. D. Hewings (2018): ‘Modelling regional productivity performance across Western Europe’. *Regional Studies*, vol. 52, pp. 1372-1387.

Sealey Jr, C. W. 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.

Semih Yildirim, H. and G. C. Philippatos (2007): ‘Competition and contestability in Central and Eastern European banking markets’. *Managerial Finance*, vol. 33, pp. 195-209.

Shaffer, S. and L. Spierdijk (2020): ‘Measuring multi-product banks’ market power using the Lerner index’. *Journal of Banking and Finance*, vol. 117, p. 105859.

Shaffer, S. and L. Spierdijk (2015): ‘The Panzar–Rosse revenue test and market power in banking’. *Journal of Banking and Finance*, vol. 61, pp. 340-347.

Skevas, I. and T. Skevas (2021): ‘A generalized true random-effects model with spatially autocorrelated persistent and transient inefficiency’. *European Journal of Operational Research*, vol. 293, pp. 1131-1142.

Tran, K. C. and M. G. Tsionas (2023): ‘Semiparametric estimation of a spatial autoregressive nonparametric stochastic frontier model’. *Journal of Spatial Econometrics*, vol. 4.

Tran, K. C., M. G. Tsionas and A. B. Prokhorov (2023): ‘Semiparametric estimation of spatial autoregressive smooth-coefficient panel stochastic frontier models’. *European Journal of Operational Research*, vol. 304, pp. 1189-1199.

Wang, H-J. and C-W. Ho (2010): ‘Estimating fixed-effect panel stochastic frontier models by model transformation’. *Journal of Econometrics*, vol. 157, pp. 286-296.

Wang, X., L. Han and X. Huang (2020): ‘Bank market power and SME finance: Firm-bank evidence from European countries’. *Journal of International Financial Markets, Institutions and Money*, vol. 64, p. 101162.

Wang, Y., T. M. Whited, Y. Wu and K. Xiao (2022): ‘Bank market power and monetary policy transmission: Evidence from a structural estimation’. *Journal of Finance*, vol. 77, pp. 2093-2141.

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.

Appendix 1

Candidate bank market power measures for extension to the spatial setting

We review the three main approaches to measure non-spatial bank market power. As all three measures are model-based they can be extended to the spatial setting. This is because for our spatial extension we need to use a model-based approach to estimate unobserved spillovers. For brevity, we list a small selection of studies for illustrative purposes that have applied the three approaches.

The first approach is the Panzar and Rosse H statistic (e.g., Coccorese, 2009; Bikker *et al.*, 2012; Mi *et al.*, 2024). H_i is calculated using the alternative revenue function, $TR_i = f(Q_i, M_i)$, where TR_i is the total revenue of the ith bank, M_i is the vector of input prices indexed $g \in 1, ..., G$ and Q_i is as defined for Eq. 1. $H_i = \sum_g \frac{\partial \ln TR_i}{\partial \ln M_{ig}}$ represents the ratio of the change in the output price to the change in input prices as outputs are held constant. The lower limit of H is negative and the upper limit is 1. The higher H is for a bank, the less market power it has in terms of its ability to set price independently of cost. It would be econometrically feasible to extend the H statistic to the spatial setting, which would first involve estimating an appropriate spatial alternative revenue function. From thereon the approach would be similar to the one we propose here for our different line of inquiry. We follow a different course because, first, we want to calculate market power spillovers with and without an adjustment for inefficiency. It not possible though to incorporate an adjustment for inefficiency into the above approach to calculate the H statistic. This is because the revenue elasticities with respect to the input prices are not affected by the presence of inefficiency. Second, Shaffer and Spierdijk (2015) show that the H statistic cannot reliably measure market power. They reach this conclusion based on counterintuitive results of $H > 0$ for five standard highly non-competitive oligopoly settings.

The second approach uses the *Conduct* parameter at the level of bank loans (e.g., Coccorese, 2009; Delis and Tsionas, 2009) and across banks at the country level (e.g., Coccorese *et al.*, 2021). Coccorese (2009) reports mean *Conduct* parameters across the sampled banks for the full study period and subperiods, while Delis and Tsionas (2009) estimate time-varying *Conduct* parameters for each bank. The conduct measure is the conjectured response of the industry's kth output to variation in the same output at the ith bank (Degryse *et al.*, 2019). This is a spillover to a more aggregate level and, therefore, differs from the usual interpretation of spillovers between units at the same level of aggregation we consider here, i.e., spillovers between banks and between the corresponding products of banks.

To tie in with the brief below discussion of the Lerner index in the third approach, we recognize that the conduct measure can be expressed as $Conduct_{ik} = e_{ik} L_{ik}$ (e.g., Degryse *et al.*, 2019). e_{ik} is the price elasticity of demand for the bank's kth product and $L_{ik} = \frac{P_{ik} - MC_{ik}}{P_{ik}}$ is the non-spatial Lerner index defined in Eq. 1. $Conduct_{ik}$ is therefore an elasticity-adjusted Lerner index and so the higher the $Conduct_{ik}$ parameter for a bank, the more market power it has. MC_{ik} is at the bank product level and can be estimated from a multi-product bank level cost model (e.g., Shaffer and Spierdijk, 2020). In line with the discussion of L_{ik} in Eq. 1, for $P_{ik} \geq MC_{ik}$, the value of $Conduct_{ik}$ will be in the range 0 to 1, and for $P_{ik} < MC_{ik}$, $Conduct_{ik} < 0$. Whilst a case can equally be made to extend $Conduct_{ik}$ to the spatial setting, we extend the well-established Lerner index. A motivation for this is the practical usage of the Lerner index. This is because of the three main approaches to measure non-spatial bank market power we review, the Lerner index is the only (country-level) bank market power measure which the World Bank reports in its Global Financial Development Database.

We refer to L_{ik} (e.g., Shaffer and Spierdijk, 2020; Wang *et al.*, 2020; Mi *et al.*, 2024) and $Markup_{ik} = \frac{P_{ik} - MC_{ik}}{MC_{ik}}$ as the third approach as the latter is interpreted in the same way as L_{ik} (see the above discussion of Eq. 1). There are some differences though between these two measures. To illustrate, as a cost function is monotonically increasing in each output $MC_{ik} > 0$, so $L_{ik} < 1$, but $Markup_{ik} < 1$ may or may not be the case. Whilst we consider only the Lerner index due to space constraints and its appeal and widespread use, the spatial approach we introduce can also be applied to obtain markup spill-ins and spill-outs.

Appendix 2

Estimation procedure for the SAR stochastic frontier models

We estimate Eq. 5 using within maximum likelihood (ML) estimation with the closed skew normal distribution. This involves adapting the non-spatial estimation procedure in Chen *et al.* (2014) to the case of SAR dependence. The procedure begins with the following within transformation of Eq. 5.

$$\widetilde{c}_{it} = \widetilde{x}_{it}\beta' + \delta \sum_{j=1}^{N_t} w_{ijt}\widetilde{c}_{jt} + \widetilde{\varepsilon}_{it}, \tag{A1}$$

where the within transformed observations and composed error ($\widetilde{c}_{it}$, $\widetilde{x}_{it}$ and $\widetilde{\varepsilon}_{it}$) are deviations from their respective means ($\widetilde{c}_{it} = c_{it} - \frac{1}{T_i}\sum_t c_{it}$ and similarly for $\widetilde{x}_{it}$ and $\widetilde{\varepsilon}_{it}$). This transformation removes d_i and α, where we do not then demean by time period to eliminate b_t as this would also eliminate the important t and t^2 components of the non-linear time trend. To simplify the notation in Eq. A1 we subsume b_t into $\widetilde{x}_{it}$. As noted previously, the closed skew normally distributed error term ε_{it} has two components, $v_{it} + u_{it}$. As is standard, let $\lambda = \frac{\sigma_u}{\sigma_v}$ and $\sigma_{uv}^2 = \sigma_u^2 + \sigma_v^2$. The probability density of the composed error is then

$$f(\widetilde{\varepsilon}) = \frac{2}{\sigma}\varphi\left(\frac{\widetilde{\varepsilon}}{\sigma}\right)\Phi\left(-\frac{\lambda\widetilde{\varepsilon}}{\sigma}\right), \tag{A2}$$

where φ and Φ are the standard normal probability and cumulative density functions (pdf and cdf), respectively.

Denote two vectors of within transformed composed errors for the *ith* bank as $\widetilde{\varepsilon}_i = (\widetilde{\varepsilon}_{i1}, ..., \widetilde{\varepsilon}_{iT_i})'$ and $\widetilde{\varepsilon}_i^* = (\widetilde{\varepsilon}_{i1}, ..., \widetilde{\varepsilon}_{iT_i-1})'$, where the number of periods which the *ith* bank is in the sample for is T_i. Using Theorem 3 in Chen *et al.* (2014), the log-likelihood function for the within transformed model is as follows.

$$\ln L_{within} = \eta + \sum_i \left[\ln \varphi_{T_i-1}\left(\widetilde{c}_{it}^* - \widetilde{x}_{it}^*\beta' - \delta\sum_{j=1}^{N_t} w_{ijt}\widetilde{c}_{jt}^*;\ 0, \sigma^2\left(\mathbf{I}_{T_i-1} - \frac{1}{T_i}\Omega_{T_i-1}\right)\right)\right] +$$
$$\sum_i \left[\ln \Phi_{T_i}\left(-\frac{\lambda}{\sigma}(\widetilde{c}_{it} - \widetilde{x}_{it}\beta' - \delta\sum_{j=1}^{N_t} w_{ijt}\widetilde{c}_{jt});\ 0_{T_i}, \mathbf{I}_{T_i} + \frac{\lambda^2}{T_i}\Omega_{T_i}\right)\right] + \ln|\mathbf{I}^* - \delta\mathbf{W}^*|, \tag{A3}$$

where to fix ideas the first summation sums the $T_i - 1$ residuals across all the banks. η is the constant, $\Omega_{T_i-1} = \iota_{T_i-1}\iota'_{T_i-1}$ is the $T_i - 1\times$ $T_i - 1$ matrix of ones, and $\mathbf{I}_{T_i-1}$ is the corresponding identity matrix. To account for the endogeneity of the SAR variable and $\widetilde{\varepsilon}$ being unobserved, Eq. A3 includes $\ln|\mathbf{I}^* - \delta\mathbf{W}^*|$, which is the logged determinant of the Jacobian of the transformation from $\widetilde{\varepsilon}$ to $\widetilde{c}$ (e.g., Elhorst, 2009). $\mathbf{I}^*$ is the block diagonal identity matrix with T blocks comprising the N_t dimensional $\mathbf{I}_{N_t}$ for all $t \in 1, ..., T$. Likewise, $\mathbf{W}^*$ is the block diagonal spatial weights matrix, where the T blocks are the N_t dimensional $\mathbf{W}_t$ for all $t \in 1, ..., T$. This within log-likelihood function is maximized with respect to β, δ, λ and σ^2.

To evaluate the within log-likelihood function we use the following approach. We estimate the cdf in Eq. A3 by following appendix C in Chen *et al.* (2014), which involves evaluating a single integral. As suggested by Pace and Barry (1997), we also pre-calculate $\ln|\mathbf{I}^* - \delta\mathbf{W}^*|$ for a vector of values of δ based on 0.001 increments over the interval $\left(\frac{1}{\min(h_1^{\min},...,h_T^{\min})}, \frac{1}{\max(h_1^{\max},...,h_T^{\max})}\right)$. The estimate of the asymptotic variance is obtained by taking the inverse of the information matrix of the ML estimate of Eq. A3. To estimate u_{it}, we use the standard approach in the literature in Eq. A4 to predict u_{it} conditional on ε_{it} (Jondrow *et al.*, 1982).

$$\widehat{u}_{it} = \mathrm{E}\left(u_{it}|\widetilde{\varepsilon}_{it}\right) = \frac{\sigma_u \sigma_v}{\sigma_{uv}} \left(\frac{\varphi_{it}}{1-\Phi_{it}} - \frac{\widetilde{\varepsilon}_{it}\lambda}{\sigma_{uv}}\right), \quad \text{(A4)}$$

where $\Phi_{it} = \Phi\left(\widetilde{\varepsilon}_{it}\lambda/\sigma_{uv}\right)$ and $\varphi_{it} = \varphi\left(\widetilde{\varepsilon}_{it}\lambda/\sigma_{uv}\right)$.

Appendix 3

Table A: Further SAR stochastic frontier models

SAR stochastic output distance frontier: Left-hand side output $-q_1$				**SAR stochastic output distance frontier: Left-hand side output** $-q_3$			
	Model coeff		*Model coeff*		*Model coeff*		*Model coeff*
s_1	-0.850^{***}	t	-0.002^{***}	s_1	-0.860^{***}	t	-0.002^{***}
s_2	-0.183^{***}	t^2	$0.001 \times 10^{-2***}$	s_2	-0.209^{***}	t^2	$0.003 \times 10^{-2***}$
s_3	-0.003	s_1t	-0.001^{***}	s_3	0.002	s_1t	-0.002^{***}
q_2	0.170^{***}	s_2t	-0.002^{***}	q_1	0.774^{***}	s_2t	-0.002^{***}
q_3	0.079^{***}	s_3t	$0.015 \times 10^{-2**}$	q_2	0.153^{***}	s_3t	$-0.018 \times 10^{-2*}$
s_1^2	-0.052^{***}	q_2t	-0.004×10^{-3}	s_1^2	-0.050^{***}	q_1t	-0.001^{***}
s_2^2	-0.031^{***}	q_3t	0.001^{***}	s_2^2	-0.039^{***}	q_2t	0.008×10^{-2}
s_3^2	-0.002^{***}	*LLA*	0.664^{***}	s_3^2	-0.002^{**}	*LLA*	0.564^{***}
s_1s_2	0.104^{***}	*Tier1CR*	-1.353^{***}	s_1s_2	0.100^{***}	*Tier1CR*	-1.934^{***}
s_1s_3	-0.009^{***}	*Tier2CR*	0.233	s_1s_3	-0.003	*Tier2CR*	1.369^{***}
s_2s_3	-0.001	*Equity*	0.268^{***}	s_2s_3	0.028×10^{-2}	*Equity*	0.720^{***}
q_2^2	0.014^{***}	*NPL*	0.883^{***}	q_1^2	0.027^{***}	*NPL*	0.670^{***}
q_3^2	0.009^{***}	*HHI*	-0.171^{***}	q_2^2	0.012^{***}	*HHI*	-0.196^{***}
q_2q_3	0.004^{***}	*Age*	-0.008	q_1q_2	-0.031^{***}	*Age*	-0.022^{**}
s_1q_2	-0.017^{***}	*Security*	0.073^{**}	s_1q_1	0.008^{***}	*Security*	0.138^{***}
s_1q_3	0.014^{***}	$W_t(-q_1)$	-0.231^{***}	s_1q_2	-0.024^{***}	$W_t(-q_3)$	-0.296^{***}
s_2q_2	0.007^{***}	σ	0.056^{***}	s_2q_1	0.010^{***}	σ	0.120^{***}
s_2q_3	-0.014^{***}	λ	0.663^{***}	s_2q_2	0.005^{**}	λ	0.675^{***}
s_3q_2	-0.006^{***}			s_3q_1	0.016^{***}		
s_3q_3	-0.006^{***}	*LL*	$31,565.0$	s_3q_2	-0.009^{***}	*LL*	$31,619.1$

Note: *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.