Article:

Glass, A.J. and Kenjegalieva, K. (2019) A spatial productivity index in the presence of efficiency spillovers: Evidence for U.S. banks, 1992–2015. European Journal of Operational Research, 273 (3). pp. 1165-1179. ISSN: 0377-2217

https://doi.org/10.1016/j.ejor.2018.09.011

A Spatial Productivity Index in the Presence of Allocative Inefficiency: Evidence for U.S. Banks, $1992 - 2015$

Anthony J. Glass* and Karligash Kenjegalieva†‡

May 2017

Abstract

This paper presents the methodology for a new spatial decomposition of total factor productivity (TFP) growth. We also provide an empirical application to demonstrate the steps involved in the practical implementation of our new decomposition. As a result our paper makes a substantial contribution to the literature on the spatial decomposition of TFP growth which is a vastly underdeveloped literature as there is presently just one short study in the area. In particular, our paper develops this sparse literature in four respects which are varied in nature. Firstly, we introduce a cost efficiency spillover growth component. Secondly, we include own and spillover allocative efficiency growth components. Thirdly, we provide a much more detailed coverage of the spatial decomposition of TFP growth than the short communication in the extant literature. Fourthly, in contrast to the only other study in this area where the empirical application is very traditional as it uses data for geographical areas (cities, regions, etc.), we apply our spatial TFP growth decomposition using firm level data, which suggests that there can be an important future role for spatial efficiency and productivity analysis in OR. Our empirical application focuses on U.S. banks over the period $1992 - 2015$, which is an interesting study period as it includes the period pertaining to the financial crisis. Among other things, we find that, on average, a large U.S. bank's TFP since the financial crisis has become much more dependent on the bank itself and less so on spatial spillovers.

Key words: (D) Productivity and competitiveness; Spatial decomposition of TFP growth; Allocative efficiency growth; Efficiency spillovers; U.S. banks.

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

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

‡Corresponding author

1 Introduction

Although there is a well-established non-spatial methodological literature on the decomposition of total factor productivity (TFP) growth (e.g., Diewert and Fox, 2017; O'Donnell, 2016; Sun *et al.*, 2015; Oude Lansink *et al.*, 2015; and Mukherjee *et al.*, 2001) and a burgeoning literature on spatial stochastic frontier modeling (Druska and Horrace, 2004; Glass *et al.*, 2013; 2014; 2016a; 2016b; Orea *et al.*, 2016; Tsionas and Michaelides, 2016), there is only one short paper by Glass *et al.* (2013) (GKPF from hereon) on spatial TFP growth decomposition in the presence of spillovers. Our paper therefore makes a substantive contribution to the sparse literature on the spatial decomposition of TFP growth as we extend the GKPF analysis in four respects. Using the non-spatial case as a starting point, from a fitted primal stochastic frontier model (i.e., one where input prices do not feature in the technology such as production and input and output distance frontiers), the non-spatial generalized Malmquist decomposition of TFP growth (Orea, 2002) consists of a technical change component, returns to scale change and the change in technical efficiency. There is not, however, a direct correspondence between this non-spatial decomposition and the spatial generalized Malmquist TFP growth decomposition in GKPF, which comprises own and spillover technical change components, own and spillover returns to scale changes and the change in own technical efficiency. This is because the TFP growth decomposition in GKPF does not include the change in technical efficiency spillovers so strictly speaking they propose a partial spatial TFP growth decomposition. Our first extension of GKPF therefore is a methodological one as we augment their TFP growth decomposition with an efficiency spillover change component, which in our empirical application is the cost efficiency spillover change. We introduce the cost efficiency spillover change by computing efficiency spillovers using the method set out in Glass *et al.* (2016b). As we estimate a spatial stochastic cost frontier, in the spirit of the non-spatial TFP growth decomposition in Bauer (1990), our second extension of GKPF is also methodological as we further augment their TFP growth decomposition with own and spillover allocative efficiency changes.

Our third extension of the short communication in GKPF is our detailed coverage of the spatial decomposition of TFP growth and our fourth extension relates to our empirical application. Whereas the empirical application in GKPF is traditional as it applies spatial methods using data for geographical areas (cities, regions, etc.), we, on the other hand, apply our spatial TFP growth decomposition using firm level data, which serves to highlight the future relevance of spatial efficiency and productivity analysis in OR.[1] In particular, our empirical application analyzes U.S. banks over the period $1992 - 2015$.[2]

[1] Specifically, GKPF apply their spatial TFP growth decomposition to a spatial aggregate production frontier for European countries.

[2] Outside the OR literature there are a small number of applications of spatial stochastic frontier modeling (but not a spatial TFP growth decomposition) using firm level data. These applications are

This is a very interesting application as our sample includes the period pertaining to the financial crisis. Our empirical application can therefore shed light on the implications of the tighter post-crisis regulatory regime for changes in the own and spillover cost and allocative efficiencies of U.S. banks.[3]

Since to apply the spatial TFP growth decomposition that we develop we must first estimate, for reasons that we will explain in due course, a spatial stochastic frontier model which contains at least the SAR variable (i.e., the spatial lag of the dependent variable), a discussion of the evolution of the expanding methodological literature on spatial stochastic frontier modeling is of order. Interestingly, the methodological literature on spatial stochastic frontier modeling is beginning to mirror that for non-spatial stochastic frontier models. This is because spatial stochastic frontier studies can be split into two groups. The first group of studies contain the initial contributions on spatial stochastic frontier modeling and use a different approach to the one we utilize in this paper. This group of studies estimate one-way spatial panel models and compute efficiency from the cross-sectional specific effects. The first such study is Druska and Horrace (2004). By extending the cross-sectional spatial error model in Kelejian and Prucha (1999) they develop a GMM stochastic frontier model with fixed effects. Using the fixed effects they calculate time-invariant efficiency by applying the Schmidt and Sickles (1984) efficiency estimator, which assumes a composed error structure with idiosyncratic error and time-invariant inefficiency components. GKPF is the same type of study as they use the fixed effects from a one-way SAR model to calculate time-variant efficiency using the method in Cornwell *et al.* (1990).[4]

The second group of spatial stochastic frontier studies, which is the group this paper belongs to, follows the vast majority of the non-spatial stochastic frontier literature by computing efficiency via distributional assumptions to distinguish between the idiosyncratic error and inefficiency components of the composed disturbance. One such study is Tsionas and Michaelides (2016) who develop a Bayesian estimator of what we refer to as a spatial inefficiency model, which is a form of spatial error stochastic frontier model as it contains a spatial lag of the vector of inefficiencies. Two other studies that belong to this second group are both by Glass *et al.* (2016a; 2016b). Both these studies propose a panel data spatial Durbin stochastic frontier model which is a SAR stochastic frontier model augmented with exogenous spatial lags of the independent variables.[5] To minimize issues

exclusively in the economics literature and include Druska and Horrace (2004) who focus on Indonesian rice farms and Orea *et al.* (2016) who use data on Norwegian electricity distribution utilities. The empirical application part of our paper, however, represents the first application of spatial stochastic frontier modeling to banks.

[3] The post-crisis tightening of the regulation of U.S. banks is the result of the Dodd-Frank regulatory reforms which came into being in 2010.

[4] In spatial stochastic frontier modeling the spatial autocorrelated error (see Druska and Horrace, 2004) and the SAR variable (see GKPF) are endogenous which is accounted for in the estimation.

[5] Orea *et al.* (2016) also estimate a spatial Durbin stochastic frontier specification by making distributional assumptions to distinguish between the idiosyncratic error and inefficiency. As they note,

relating to convergence both these studies adopt a pseudo maximum likelihood (PML) estimator. In contrast to one-step full information ML (FML) estimation of the models PML involves estimating the models in steps.[6] In Glass *et al.* (2016a) a two-step PML procedure is used which involves estimating a non-frontier spatial Durbin model in the first step and in the second step splitting the composed disturbance into the idiosyncratic error and time-variant inefficiency. This model has been extended by Glass *et al.* (2016b) to simultaneously include time-invariant and time-variant inefficiency components and via random effects is to the best of our knowledge the first spatial stochastic frontier model to account for unobserved heterogeneity. We therefore use this set-up as the basis for the development of our spatial TFP growth decomposition.

More specifically, we estimate a spatial Durbin stochastic frontier specification for four reasons. Firstly, it is well-established that the spatial Durbin specification nests the SAR and spatial error/inefficiency specifications so a spatial Durbin model is robust to misspecification of the global spatial dependence (i.e., modeling spatial error autorcorrelation instead of the true SAR dependence and vice-versa). Secondly, although a spatial error/inefficiency specification and SAR and spatial Durbin specifications all account for global spatial dependence (1st order through to $(N-1)$th order spatial interaction), with the spatial error/inefficiency specification the spillover elasticity relates to the disturbance/inefficiency, whereas as we require for our spatial TFP growth decomposition, the spillover elasticities from the SAR and spatial Durbin specifications can be related to the exogenous regressors. As a result of this property from SAR and spatial Durbin specifications we obtain the spillover technical change and change in spillover returns of scale components of our spatial TFP growth decomposition. Thirdly, we favor a spatial Durbin specification over a SAR model because with the latter the ratio of the own and spillover elasticities (referred to in the spatial econometrics literature as direct and indirect elasticities, respectively) is the same for all the exogenous regressors which is implausible. This is not the case with the spatial Durbin specification because of the presence of the spatial lags of the exogenous regressors in the model. Fourthly, first impressions may suggest that we can only compute the change in spillover cost efficiency component of our spatial TFP growth decomposition from a spatial inefficiency model but this is not the case because, as we show in this paper, spillover cost efficiency can be obtained from the reduced form of the spatial Durbin specification.

By way of an insight into our empirical findings, among other things, we find that in terms of allocative efficiency large U.S. banks responded appropriately to the financial crisis. This is because the crisis, which large U.S. banks played a key role in, marked the beginning of a period of annual increases in the change in the average allocative efficiency

however, the spatial variables in their model do not have an economic interpretation as they are used as predictors of omitted non-spatial variables, whereas in Glass *et al.* (2016a) this is not the case so the spatial variables are causal.

[6] PML is sometimes referred to as quasi ML.

of large banks. Turning now to the structure of the remainder of this paper. In section 2 we set out the modeling framework which has four parts. In the first part we present the structural form of our random effects spatial Durbin stochastic cost frontier (SDCF) and in the second part we set out our PML procedure to estimate this model. The third part shows how we transform the structure form of our model into its reduced form which we then use to compute asymmetric flows of cost efficiency spillovers. These cost efficiency spillovers feature in our spatial TFP growth decomposition which we present in the fourth part of section 2. Moving on to section 3 which applies our modeling framework to U.S. banks over the period $1992 - 2015$. We then conclude in section 4.

2 Modeling Framework

2.1 Spatial Durbin Stochastic Cost Frontier (SDCF) Model with Random Effects

The structural form of the SDCF model with random effects that we estimate, where lower case letters denote logged variables, is as follows:

$$\begin{aligned}\widetilde{c}_{it} &= \alpha + TL\left(y_{it}, \widetilde{p}_{it}, t\right) + \gamma' z_{it} + STL\left(\sum_{j=1}^{N} w_{ij} y_{jt}, \sum_{j=1}^{N} w_{ij} \widetilde{p}_{jt}\right) + \gamma_s' \sum_{j=1}^{N} w_{ij} z_{jt} + \\ &\quad \delta \sum_{j=1}^{N} w_{ij} \widetilde{c}_{jt} + \kappa_i + v_{it} + \eta_i + u_{it}; \qquad (1) \\ &\kappa_i \sim N\left(0, \sigma_\kappa^2\right);\ \eta_i \sim N^+\left(0, \sigma_\eta^2\right);\ v_{it} \sim N\left(0, \sigma_v^2\right);\ u_{it} \sim N^+\left(0, \sigma_u^2\right).\end{aligned}$$

In each cross-section there are N units indexed $i = 1, ..., N$ that operate over T periods indexed $t = 1, ..., T$. Following the spatial econometrics literature and also the typical case that is encountered when using firm level data we focus on large N and small T. $TL\left(y_{it}, \widetilde{p}_{it}, t\right) = \rho t + \frac{1}{2}\varsigma t^2 + \zeta' \widetilde{p}_{it} + \varphi' y_{it} + \frac{1}{2} \widetilde{p}_{it}' \mathbf{\Theta} \widetilde{p}_{it} + \frac{1}{2} y_{it}' \mathbf{\Gamma} y_{it} + \widetilde{p}_{it}' \mathbf{\Psi} y_{it} + \varrho' \widetilde{p}_{it} t + \varpi' y_{it} t$ represents the variable returns to scale translog approximation of the log of the cost function technology. For the ith unit in period t p_{it} is the vector of observations for the input prices, which are indexed $k = 1, ..., K.$, and $\widetilde{p}_{it} = p_{it} - p_{Kit}$ denotes the $(1 \times (K-1))$ vector of observations for the normalized input prices. y_{it} is the $(1 \times M)$ vector of observations for the outputs, which are indexed $m = 1, ..., M$, $\widetilde{c}_{it} = c_{it} - p_{Kit}$ is an observation for normalized total cost and α is the intercept.

$\mathbf{W_N}$ is the $(N \times N)$ exogenous spatial weights matrix of non-negative constants w_{ij}. As $\mathbf{W_N}$ represents the spatial arrangement of the cross-sectional units and also the strength of the interaction among the units all the elements on the main diagonal of $\mathbf{W_N}$ are set to zero as a unit cannot be in its own neighborhood set. $\mathbf{W_N}$ is often populated using some measure of geographical proximity and must be specified *a priori* before estimation. Having specified $\mathbf{W_N}$ we can construct $\sum_{j=1}^{N} w_{ij}\widetilde{c}_{jt}$ which is the spatial lag of the dependent variable. This SAR variable is endogenous which we account for in the estimation and the associated SAR parameter $\delta \in (1/r_{\min}, 1/r_{\max})$, where $r_{\min}$ and $r_{\max}$ are the most negative and most positive real characteristic roots of $\mathbf{W_N}$, respectively. In our application $r_{\max} = 1$ as we use a row-normalized specification of $\mathbf{W_N}$.

In our model specification z_{it} is a vector of observations for the non-spatial regressors, $\sum_{j=1}^{N} w_{ij} z_{jt}$ is its spatial lag and $STL\left(\sum_{j=1}^{N} w_{ij} y_{jt}, \sum_{j=1}^{N} w_{ij}\widetilde{p}_{jt}\right)$ is the spatial lag of $TL\left(y_{jt}, \widetilde{p}_{jt}, t\right)$. We account for technical change through $TL\left(y_{jt}, \widetilde{p}_{jt}, t\right)$ via a non-linear time trend by including t and t^2 and by including the interactions ty_{it} and $t\widetilde{p}_{it}$ technical change is non-neutral. $\mathbf{W_N}t$ and $\mathbf{W_N}t^2$ are omitted from STL as they are collinear with t and t^2 in $TL\left(y_{jt}, \widetilde{p}_{jt}, t\right)$ (i.e., $t = \mathbf{W_N}t$ and $t^2 = \mathbf{W_N}t^2$) but the interactions with $\mathbf{W_N}t$ are retained. z_{it}, $\sum_{j=1}^{N} w_{ij} z_{jt}$, STL and $\sum_{j=1}^{N} w_{ij}\widetilde{c}_{jt}$ all shift the cost frontier technology in Eq. 1 but whereas the SAR variable accounts for endogenous global spatial dependence (1st order through to $(N-1)$th order neighbor effects), $\sum_{j=1}^{N} w_{ij} z_{jt}$ and STL are exogenous and account for only local spatial dependence (1st order neighbor effects).

In Eq. 1: ρ and $\frac{1}{2}\varsigma$ are regression parameters; ζ', φ', ϱ', ϖ', γ' and γ_s' are vectors of regression parameters, where a subscript s denotes local spatial parameters; and $\frac{1}{2}\mathbf{\Theta}$, $\frac{1}{2}\mathbf{\Gamma}$ and $\mathbf{\Psi}$ are matrices of the regression parameters $\frac{1}{2}\theta$, $\frac{1}{2}\tau$ and ψ, respectively. We also estimate the corresponding local spatial parameters for STL with the exception of ρ and ς. From the properties of the translog functional form (Christensen *et al.*, 1973) Eq. 1 is twice differentiable with respect to an output, a normalized input price and the spatial lags of an output and normalized input price, where the associated Hessians are symmetric because of the symmetry restrictions that are placed on the associated matrices of parameters (e.g., $\frac{1}{2}\tau_{1M} = \frac{1}{2}\tau_{M1}$ in $\frac{1}{2}\mathbf{\Gamma}$).

Our model specification is characterized by a four component error structure, $\varepsilon_{it}^{*} = \varepsilon_i + \varepsilon_{it} = \kappa_i + v_{it} + \eta_i + u_{it}$, where $\varepsilon_i = \kappa_i + \eta_i$ is the time-invariant component and $\varepsilon_{it} = v_{it} + u_{it}$ is the time-variant component. Since the estimator of our model rests on κ_i, v_{it}, η_i and u_{it} being i.i.d. across i and t or just i as is appropriate, where distributional assumptions distinguish between each error component, we account for unobserved heterogeneity using random effects. In Eq. 1 v_{it} is the idiosyncratic error and as is standard when modeling unobserved heterogeneity using random effects, the unit specific effect, κ_i, is a time-invariant random error. η_i is net time-invariant inefficiency (NII_i) and u_{it} is net time-variant inefficiency (NVI_{it}), both of which are bounded in the interval $[0, 1]$. Since Eq. 1 is in log form, using NVI_{it} and NII_i yields the combined measure of gross inefficiency, $GVI_{it} = \eta_i + u_{it} = NII_i + NVI_{it}$. GVI_{it} is also bounded in the interval $[0, 1]$ and is time-variant which emanates from NVI_{it}.[7] Both η_i and u_{it} are assumed to have

[7] In the literature on the corresponding non-spatial specification of our model GVI is referred to as overall time-variant inefficiency (Kumbhakar *et al.*, 2014). We, however, refer to it as gross inefficiency

a half-normal distribution which is a common distributional assumption for inefficiency in the stochastic frontier literature (e.g., Bos *et al.*, 2009; Greene, 2004). Our estimation procedure, however, is sufficiently general to accommodate alternative distributional assumptions for u_{it} and η_i. In the next subsection we provide an overview of technical issues relating to estimation where we cover how we obtain the estimates of u_{it} and η_i.

Testing the appropriateness of the error structure in our model specification for an empirical application involves applying the one-sided hypothesis test in Gouriéroux *et al.* (1982) to test for the presence of each of the four error components (κ, v, η and u). The test statistic has an asymptotic distribution that is a mixture of chi-squared distributions, $\frac{1}{2}\chi^2(0) + \frac{1}{2}\chi^2(1)$. For $G \in \{\kappa, v, \eta, u\}$ rejection of the null, $\widehat{\sigma}_G^2 = 0$, in favor of the alternative hypothesis, $\widehat{\sigma}_G^2 > 0$, constitutes evidence of the presence of the error component.[8] In an empirical setting the nulls for the inefficiency components may not both be rejected so the model specification in Eq. 1 has the appealing feature that it nests other models. For example, failure to reject the absence of η would lead to Eq. 1 collapsing to the SDCF extension of the non-spatial true random effects frontier model (Greene, 2005).

2.2 Overview of the Pseudo Maximum Likelihood (PML) Estimator and Estimation of the Own Cost Efficiencies

As this paper focuses on the development and application of the methodology for a spatial TFP growth decomposition in the presence of allocative inefficiency we only provide an overview of the procedure we use to estimate Eq. 1. See Glass *et al.* (2016b) for a detailed presentation of this estimation procedure. For simplicity we collect the observations for unit i in period t for all the exogenous regressors in a single vector x_{it} and we also collect the associated regression parameters in the vector β, where everything else is as previously defined for Eq. 1.

$$\widetilde{c}_{it} = \alpha + \beta' x_{it} + \delta \sum_{j=1}^{N} w_{ij} \widetilde{c}_{jt} + \kappa_i + v_{it} + \eta_i + u_{it}. \tag{2}$$

because based on the terminology used in the spatial econometrics literature having computed GVI we proceed to compute, among other things, direct, indirect and total GVI (see Glass *et al.*, 2016b). This avoids referring to the total GVI as the total overall time-variant inefficiency which is very confusing. Consequently, we refer to the η_i and u_{it} components of GVI as net inefficiencies as the former is net of time-variant inefficiency and the latter is net of time-invariant inefficiency. NII, NVI and GVI should also not be confused with the net and gross inefficiencies in Coelli *et al.* (1999) as the interpretations of net and gross in their set-up are entirely different.

[8] Andrews (2001) derives another relevant approach to test for the presence of each component of our error structure. The test statistic he derives allows for, firstly, the possibility that the parameter value lies on the boundary of the parameter space under the null and, secondly, the possible presence of a nuisance parameter under the alternative hypothesis. The asymptotic distribution of this test statistic is not a chi-squared distribution and involves semi-parametric simulation.

The general form of our model in Eq. 2 is the spatial Durbin counterpart of the non-spatial stochastic frontier model in Badunenko and Kumbhakar (2016; 2017), Colombi *et al.* (2014), Kumbhakar *et al.* (2014), Tsionas and Kumbhakar (2014) and Filippini and Greene (2016). Alternatively, if the local spatial regressors are omitted from Eq. 1 and thus x_{it} in Eq. 2 this general form becomes the SAR counterpart of the non-spatial specification.

Tsionas and Kumbhakar (2014) develop a one-step Bayesian estimator of the non-spatial counterpart of Eq. 2 and Colombi *et al.* (2014) estimate this non-spatial model using FML. Due to the sensitivity of the Bayesian approach to the choice of informative priors for the main objects of the estimation and concerns about the empirical tractability of the FML procedure, Filippini and Greene (2016) and Badunenko and Kumbhakar (2016; 2017) estimate this non-spatial model using one-step simulated ML. Eq. 2 is of course even more complex than its non-spatial counterpart because of the presence of the additional SAR parameter. For this reason we employ a PML estimator that involves estimating our model in steps using ML. See Kumbhakar *et al.* (2014) for the corresponding non-spatial estimator. In particular, we estimate Eq. 1 by maximizing three log-likelihood functions, one for each step. Step 1 estimates the non-frontier random effects spatial Durbin model which distinguishes between the time-invariant and time-variant components of the composed error. Step 2 splits the time-variant error into its constituent parts, v_{it} and u_{it}, and step 3 splits the time-invariant error into κ_i and η_i.

We begin our overview of the estimator we employ by applying a standard reparameterization in the stochastic frontier literature to the case in hand in Eq. 2 by using $\sigma^2_{\eta\kappa} = \sigma^2_\eta + \sigma^2_\kappa$ and $\lambda_{\eta\kappa} = \sigma_\eta/\sigma_\kappa$, and $\sigma^2_{uv} = \sigma^2_u + \sigma^2_v$ and $\lambda_{uv} = \sigma_u/\sigma_v$. Therefore, $\sigma^2_\eta = \sigma^2_{\eta\kappa}/\left(1+\lambda^2_{\eta\kappa}\right)$, $\sigma^2_\kappa = \sigma^2_{\eta\kappa}\lambda^2_{\eta\kappa}/\left(1+\lambda^2_{\eta\kappa}\right)$, $\sigma^2_u = \sigma^2_{uv}/\left(1+\lambda^2_{uv}\right)$ and $\sigma^2_v = \sigma^2_{uv}\lambda^2_{uv}/\left(1+\lambda^2_{uv}\right)$. We then transform in Eq. 2 the positively skewed time-invariant error, ε_i, the positively skewed time-variant error, ε_{it}, and the intercept as follows.

$$\widetilde{c}_{it} = \alpha^\circ + \beta' x_{it} + \delta \sum_{j=1}^{N} w_{ij}\widetilde{c}_{jt} + \varepsilon^\circ_i + \varepsilon^\circ_{it}, \tag{3}$$

where $\alpha^\circ = \alpha + \mu_{\varepsilon_i} + \mu_{\varepsilon_{it}}$, $\varepsilon^\circ_i = \kappa_i + \eta_i - \mu_{\varepsilon_i}$, $\varepsilon^\circ_{it} = v_{it} + u_{it} - \mu_{\varepsilon_{it}}$, $\mu_{\varepsilon_{it}} = \mathrm{E}(u_{it})$ and $\mu_{\varepsilon_i} = \mathrm{E}(\eta_i)$. Eq. 3 therefore has the form of the non-frontier random effects spatial Durbin model with time-invariant and time-variant error components, ε°_i and ε°_{it}, which satisfy the zero mean condition by construction.

The log-likelihood function to estimate Eq. 3 in step 1 is:

$$LL = -\frac{NT}{2}\log\left(2\pi\sigma^2\right) + T\log\left|\mathbf{I_N} - \delta\mathbf{W_N}\right| - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\widetilde{c}^{\bullet}_{it} - \delta\left(\sum_{j=1}^{N} w_{ij}\widetilde{c}_{jt}\right)^{\bullet} - \beta' x^{\bullet}_{it}\right]^2, \tag{4}$$

where $T \log |\mathbf{I_N} - \delta \mathbf{W_N}|$ is the contribution to the log-likelihood from the scaled logged determinant of the Jacobian of the transformation from $\varepsilon_{it}^{\bullet}$ to $\widetilde{c}_{it}^{\bullet}$. It is now standard in the spatial literature to use Eq. 4 for ML estimation of non-frontier random effects SAR and spatial Durbin models where the transformation from $\varepsilon_{it}^{\bullet}$ to $\widetilde{c}_{it}^{\bullet}$ accounts for the endogeneity of the SAR variable (Elhorst, 2009). $\bullet$ denotes a transformation of variables which is dependent on the weight that is attached to the cross-sectional component of the data, ϑ, where $0 < \vartheta^2 = \sigma_{uv}^2 / \left(T\sigma_{\eta\kappa}^2 + \sigma_{uv}^2 \right) \leq 1$. Using this notation the transformed data is expressed in quasi-differenced form as:

$$\widetilde{c}_{it}^{\bullet} = \widetilde{c}_{it} - (1 - \vartheta) \frac{1}{T} \sum_{t=1}^{T} \widetilde{c}_{it}, \tag{5}$$

$$\left(\sum_{j=1}^{N} w_{ij} \widetilde{c}_{jt} \right)^{\bullet} = \sum_{j=1}^{N} w_{ij} \widetilde{c}_{jt} - (1 - \vartheta) \frac{1}{T} \sum_{j=1}^{N} w_{ij} \widetilde{c}_{jt}, \tag{6}$$

$$x_{it}^{\bullet} = x_{it} - (1 - \vartheta) \frac{1}{T} \sum_{t=1}^{T} x_{it}. \tag{7}$$

Step 2 of our PML estimator estimates λ_{uv} and u_{it}. To estimates λ_{uv} we maximize the concentrated log-likelihood function in Eq. 8 which involves substituting in from step 1 $\widehat{\varepsilon}_{it}^{\circ}$ for ε_{it}° and substituting in for δ and β' in Eq. 9 $\widehat{\delta}$ and $\widehat{\beta}'$ to obtain $\widehat{\sigma}_{uv}$ which we then substitute in for σ_{uv} in Eq. 8.

$$LL(\lambda_{uv}) = -NT \ln \sigma_{uv} + \sum_{i=1}^{N} \sum_{t=1}^{T} \ln \left[1 - \Phi \left(\frac{\varepsilon_{it}^{\circ} \lambda_{uv}}{\sigma_{uv}} \right) \right] - \frac{1}{2\sigma_{uv}^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{it}^{\circ 2}, \tag{8}$$

where Φ denotes the standard normal cumulative distribution function and

$$\widehat{\sigma}_{uv} = \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\widetilde{c}_{it}^{\bullet} - \delta \left(\sum_{j=1}^{N} w_{ij} \widetilde{c}_{jt} \right)^{\bullet} - \beta' x_{it}^{\bullet} \right]^2 / \left[1 - 2\lambda_{uv}^2 / \pi \left(1 + \lambda_{uv}^2 \right) \right] \right)^{1/2}. \tag{9}$$

The solution to $LL(\lambda_{uv})$ is the ML estimate $\widehat{\lambda}_{uv}$ which we substitute into Eq. 9 to obtain the ML estimate $\widehat{\sigma}_{uv}$.

The consistent estimate of the constant term, α, is then scaled up by the value of $\widehat{\mu}_{e_{it}} \left(\widehat{\lambda}_{uv}, \widehat{\sigma}_{uv}^2 \right) = \widehat{u}_{it}$. For stochastic cost and output distance frontiers the productive unit is assumed to be minimizing the objective variable or vector of objective variables, as is appropriate, so for these technologies the constant is scaled up. In contrast, for stochastic production, revenue, profit and input distance frontiers the constant is scaled down by $\widehat{\mu}_{\varepsilon_{it}} \left(\widehat{\lambda}_{uv}, \widehat{\sigma}_{uv}^2 \right) = \widehat{u}_{it}$. To compute $\widehat{u}_{it}$ we use the following Jondrow *et al.* (1982) (JMLS) method, which involves, as we have done in other parts of our PML estimator above, substituting in the relevant estimates into Eq. 10.

$$\widehat{u}_{it} = \mathrm{E}\left(u_{it}|\varepsilon_{it}^{\circ}\right) = \frac{\sigma_u \sigma_v}{\sigma_{uv}} \left(\frac{\phi_{it}}{1-\Phi_{it}} - \frac{\varepsilon_{it}^{\circ}\lambda_{uv}}{\sigma_{uv}} \right), \tag{10}$$

where $\Phi_{it} = \Phi\left(\varepsilon_{it}^{\circ}\lambda_{uv}/\sigma_{uv}\right)$, $\phi_{it} = \phi\left(\varepsilon_{it}^{\circ}\lambda_{uv}/\sigma_{uv}\right)$, Φ is as previously defined and ϕ is the probability density function for the standardized normal distribution.

In step 3 of our PML estimation routine we estimate $\lambda_{\eta\kappa}$ and η_i using the corresponding step 2 procedures to estimate λ_{uv} and u_{it}. In short this involves using a similar ML estimator to that in step 2 to estimate $\lambda_{\eta\kappa}$ using $\widehat{\varepsilon}_i^{\circ}$ from step 1. The ML estimate $\widehat{\lambda}_{\eta\kappa}$ is then used to compute the ML estimate $\widehat{\sigma}_{\eta\kappa}$. There is a also a further scaling up of α by $\widehat{\mu}_{\varepsilon_i}\left(\widehat{\lambda}_{\eta\kappa}, \widehat{\sigma}_{\eta\kappa}^2\right) = \widehat{\eta}_i$ and as in step 2 we compute the estimate of η_i using the JMLS method.

As a result of our SDCF in Eq. 1 being in logged form $NVE_{it} = \exp\left(u_{it}\right)$, $NIE_i = \exp\left(\eta_i\right)$ and $GVE_{it} = \exp\left(\eta_i + u_{it}\right) = NIE_i * NVE_{it}$ are the net time-variant, net time-invariant and gross time-variant efficiencies. Even though we compute NVE_{it}, NIE_i and GVE_{it} from a spatial model it should be noted that they are own efficiency measures and do not include any efficiency spillovers across the system/network as they are computed from the structural form of our model specification. Efficiency spillovers across the system/network are obtained from the reduced form of our spatial model, which is what we provide an overview of in the next subsection.

2.3 Elasticities and Asymmetric Flows of Cost Efficiency Spillovers

It is now well-established for models that contain the SAR variable such as Eq. 1 that the fitted parameters for the exogenous regressors are not elasticities. This is because the elasticity for an exogenous regressor is a function of the SAR parameter. To disentangle the effect of an exogenous regressor from the effect of the SAR variable it is now standard in spatial econometrics to calculate direct, indirect and total elasticities using the fitted parameters from a model such as Eq. 1. A direct elasticity is interpreted in the same way as an elasticity from a non-spatial model, although a direct elasticity takes into account feedback effects which occur via the spatial multiplier matrix. Feedback is the effect of a change in an independent variable of a particular unit which reverberates back to the same unit's dependent variable through its effect on the dependent variables of the other units in the sample. An indirect elasticity can be calculated in two ways yielding the same numerical value. This leads to two interpretations of an indirect elasticity: (i) average change in the dependent variable of all the other units in the sample following a change in an independent variable for one particular unit; or (ii) average change in the dependent variable for a particular unit following a change in an independent variable for all the other units in the sample. The total elasticity is the sum of the direct and indirect elasticities.

Calculation of the direct, indirect and total elasticities is based on the reduced form of a spatial model. Accordingly, we rewrite Eq. 1 in its reduced form, where we drop the i subscripts to denote vectors of successively stacked cross-sectional observations. This reduced form in Eq. 11 is also used in part to obtain the absolute direct, indirect and total efficiencies.

$$\widetilde{c}_t = \left(\mathbf{I_N} - \delta \mathbf{W_N}\right)^{-1} \begin{pmatrix} \alpha\iota + TL\left(y_t, \widetilde{p}_t, t\right) + \gamma' z_t + STL\left(\mathbf{W_N} y_t, \mathbf{W_N}\widetilde{p}_t\right) + \\ \xi' \mathbf{W_N} z_t + \kappa + v_t + \eta + u_t \end{pmatrix}, \tag{11}$$

where $\left(\mathbf{I_N} - \delta \mathbf{W_N}\right)^{-1}$ is the spatial multiplier matrix, ι is an $(N \times 1)$ vector of ones and everything else is as previously defined for Eq. 1.

We set out the approach to calculate the direct, indirect and total elasticities at the sample mean for a first order output which we denote $y_{m,t}$. From the local spatial counterpart of Eq. 1 (i.e., Eq. 1 with the SAR variable omitted), which would only capture first order neighbor effects, if we use mean adjusted data all the fitted parameters from this local spatial model are elasticities. This is because at the sample mean the own and local spatial quadratic and interaction terms are zero. Extending this to Eq. 11 the fitted φ_m and $\varphi_{s,m}$ parameters for $y_{m,t}$ and $\mathbf{W_N} y_{m,t}$ can be used to directly calculate the direct, indirect and total elasticities for $y_{m,t}$ at the sample mean. Differentiating Eq. 11 with respect to $y_{m,t}$ as follows yields a matrix of direct and indirect elasticities for each unit, where the right-hand side of Eq. 13 is independent of the time index.

$$\left[\begin{array}{ccc} \frac{\partial \widetilde{c}}{\partial y_{m,1}}, & \cdots & , \frac{\partial \widetilde{c}}{\partial y_{m,N}} \end{array} \right]_t = \begin{bmatrix} \frac{\partial \widetilde{c}_1}{\partial y_{m,1}} & \cdots & \frac{\partial \widetilde{c}_1}{\partial y_{m,N}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \widetilde{c}_N}{\partial y_{m,1}} & \cdots & \frac{\partial \widetilde{c}_N}{\partial y_{m,N}} \end{bmatrix}_t \tag{12}$$

$$= \left(\mathbf{I_N} - \delta \mathbf{W_N}\right)^{-1} \begin{bmatrix} \varphi_m & \cdots & w_{1N}\varphi_{s,m} \\ \vdots & \ddots & \vdots \\ w_{N1}\varphi_{s,m} & \cdots & \varphi_m \end{bmatrix}. \tag{13}$$

Since Eq. 13 yields different direct and indirect elasticities for each unit, to facilitate interpretation we report the mean direct elasticity (average of the diagonal elements of Eq. 13) and the mean indirect elasticity (this average spillover elasticity to a unit is the average row sum of the off-diagonal elements of Eq. 13 and is numerically the same as the average spillover elasticity from a unit which is the average column sum of the off-diagonal elements of Eq. 13).[9] For t and t^2, whose spatial lags are omitted from Eq. 1, the mean direct, mean indirect and mean total elasticities are calculated using Eq. 13 but with the off-diagonal elements in the second matrix set equal to zero by construction.

[9] We compute the associated $t-$statistics by Monte Carlo simulation of the distributions of the mean direct, mean indirect and mean total elasticities.

Turning now to an overview of the salient features of absolute direct, indirect and total efficiencies. For a detailed coverage of these efficiencies see Glass *et al.* (2016b), which is an extension of the first contribution on efficiency spillovers (Glass *et al.*, 2016a). Direct efficiency for a unit is interpreted in the same way as own efficiency from a non-spatial model. In contrast to own efficiency from a non-spatial or a spatial stochastic frontier model, however, direct efficiency is own efficiency plus efficiency feedback. Efficiency feedback is the component of a unit's direct efficiency which due to the spatial multiplier matrix has rebounded back to the unit having passed through 1st order and higher order neighbors. As was the case for an indirect elasticity, indirect efficiency can be interpreted in two ways: (i) the sum of the efficiency spillovers to a unit from all the other units in the sample; and (ii) the sum of efficiency spillovers from a unit to all the other units in the sample. When these two indirect efficiencies are averaged across the sample they will yield the same numerical value but they will be asymmetric for individual units. In the same way as a total elasticity is calculated, a unit's total efficiency is the sum of its direct and indirect efficiencies. Due to there being two asymmetric indirect efficiencies for a unit, there are two asymmetric total efficiencies for each unit. When the two total efficiencies are averaged across the sample, however, they yield the same numerical value.

From the reduced form of our model in Eq. 11 we recognize that $(\mathbf{I_N} - \delta\mathbf{W_N})^{-1}\eta = \eta_{To}^{Tot}$ and $(\mathbf{I_N} - \delta\mathbf{W_N})^{-1}u_t = u_{t,\,To}^{Tot}$, where η_{To}^{Tot} and $u_{t,\,To}^{Tot}$ are $(N \times 1)$ vectors of total NII and total NVI, respectively. The subscript To denotes that the inefficiency spillovers used in the calculation of these total inefficiency vectors are inefficiency spillovers which come to the ith unit from all the jth units in the sample for $i \neq j$. Since from above $GVI_t = NII + NVI_t$ is own gross inefficiency, $(\mathbf{I_N} - \delta\mathbf{W_N})^{-1}(\eta + u_t) = GVI_{t,\,To}^{Tot}$ is the $(N \times 1)$ vector of total GVI. Based on own $NVE_t = \exp(u_t)$ and own $NIE = \exp(\eta)$ from above, the vectors of total efficiencies that directly correspond to the vectors of total inefficiencies η_{To}^{Tot} and $u_{t,\,To}^{Tot}$ are $(\mathbf{I_N} - \delta\mathbf{W_N})^{-1}\exp(\eta) = NIE_{To}^{Tot}$ and $(\mathbf{I_N} - \delta\mathbf{W_N})^{-1}\exp(u_t) = NVE_{t,\,To}^{Tot}$. In similar vein, based on $GVE_t = NIE * NVE_t = \exp(\eta + u_t)$ from above, the vector of total gross efficiencies that directly corresponds to $GVI_{t,\,To}^{Tot}$ is $(\mathbf{I_N} - \delta\mathbf{W_N})^{-1}\exp(\eta + u_t) = GVE_{t,\,To}^{Tot}$. Intuitively, a unit's NIE_{To}^{Tot}, $NVE_{t,\,To}^{Tot}$ and $GVE_{t,\,To}^{Tot}$ measure its NIE, NVE and GVE across a system/network, where in our empirical analysis we apply these measures to obtain cost efficiencies for individual banks' across the U.S. banking system. An appealing feature of the computation of these total efficiencies is that they can be additively decomposed into their direct and indirect components. See Glass *et al.* (2016b) for details on this.

In our empirical application $\mathbf{W_N}$ is asymmetric which is typically the case in spatial econometrics. If $\mathbf{W_N}$ is asymmetric $(\mathbf{I_N} - \delta\mathbf{W_N})^{-1}$ will also be asymmetric resulting in asymmetric indirect NIE, NVE and GVE spillovers to and from a unit. Furthermore, in contrast to the own NIE, NVE and GVE from Eq. 1, the direct, indirect and total NIE, NVE and GVE from the reduced form of our model all include some form of efficiency spillover. In line with the own NIE, NVE and GVE from Eq. 1, the lower bound of the direct, indirect and total NIE, NVE and GVE scores from the reduced form of our model is of course 0. Other than that direct, indirect and total NIE, NVE and GVE are unbounded. This in no way precludes these efficiencies from being incorporated into a spatial TFP growth decomposition because our direct, indirect and total NIE, NVE and GVE measures are easily interpretable as they are percentages. This is because direct, indirect and total NIE, NVE and GVE are scaled own NIE, NVE and GVE. As a result these direct, indirect and total efficiencies have a simple interpretation as they are relative to the own NIE, NVE and GVE benchmarks. The magnitude of the scaling relates to the magnitude of the efficiency spillover that is included in the direct, indirect and total NIE, NVE and GVE. If the magnitude of the efficiency spillover is sufficiently large a direct/indirect/total NIE, NVE or GVE score will be greater than 1. If this is the case the efficiency spillover has pushed the unit beyond the best practice frontier for the relevant own efficiency from Eq. 1.

2.4 Spatial TFP Growth with Spatial Efficiency Change Components

The starting point for the calculation of the components of the spatial TFP index that we propose are the estimated direct, indirect and total models which have the following forms. See Glass *et al.* (2015) for details on how the presence of the SAR variable in a model gives rise to estimates of the direct, indirect and total models (referred to in their paper as estimates of the internal, external and total models):

$$\begin{aligned} c_{it}^{Dir} = {} & \rho^{Dir} t + \frac{1}{2}\varsigma^{Dir} t^2 + \zeta^{Dir'} \widetilde{p}_{it} + \varphi^{Dir'} y_{it} + \frac{1}{2}\widetilde{p}_{it}' \mathbf{\Theta}^{\mathbf{Dir}} \widetilde{p}_{it} + \frac{1}{2} y_{it}' \mathbf{\Gamma}^{\mathbf{Dir}} y_{it} + \\ & p_{it}' \mathbf{\Psi}^{\mathbf{Dir}} y_{it} + \varrho^{Dir'} \widetilde{p}_{it} t + \varpi^{Dir'} y_{it} t + \gamma^{Dir'} z_{it} + GVI_{it}^{Dir}. \end{aligned} \tag{14}$$

$$\begin{aligned} c_{it}^{Ind} = {} & \rho^{Ind} \sum_{j=1}^{N} w_{ij} t_j + \frac{1}{2}\varsigma^{Ind} \sum_{j=1}^{N} w_{ij} t_j^2 + \zeta^{Ind'} \sum_{j=1}^{N} w_{ij} \widetilde{p}_{jt} + \varphi^{Ind'} \sum_{j=1}^{N} w_{ij} y_{jt} + \\ & \frac{1}{2} \sum_{j=1}^{N} w_{ij} \widetilde{p}_{jt}' \mathbf{\Theta}^{\mathbf{Ind}} \sum_{j=1}^{N} w_{ij} \widetilde{p}_{jt} + \frac{1}{2} \sum_{j=1}^{N} w_{ij} y_{jt}' \mathbf{\Gamma}^{\mathbf{Ind}} \sum_{j=1}^{N} w_{ij} y_{jt} + \sum_{j=1}^{N} w_{ij} p_{jt}' \mathbf{\Psi}^{\mathbf{Ind}} \sum_{j=1}^{N} w_{ij} y_{jt} + \\ & \varrho^{Ind'} \sum_{j=1}^{N} w_{ij} \widetilde{p}_{jt} \sum_{j=1}^{N} w_{ij} t_j + \varpi^{Ind'} \sum_{j=1}^{N} w_{ij} y_{jt} t_j + \gamma^{Ind'} \sum_{j=1}^{N} w_{ij} z_{jt} + GVI_{t,\,To}^{Ind} \\ & \forall\; i \neq j. \end{aligned} \tag{15}$$

$$c_{it}^{Tot} = \rho^{Tot}\sum_{j=1}^{N} w_{ij}t_j + \frac{1}{2}\varsigma^{Tot}\sum_{j=1}^{N} w_{ij}t_j^2 + \zeta^{Tot'}\sum_{j=1}^{N} w_{ij}\widetilde{p}_{jt} + \varphi^{Tot'}\sum_{j=1}^{N} w_{ij}y_{jt}+$$
$$\frac{1}{2}\sum_{j=1}^{N} w_{ij}\widetilde{p}'_{jt}\mathbf{\Theta}^{\mathbf{Tot}}\sum_{j=1}^{N} w_{ij}\widetilde{p}_{jt} + \frac{1}{2}\sum_{j=1}^{N} w_{ij}y'_{jt}\mathbf{\Gamma}^{\mathbf{Tot}}\sum_{j=1}^{N} w_{ij}y_{jt} + \sum_{j=1}^{N} w_{ij}p'_{jt}\mathbf{\Psi}^{\mathbf{Tot}}\sum_{j=1}^{N} w_{ij}y_{jt}+$$
$$\varrho^{Tot'}\sum_{j=1}^{N} w_{ij}\widetilde{p}_{jt}\sum_{j=1}^{N} w_{ij}t_j + \varpi^{Tot'}\sum_{j=1}^{N} w_{ij}y_{jt}t_j + \gamma^{Tot'}\sum_{j=1}^{N} w_{ij}z_{jt} + GVI_{t,\ To}^{Tot}\ \forall i,\ j, \quad (16)$$

where $c_{it}^{Tot} = c_{it}^{Dir} + c_{it}^{Ind}$ and although we do not observe c_{it}^{Dir}, c_{it}^{Ind} and c_{it}^{Tot}, if necessary, they can be estimated using the estimates of Eqs. $14 - 16$.

The decomposition of our spatial TFP index is in the spirit of the non-spatial TFP growth decompositions in Bauer (1990) and Orea (2002). In particular, we decompose spatial total TFP growth, which we denote as ΔTFP_{it+1}^{Tot}, into the four total components in Eq. 17, where the corresponding four non-spatial components represent the non-spatial version of our TFP growth decomposition in the literature. Each of these total components can be decomposed into a direct component for a unit which includes feedback and an indirect component that accounts for spatial spillovers.

$$\Delta TFP_{it+1}^{Tot} = \Delta TA_{it+1}^{Tot} + \Delta GVE_{it+1,To}^{Tot} + \Delta AE_{it+1}^{Tot} + \Delta RSE_{it+1}^{Tot}, \quad (17)$$

where ΔTA_{it+1}^{Tot} is the total technical advancement component; $\Delta GVE_{it+1,To}^{Tot}$ is total gross cost efficiency change, which we calculate using the indirect GVE spillovers that come to the ith unit from all the jth units as we are interested in TFP growth for the ith unit; ΔAE_{it+1}^{Tot} is total allocative efficiency change; and ΔRSE_{it+1}^{Tot} is returns to scale efficiency change. We calculate these components of ΔTFP_{it+1}^{Tot} as follows:

1.

$$\Delta TA_{it+1}^{Tot} = -\frac{1}{2}\left(\frac{\partial c_{it+1}^{Tot}}{\partial t} + \frac{\partial c_{it}^{Tot}}{\partial t}\right) = -\frac{1}{2}\left[\left(\frac{\partial c_{it+1}^{Dir}}{\partial t} + \frac{\partial c_{it+1}^{Ind}}{\partial t}\right) + \left(\frac{\partial c_{it}^{Dir}}{\partial t} + \frac{\partial c_{it}^{Ind}}{\partial t}\right)\right]. \quad (18)$$

2.

$$\Delta GVE_{it+1,To}^{Tot} = \left(GVE_{it+1,To}^{Dir} + GVE_{it+1,To}^{Ind}\right) - \left(GVE_{it,To}^{Dir} + GVE_{it,To}^{Ind}\right). \quad (19)$$

3.

$$\Delta AE^{Tot}_{it+1} = \frac{1}{2}\sum_{k=1}^{K}\left[\left(s^{Tot}_{kit+1} - e^{Tot}_{kit+1}\right) + \left(s^{Tot}_{kit} - e^{Tot}_{kit}\right)\right]\left(p^{Tot}_{kit+1} - p^{Tot}_{kit}\right) \tag{20}$$

$$= \frac{1}{2}\sum_{k=1}^{K}\left(\begin{array}{c}\left[\left(s^{Dir}_{kit+1} - e^{Dir}_{kit+1}\right) + \left(s^{Dir}_{kit} - e^{Dir}_{kit}\right)\right]\left(p^{Dir}_{kit+1} - p^{Dir}_{kit}\right) + \\ \left[\left(s^{Ind}_{kit+1} - e^{Ind}_{kit+1}\right) + \left(s^{Ind}_{kit} - e^{Ind}_{kit}\right)\right]\left(p^{Ind}_{kit+1} - p^{Ind}_{kit}\right)\end{array}\right), \tag{21}$$

where $p^{Dir}_{kit} = p_{kit}$, $p^{Ind}_{kit} = \sum_{j=1}^{N} w_{ij}p_{kjt}\ \forall\ i \neq j$ and $p^{Tot}_{kit} = p_{kit} + \sum_{j=1}^{N} w_{ij}p_{kjt}\ \forall\ i \neq j$ are the direct, indirect and total prices, respectively. This is to say that we compute p^{Dir}_{kit}, p^{Ind}_{kit} and p^{Tot}_{kit} from the data, where p^{Dir}_{kit} includes a component which feeds back to the ith unit from the other units, which will be inherent in the data for p_{kit}. Moreover, e^{Dir}_{kit}, e^{Ind}_{kit} and e^{Tot}_{kit} are scaled direct, indirect and total elasticities for the kth input price. These elasticities are the optimal cost shares as the direct, indirect and total input price elasticities from Eqs. $14 - 16$ are scaled proportionally to sum to 1. We must scale up/down the direct, indirect and total input price elasticities because although the structural form of our model in Eq. 1 is characterized by homogeneity of degree 1 in input prices this is not the case for the estimated direct, indirect and total models.

Turning now to s^{Dir}_{kit}, s^{Ind}_{kit} and s^{Tot}_{kit} which are direct, indirect and total input expenditure share weights from the data. The calculation of s^{Dir}_{kit} is based on $p_{kit}q_{kit}$, where q_{kit} is the kth input quantity. Additionally because in the empirical application we use specifications of $\mathbf{W_N}$ that reflect the systemic nature of the U.S. banking industry by assuming some degree of spatial interaction between every bank in each sample, $\sum_{j=1}^{N} w_{ij}p_{kjt}q_{kjt}$ is the basis of the calculation of $s^{Ind}_{kit}\ \forall\ i \neq j$. We are therefore taking the view that the indirect spillover share for an individual bank is a spatially weighted average of the shares of other banks in the sample. From the calculations of s^{Dir}_{kit} and s^{Ind}_{kit} it follows that $p_{kit}q_{kit} + \sum_{j=1}^{N} w_{ij}p_{kjt}q_{kjt}\ \forall\ i, j$ is the basis of the calculation of s^{Tot}_{kit}.[10]

[10] Although outside the scope of this paper an alternative possible approach to compute the data informed actual indirect input expenditure weights, s^{Ind}_{kit}, which are needed to calculate ΔAE^{Ind}_{it+1}, is the least absolute shrinkage and selection operator (Lasso). Such an approach would resemble the use of this estimator in the spatial literature to obtain a data informed specification of $\mathbf{W_N}$. That said, since the Lasso method has been exclusively used in this context in methodological spatial studies its empirical tractability needs to be established.

4.

$$\Delta RSE_{it+1}^{Tot} = \frac{1}{2}\sum_{m=1}^{M}\left(\begin{array}{c}\left[\left(1-\Omega_{it+1}^{Tot}\right)e_{mit+1}^{Tot}/\Omega_{it+1}^{Tot}\right]+\\ \left[\left(1-\Omega_{it}^{Tot}\right)e_{mit}^{Tot}/\Omega_{it}^{Tot}\right]\end{array}\right)\left(\sum_{j=1}^{N}y_{mjt+1}-\sum_{j=1}^{N}y_{mjt}\right)\ \forall\ i,j \tag{22}$$

$$= \frac{1}{2}\sum_{m=1}^{M}\left[\begin{array}{c}\left(\begin{array}{c}\left[\left(1-\Omega_{it+1}^{Tot}\right)e_{mit+1}^{Dir}/\Omega_{it+1}^{Tot}\right]+\\ \left[\left(1-\Omega_{it}^{Tot}\right)e_{mit}^{Dir}/\Omega_{it}^{Tot}\right]\end{array}\right)\left(y_{mit+1}-y_{mit}\right)+\\ \left(\begin{array}{c}\left[\left(1-\Omega_{it+1}^{Tot}\right)e_{mit+1}^{Ind}/\Omega_{it+1}^{Tot}\right]+\\ \left[\left(1-\Omega_{it}^{Tot}\right)e_{mit}^{Ind}/\Omega_{it}^{Tot}\right]\end{array}\right)\left(\sum_{j=1}^{N}y_{mjt+1}-\sum_{j=1}^{N}y_{mjt}\right)\end{array}\right]\ \forall\ i\neq j, \tag{23}$$

where e_{mit}^{Dir}, e_{mit}^{Ind} and e_{mit}^{Tot} are the direct, indirect and total elasticities for the mth output which we obtain directly from Eqs. $14-16$, respectively. In addition, $\left(1-\Omega_{it}^{Tot}\right)/\Omega_{it}^{Tot}$ is the total scale factor where $RTS_{it}^{Tot} = 1/\Omega_{it}^{Tot} = 1/\sum_{m=1}^{M}\sum_{j=1}^{N}e_{mjt}^{Tot}\ \forall\ i,j$ is total returns to scale. See Glass *et al.* (2015) for details on how the presence of a SAR variable that shifts the production technology gives rise to direct, indirect and total returns to scale (referred to in their paper as internal, external and total returns). Intuitively, total returns to scale measure the percentage change in the ith unit's cost following a simultaneous 1% increase in the output(s) of all the units in the sample (the ith unit and all the other $N-1$ units). These total returns to scale are the sum of direct and indirect returns to scale, $RTS_{it}^{Dir} = 1/\Omega_{it}^{Dir} = 1/\sum_{m=1}^{M}e_{mit}^{Dir}$ and $RTS_{it}^{Ind} = 1/\Omega_{it}^{Ind} = 1/\sum_{m=1}^{M}\sum_{j=1}^{N}e_{mjt}^{Ind}$ where for RTS_{it}^{Dir} and RTS_{it}^{Ind} $i\neq j$. Direct returns to scale from a spatial cost function with a SAR environmental variable and returns to scale from a non-spatial cost function are interpreted in the same way and measure the percentage change in the ith unit's cost due to a 1% increase in the ith unit's output(s). Unlike non-spatial returns to scale, however, direct returns to scale also include feedback effects (i.e., when the ith unit's output(s) change, which via the spatial multiplier matrix affect neighbors' costs, some of this effect on neighbors' costs rebounds and affects the cost of the ith unit). Indirect returns to scale from a spatial cost function with a SAR environmental variable measure the percentage change in the ith unit's cost as a result of a simultaneous 1% increase in the output(s) of all the other $N-1$ units. We observe decreasing direct, indirect and total returns if $RTS_{it}^{Dir} < 1$, $RTS_{it}^{Ind} < 1$ and $RTS_{it}^{Tot} < 1$, constant direct, indirect and total returns if $RTS_{it}^{Dir} = 1$, $RTS_{it}^{Ind} = 1$ and $RTS_{it}^{Tot} = 1$, and increasing direct, indirect and total returns if $RTS_{it}^{Dir} > 1$, $RTS_{it}^{Ind} > 1$ and $RTS_{it}^{Tot} > 1$. The classification of RTS_{it}^{Dir}, RTS_{it}^{Ind} and RTS_{it}^{Tot} need not of course be the same.

It is important to note that in Eq. 23 to additively decompose ΔRSE_{it+1}^{Tot} into its direct and indirect components, ΔRSE_{it+1}^{Dir} and ΔRSE_{it+1}^{Ind}, the total scale factor is applied to the output direct and indirect elasticities. This is to weight the contribution of the change in the direct and indirect output elasticities by the change in total returns to scale.

3 Application to U.S. Banks

3.1 Data, Variables and Spatial Weights Matrices

Due to the extent of the heterogeneity in the U.S. banking industry we estimate Eq. 1 separately for large and medium-sized banks. Among other things, this has the benefit of allowing us to examine whether the degree of SAR dependence between medium-sized banks differs from that between large banks. Based on the classification of large and medium-sized U.S. banks in Berger and Roman (2016) we classify a large bank as having total assets greater than \$3 billion in 2015 and a medium-sized bank as having total assets in 2015 between \$1 billion and \$3 billion. For large and medium-sized banks the data samples comprise annual observations for the period $1992 - 2015$, which is a particularly interesting period as it includes the period covering the financial crisis as well sufficiently long pre and post-crisis periods. Furthermore, both our data samples are balanced panels because we analyze continuously operating banks to avoid the impact of entry and exit and to focus on the performance of the core groups of surviving large and medium-sized institutions.

All the data for the variables was either extracted directly from the Reports of Condition and Income (i.e., the Call Reports) of the Federal Reserve System, which we obtain from the Federal Deposit Insurance Corporation (FDIC), or was constructed by the authors using data from this source. Monetary volumes were then deflated to 2005 prices using the consumer price index. After omitting banks where there is a missing bank-year observation for a variable we are left with rich samples of 192 large banks and 299 medium-sized banks.[11] To make appropriate comparisons between large and medium-sized banks we use the same set of variables to estimate the model for each data sample. The outputs and input prices in our SDCF specification are based on the Sealey and Lindley (1977) intermediation approach to banking. We therefore assume, firstly, that banks use the savings of consumers and firms to make investments, which are the outputs, and, secondly, that banks seek to minimize the costs associated with the production of these outputs. In table 1 we describe the variables we use to estimate the models and provide summary statistics for these variables. In short, the three outputs in the models, which reflect the lending and non-lending activities of banks are loans, total securities and total non-interest income ($y_1 - y_3$), and the three input prices relate to the cost of fixed assets,

[11] Using the same approach for small U.S. banks we are left with in excess of $2,900$ small banks. Based on the classification in Berger and Roman (2016) we classify a small U.S. bank as having total assets less than \$1 billion in 2015. We do not, however, estimate the corresponding SDCF for small banks because model estimation time was excessive. To illustrate, even after an entire week of estimating we had still not obtained the results for the non-frontier model for step 1 for the small banks sample.

labor and deposits ($p_1 - p_3$), where p_1 is the normalizing input price. We therefore sum the expenditures on these three inputs to obtain the dependent variable total operating cost (TC), where TC is also normalized by p_1. The remaining 12 variables in table 1 plus a further variable are the z variables that shift the cost frontier technology. This further variable is a financial crisis dummy variable ($2008Dum$) which takes a value of 1 in 2008 and thereafter and 0 before.

We specify a single corresponding $\mathbf{W_N}$ for each of our models for large and medium-sized banks. As we want to account for the entire large and medium-sized bank networks, for each sample the $\mathbf{W_N}$ is based on the inverse distance between the zip codes of the headquarters of each pair of banks. We are therefore taking the view that- the shorter (larger) the distance between banks' headquarters, the greater (smaller) the degree of spatial interaction between the banks, which is entirely reasonable because there is likely to be a greater overlap of banks' branch networks if their headquarters are in close proximity. Since we use distances between the headquarters of all the banks in each sample to calculate the spatial weights, the specifications of $\mathbf{W_N}$ that we use are denoted $\mathbf{W_{All}^{Large}}$ and $\mathbf{W_{All}^{Med}}$. $\mathbf{W_{All}^{Large}}$ and $\mathbf{W_{All}^{Med}}$ therefore avoid omitting any meaningful spatial interaction by applying an arbitrary cut-off to a unit's neighborhood set. Applying a cut-off would involve making an arbitrary assumption about the number of nearest neighbors in a unit's neighborhood set (i.e., three, four, five, etc.) or, alternatively, assuming that all units within a arbitrary distance of one another are neighbors. Also, as we use geographical location rather than financial linkages (e.g., inter-bank lending) to construct $\mathbf{W_{All}^{Large}}$ and $\mathbf{W_{All}^{Med}}$ the spatial weights are exogenous.[12] In regional science and urban economics the analogous $\mathbf{W_N}$ is often used and is constructed using the inverse distances between the centroids of each pair of geographical areas in the sample.

Furthermore, our specifications of $\mathbf{W_N}$ are row-normalized which preserves the scaling of the data. This is because for a particular bank the SAR variable will be a weighted average of the dependent variable observations for all the other banks in the bank's neighborhood set. When a row-normalized inverse distance specification of $\mathbf{W_N}$ is used spillovers are inversely related to the relative distance between the units. Viewing geographical distance as a relative measure is reasonable because it allows the interpretation of distance to vary from bank-to-bank depending on how remote or central the location of a bank's headquarters compared to the headquarters of every other bank in the sample. In addition, a relative interpretation of geographical distance is particularly appropriate for the banking industry because the financial connectedness of banks ensures that no bank's headquarters are isolated within the banking system.

[12] Data on inter-bank lending is not publicly available but if it was possible to access this data or alternatively use some other measure of the financial interconnectedness of banks it would be possible to construct a financial distance based specification of $\mathbf{W_N}$. There would, however, be the added complication of endogenous spatial weights with this type of $\mathbf{W_N}$. To account for endogenous spatial weights our modeling framework for step 1 would need to be adapted.

Table 1: Variable descriptions and summary statistics

Variable description	Model notation	Medium-Sized Banks (299 banks) Mean	Medium-Sized Banks (299 banks) St. Dev.	Large Banks (192 banks) Mean	Large Banks (192 banks) St. Dev.
Operating cost (in 000s 2005 U.S. $): Sum of salaries, interest expenses on deposits and expenditure on fixed assets	TC	27,671	18,657	686,945	2,953,757
Cost of fixed assets: Expenditure on fixed assets divided by the sum of the value of premises and fixed assets	p_1	0.33	0.64	1.06	19.67
Cost of labor: Salaries divided by number of full-time equivalent total employees	p_2	52.80	19.07	58.95	20.09
Cost of deposits: Interest expenses on deposits divided by total deposits	p_3	0.02	0.01	0.02	0.01
Loans: Net loans and leases (in 000s 2005 U.S. $)	y_1	519,122	382,209	13,941,319	59,409,848
Total securities (in 000s 2005 U.S. $)	y_2	193,002	167,684	4,676,564	22,249,192
Total non-interest income (in 000s 2005 U.S. $)	y_3	9,002	17,206	519,222	2,458,490
Return on assets	ROA	1.062	0.840	1.145	1.171
Debt securities: Ratio of debt securities to total securities (y_2)	$DebtSec$	0.967	0.107	0.975	0.075
Loan loss allowance as a share of loans (y_1)	LLA	0.014	0.009	0.016	0.009
Tier 1 capital ratio: Tier 1 capital divided by total assets	CR_1	0.091	0.023	0.086	0.032
Tier 2 capital ratio: Tier 2 capital divided by total assets	CR_2	0.008	0.004	0.011	0.008
Equity ratio: Total equity capital divided by total assets	ER	0.098	0.027	0.100	0.038
Hirschman-Herfindahl Index (HHI) of each bank's asset portfolio across real estate loans, farm loans, commercial and industrial loans, loans to individuals and other loans as ratios of total loans	$Scope$	0.609	0.188	0.574	0.200
Bank asset market share: Each bank's total assets as a share of industry assets	MS	0.008	0.005	0.243	1.082
Security share: Securities (y_2) as a share of of total assets	SEC	0.250	0.137	0.238	0.139
Asset quality: Ratio of non-performing loans to total loans	NPL	0.003	0.009	0.003	0.007
Number of domestic U.S. branches	$Branches$	16	17	165	555
Number of years the institution has been established	Age	81	44	89	48

3.2 Estimated Models

In table 2 we present the fitted $\mathbf{W}^{\mathbf{Med}}_{\mathbf{All}}$ and $\mathbf{W}^{\mathbf{Large}}_{\mathbf{All}}$ SDCF models.[13] The standard interpretation of such models is to recognize that the SAR parameter, δ, is not a spillover elasticity. Spillover elasticities from models that contain the SAR variable are the indirect elasticities, which as is apparent from Eq. 13 depend on, among other things, δ. In table 3 we present the direct, indirect and total elasticities from the fitted $\mathbf{W}^{\mathbf{Med}}_{\mathbf{All}}$ and $\mathbf{W}^{\mathbf{Large}}_{\mathbf{All}}$ SDCF models. An estimate of δ, however, does have an informative interpretation as it represents the degree of SAR dependence across the cross-sectional units. We can see from table 2 that the estimates of δ from the $\mathbf{W}^{\mathbf{Med}}_{\mathbf{All}}$ and $\mathbf{W}^{\mathbf{Large}}_{\mathbf{All}}$ SDCF models are 0.34 and 0.13, respectively, both of which are significant at the 0.1% level. The estimates of δ are interesting as they indicate that there is non-negligible positive SAR dependence across large banks and substantial positive SAR dependence between medium-sized banks, which justifies modeling the cross-sectional SAR dependence in each sample. In particular, a one-sided t-test indicates that the estimate of δ from the $\mathbf{W}^{\mathbf{Med}}_{\mathbf{All}}$ SDCF is significantly larger than that from the $\mathbf{W}^{\mathbf{Large}}_{\mathbf{All}}$ SDCF at the 0.1% level. We suggest that there is a higher degree of SAR dependence across medium-sized banks than there is between large banks because the business of medium-sized banks is more regionally oriented than that of large banks. An interesting issue for further work when computing speed makes it feasible would be to investigate if the estimate of δ for small banks from the corresponding SDCF is greater than we observe here for medium-sized banks because the business of small banks is more localized than the business of their medium-sized counterparts. Furthermore, we can see from table 2 that our preference for a spatial Durbin specification over a SAR specification for medium-sized banks and large banks is supported by a number of significant local spatial parameters at the 5% level or lower in our fitted $\mathbf{W}^{\mathbf{Med}}_{\mathbf{All}}$ and $\mathbf{W}^{\mathbf{Large}}_{\mathbf{All}}$ SDCF models (e.g., the $\mathbf{W}y_1$, $\mathbf{W}\widetilde{p}_2$ and $\mathbf{W}\widetilde{p}_3$ parameters in the $\mathbf{W}^{\mathbf{Med}}_{\mathbf{All}}$ model and the $\mathbf{W}y_3$ and $\mathbf{W}\widetilde{p}_3$ parameters in the $\mathbf{W}^{\mathbf{Large}}_{\mathbf{All}}$ model).

In table 3 we present the direct, indirect and total elasticities from the fitted $\mathbf{W}^{\mathbf{Large}}_{\mathbf{All}}$ and $\mathbf{W}^{\mathbf{Med}}_{\mathbf{All}}$ SDCF models. In line with production theory, the first order direct output and input price parameters from the fitted $\mathbf{W}^{\mathbf{Large}}_{\mathbf{All}}$ and $\mathbf{W}^{\mathbf{Med}}_{\mathbf{All}}$ SDCF models are positive. We can therefore conclude that the fitted $\mathbf{W}^{\mathbf{Large}}_{\mathbf{All}}$ and $\mathbf{W}^{\mathbf{Med}}_{\mathbf{All}}$ SDCF models satisfy the monotonicity property of the translog cost function at the sample mean. All the direct output and input price elasticities at the sample mean are also significant at the 0.1% level. Moreover, we can see from table 3 that there are a number of significant indirect elasticities at the 5% level or lower, which further justifies the spatial Durbin modeling of the costs of medium-sized and large banks in this application.

[13] As is the case with standard non-spatial non-frontier random effects models, the $\log_{10}\vartheta$ parameter in the reported $\mathbf{W}^{\mathbf{Med}}_{\mathbf{All}}$ and $\mathbf{W}^{\mathbf{Large}}_{\mathbf{All}}$ SDCF models is the weight that is attached to the cross-sectional component of the data. In both models this parameter is significant at the 0.1% level.

Table 2: Estimated spatial Durbin stochastic cost frontier models for medium-sized and large U.S. banks

	Medium banks	**Large banks**		**Medium banks**	**Large banks**		**Medium banks**	**Large banks**
y_1	0.335***	0.444***	LLA	0.539**	−0.216	$\mathbf{W}y_2\widetilde{p}_3$	0.022	0.020*
y_2	0.252***	0.178***	CR_1	−0.634***	−0.699***	$\mathbf{W}y_3\widetilde{p}_2$	−0.033	−0.012
y_3	0.156***	0.244***	CR_2	−0.251	−0.192	$\mathbf{W}y_3\widetilde{p}_3$	0.039***	0.024
$\widetilde{p}_2$	0.524***	0.494***	ER	0.394***	0.239	$\mathbf{W}y_1t$	−0.014***	−0.019***
$\widetilde{p}_3$	0.420***	0.431***	$Scope$	−0.003	−0.045	$\mathbf{W}y_2t$	−0.001	−0.002
y_1^2	0.041***	0.047***	MS	10.325***	0.023***	$\mathbf{W}y_3t$	0.005**	0.011***
y_2^2	0.030***	0.007***	SEC	−0.452***	0.069	$\mathbf{W}\widetilde{p}_2t$	−0.012**	−0.003
y_3^2	0.031***	0.041***	NPL	0.896***	1.287***	$\mathbf{W}\widetilde{p}_3t$	0.010***	0.000
y_1y_2	−0.086***	−0.036***	Age	0.018*	−0.027*	$\mathbf{W}ROA$	0.030***	0.017*
y_1y_3	−0.056***	−0.082***	$Branches$	0.259***	0.178***	$\mathbf{W}DebtSec$	0.126	−0.071
y_2y_3	0.005*	0.012***	$2008Dum$	−0.016	−0.046**	$\mathbf{W}LLA$	−1.900*	−1.468
$\widetilde{p}_2^2$	0.017***	0.015*	$Constant$	−0.271	0.169	$\mathbf{W}CR_1$	0.695	0.066
$\widetilde{p}_3^2$	0.016***	0.053***	$\mathbf{W}y_1$	−0.188***	0.019	$\mathbf{W}CR_2$	3.385	−3.982*
$\widetilde{p}_2\widetilde{p}_3$	−0.042***	−0.080***	$\mathbf{W}y_2$	−0.075	−0.045	$\mathbf{W}ER$	−0.058	−0.058
$y_1\widetilde{p}_2$	−0.034***	−0.055***	$\mathbf{W}y_3$	0.013	−0.067***	$\mathbf{W}Scope$	0.038	0.034
$y_1\widetilde{p}_3$	0.039***	0.074***	$\mathbf{W}\widetilde{p}_2$	−0.079**	0.016	$\mathbf{W}MS$	−5.836	−0.025
$y_2\widetilde{p}_2$	0.013***	0.035***	$\mathbf{W}\widetilde{p}_3$	−0.118***	−0.088***	$\mathbf{W}SEC$	0.349	0.092
$y_2\widetilde{p}_3$	0.006*	0.003	$\mathbf{W}y_1^2$	−0.004	0.002	$\mathbf{W}NPL$	0.864	5.325***
$y_3\widetilde{p}_2$	0.044***	0.014*	$\mathbf{W}y_2^2$	−0.016*	−0.005	$\mathbf{W}Age$	−0.137***	−0.136***
$y_3\widetilde{p}_3$	−0.049***	−0.060***	$\mathbf{W}y_3^2$	−0.019*	−0.035***	$\mathbf{W}Branches$	−0.001	−0.049*
t	−0.002	−0.002	$\mathbf{W}y_1y_2$	0.018	0.025	δ	0.335***	0.133***
t^2	0.001***	0.001***	$\mathbf{W}y_1y_3$	0.017	0.035	$\log_{10}\vartheta$	−1.684***	−1.440***
y_1t	0.002*	0.005***	$\mathbf{W}y_2y_3$	0.004	−0.001	LL	5877.68	2401.91
y_2t	−0.002***	−0.001	$\mathbf{W}\widetilde{p}_2^2$	0.042	0.005	σ_v	0.44 (0.01)	0.80 (0.01)
y_3t	0.000	−0.001	$\mathbf{W}\widetilde{p}_3^2$	0.016*	0.000	σ_u	0.84 (0.02)	0.00 (0.15)
$\widetilde{p}_2t$	0.008***	0.006**	$\mathbf{W}\widetilde{p}_2\widetilde{p}_3$	−0.052*	0.016	σ_κ	0.09 (0.00)	0.10 (0.00)
$\widetilde{p}_3t$	−0.012***	−0.003	$\mathbf{W}y_1\widetilde{p}_2$	0.104***	0.081**	σ_η	0.15 (0.00)	0.15 (0.01)
ROA	−0.032***	−0.039***	$\mathbf{W}y_1\widetilde{p}_3$	−0.062**	−0.048**			
$DebtSec$	0.010	−0.127***	$\mathbf{W}y_2\widetilde{p}_2$	−0.015	−0.030*			

*, ** and *** denote statistical significance at the 5%, 1% and 0.1% levels, respectively.
Standard errors are in parentheses.

Table 3: Direct, indirect and total elasticities for medium-sized and large U.S. banks

	Direct		**Indirect**		**Total**	
	Medium banks	Large banks	Medium banks	Large banks	Medium banks	Large banks
y_1	0.334***	0.445***	−0.111	0.090	0.223**	0.535***
y_2	0.253***	0.179***	−0.016	−0.034	0.237*	0.145*
y_3	0.157***	0.244***	0.075**	−0.043*	0.232***	0.200***
$\widetilde{p}_2$	0.526***	0.495***	0.146***	0.098***	0.672***	0.593***
$\widetilde{p}_3$	0.421***	0.432***	0.033**	−0.037**	0.454***	0.395***
y_1^2	0.042***	0.048***	0.012	0.009	0.054**	0.057***
y_2^2	0.030***	0.007***	−0.009	−0.005	0.022	0.002
y_3^2	0.031***	0.041***	−0.012	−0.033***	0.019	0.008
y_1y_2	−0.087***	−0.036***	−0.014	0.025	−0.101***	−0.011
y_1y_3	−0.056***	−0.082***	0.000	0.029	−0.056*	−0.053**
y_2y_3	0.006*	0.013***	0.006	−0.001	0.012	0.012
$\widetilde{p}_2^2$	0.018***	0.015*	0.065	0.006	0.082*	0.021
$\widetilde{p}_3^2$	0.016***	0.053***	0.031***	0.008	0.047***	0.060***
$\widetilde{p}_2\widetilde{p}_3$	−0.043***	−0.080***	−0.094**	0.010	−0.137***	−0.070**
$y_1\widetilde{p}_2$	−0.032***	−0.054***	0.138***	0.084**	0.106*	0.030
$y_1\widetilde{p}_3$	0.037***	0.073***	−0.071*	−0.043*	−0.034	0.029
$y_2\widetilde{p}_2$	0.013**	0.035***	−0.016	−0.029	−0.003	0.006
$y_2\widetilde{p}_3$	0.007*	0.004	0.033	0.022*	0.040*	0.026*
$y_3\widetilde{p}_2$	0.044***	0.013*	−0.023	−0.012	0.021	0.000
$y_3\widetilde{p}_3$	−0.049***	−0.059***	0.034*	0.018	−0.015	−0.041*

	Direct		**Indirect**		**Total**	
	Medium banks	Large banks	Medium banks	Large banks	Medium banks	Large banks
t	−0.002	−0.002	−0.001	0.000	−0.002	−0.002
t^2	0.001***	0.001***	0.001***	0.000*	0.002***	0.001***
y_1t	0.001	0.005***	−0.020***	−0.022***	−0.018***	−0.017***
y_2t	−0.002***	−0.001	−0.002	−0.002	−0.004	−0.003
y_3t	0.000	−0.001	0.008**	0.012***	0.007*	0.012***
$\widetilde{p}_2t$	0.008***	0.006***	−0.014*	−0.003	−0.006	0.003
$\widetilde{p}_3t$	−0.012***	−0.003*	0.009*	0.000	−0.003	−0.002
ROA	−0.032***	−0.039***	0.026	0.012	−0.006	−0.028*
$DebtSec$	0.011	−0.128***	0.178	−0.104	0.189	−0.232
LLA	0.532**	−0.183	−2.604*	−1.795	−2.072	−1.978
CR_1	−0.613***	−0.684***	0.692	−0.085	0.079	−0.769
CR_2	−0.217	−0.270	4.598	−4.765*	4.381	−5.035*
ER	0.375**	0.218	0.130	−0.024	0.505	0.194
$Scope$	−0.002	−0.044	0.044	0.024	0.042	−0.020
MS	10.114***	0.022***	−3.243	−0.026	6.871	−0.004
SEC	−0.456***	0.061	0.341	0.146	−0.114	0.207
NPL	0.903***	1.304**	1.775	6.291***	2.678*	7.596***
Age	0.016*	−0.028*	−0.198***	−0.157***	−0.182***	−0.185***
$Branches$	0.261***	0.178***	0.131***	−0.026	0.392***	0.151***
$2008Dum$	−0.017	−0.048**	−0.008	−0.007*	−0.026	−0.055**

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

In particular, there are number of indirect output and input price spillover elasticities which are significant at the 5% level or lower from our fitted $\mathbf{W}_{\mathbf{All}}^{\mathbf{Med}}$ and $\mathbf{W}_{\mathbf{All}}^{\mathbf{Large}}$ SDCF models (i.e., those elasticities pertaining to y_3, $\widetilde{p}_2$ and $\widetilde{p}_3$ for both models). Whereas theory indicates what the sign of the direct output and input price elasticities should be, there is no such prescription for indirect output and input price elasticities. From the $\mathbf{W}_{\mathbf{All}}^{\mathbf{Med}}$ SDCF all the significant indirect output and input price elasticities are positive, whereas from the $\mathbf{W}_{\mathbf{All}}^{\mathbf{Large}}$ SDCF the indirect $\widetilde{p}_2$ elasticity is positive and the indirect y_3 and $\widetilde{p}_3$ elasticities are negative. These negative indirect elasticities are because the negative local spatial y_3 and $\widetilde{p}_3$ parameters more than offset the positive SAR coefficient. Despite these negative indirect elasticities and some of the indirect output elasticities not being significant in both models, we find for our samples of medium-sized and large banks that the direct output and input price elasticities dominate in the calculation of the corresponding total elasticities. This is evident because for both models all the total output and input price elasticities are positive and significant at the 5% level or lower.

Turning now to briefly discuss the direct first order time parameters. For both models we expect this parameter to be negative to signify that the spatial cost frontier shifts down annually for the sample average bank due to technical progress. In line with our expectations the direct first order time parameters from the $\mathbf{W}_{\mathbf{All}}^{\mathbf{Med}}$ and $\mathbf{W}_{\mathbf{All}}^{\mathbf{Large}}$ SDCF models are negative but neither are significant. Glass *et al.* (2017) find from a non-spatial random coefficients analysis of U.S. banks that the first order time trend parameter from their fitted input distance function for the $1992-2007$ pre-crisis period is, as we would expect, positive and significant. From duality theory this is equivalent to a significant negative first order time coefficient from the corresponding cost function. In contrast, they report a significant negative first order time trend parameter from the same specification of the input distance function for the $2008-2015$ period. An overriding feature of the $2008-2015$ period is a deepening of the financial crisis, which is the reason they provide for their negative first order time parameter. This is also the reason we give to explain why our negative direct first order time parameters are not significant.

The widespread differences in table 3 between the corresponding elasticities from the two models highlights the heterogeneity between our samples of medium-sized and large banks. Such widespread differences between the corresponding elasticities supports not pooling the two samples. Nowhere are these differences in the corresponding elasticities more apparent than for the z variables. We shed further light on this and conclude our analysis of the fitted models by discussing the key differences between the corresponding direct elasticities for the z variables. Of the 13 corresponding direct elasticities for the z variables for medium-sized and large banks we can see from table 3 that four have different signs where both are significant or just one is ($DebtSec$, LLA, SEC and Age), and a further three have the same sign but differ substantially in magnitude (ER, MS and NPL). In particular, we note that there is a huge difference between the magnitude

Table 4: Summary own net and gross cost efficiencies for medium-sized and large U.S. banks

	Medium-Sized Banks			Large Banks		
	NVE	NIE	GVE	NVE	NIE	GVE
4th quartile	0.904	0.975	0.864	0.998	0.966	0.964
3rd quartile	0.648	0.928	0.577	0.998	0.922	0.920
2nd quartile	0.544	0.903	0.483	0.998	0.893	0.892
1st quartile	0.412	0.868	0.366	0.998	0.861	0.859
Average	**0.524**	**0.892**	**0.466**	**0.998**	**0.883**	**0.881**

of the significant positive direct MS elasticities because for medium-sized banks this marginal effect is very large and for large banks it is very small. This suggests that, on average, there are implications for the scale of a medium-sized bank's operations following a marginal change in its market share, whereas an incremental change in the market share of a large bank has no such implications. Finally, we note that the direct dummy variable parameter pertaining to the financial crisis is negative for both medium-sized and large banks but only significant for the latter. The sign of these direct $2008Dum$ parameters is as we would expect because following the financial crisis interest rates went down which reduced banks' deposit account expenses.

3.3 Own Net and Own Gross Cost Efficiencies

In table 4 we summarize the own NIE, NVE and GVE from the fitted structural form of the $\mathbf{W}^{\mathbf{Med}}_{\mathbf{All}}$ and $\mathbf{W}^{\mathbf{Large}}_{\mathbf{All}}$ SDCF models in Eq. 1. Although the fitted structural form of the models accounts for global SAR and local spatial dependencies, the own efficiencies do not include any form of efficiency spillover. This is in contrast to the direct, indirect and total efficiencies that we present and discuss in the next subsection, which to different degrees all include efficiency spillovers. We can see from table 4 that the sample average NIE for medium-sized and large banks are similar (0.892 and 0.883, respectively), whereas the sample average NVE for large banks (0.998) is much larger than that for medium-sized banks (0.524). This indicates in terms of own cost efficiency that our samples of large and medium-sized banks are very different because, on average, the biggest source of underperformance for large banks is quite a small amount of time-invariant inefficiency, whereas for medium-sized banks it is a large amount of time-variant inefficiency. One-tailed tests for the presence of η and u provide support for the magnitude of the sample average estimates of NIE and NVE because at nominal levels of significance we reject $\widehat{\sigma}^2_\eta = 0$ and $\widehat{\sigma}^2_u = 0$ for medium-sized banks and for large banks we reject $\widehat{\sigma}^2_\eta = 0$ but not $\widehat{\sigma}^2_u = 0$. Multiplying the sample average NIE and NVE yields sample average GVE estimates for medium-sized and large banks of 0.466 and 0.881, respectively. Finally in this discussion of the own efficiencies, we observe that the NVE and NIE distributions for medium-sized banks and the NIE distribution for large banks are negatively skewed because from table 4 the 2nd quartile average is greater than the sample average.[14]

3.4 Direct, Indirect and Total Gross Cost Efficiencies

We confine our discussion of the direct, indirect and total cost efficiencies to GVE^{Dir}, GVE_{To}^{Ind} and GVE_{To}^{Tot} because these efficiencies provide a more complete picture of economic performance than the corresponding NIE and NVE. In table 5 we summarize our estimates of GVE^{Dir}, GVE_{To}^{Ind} and GVE_{To}^{Tot} from the reduced form of the fitted $\mathbf{W_{All}^{Med}}$ and $\mathbf{W_{All}^{Large}}$ SDCF models for the sample and by quartile. The sample average GVE^{Dir} for medium-sized and large banks are 0.468 and 0.883, respectively. These direct efficiencies are the sample average own GVE plus the efficiency feedback, which is a particular form of efficiency spillover. Specifically, this feedback is the component of a unit's direct efficiency which via the spatial multiplier matrix, $(\mathbf{I_N} - \delta\mathbf{W_N})^{-1}$, passes through a unit's 1st order and higher order neighbors and rebounds back to the unit. For large banks the magnitude of the sample average GVE^{Dir} is exactly that of the sample average own GVE (see table 4), which suggests there is zero efficiency feedback. In contrast, for medium-sized banks the sample average GVE^{Dir} is of the order of 0.056 smaller than the sample average own GVE, which indicates in this case that there is negative efficiency feedback. Also, we can see from table 5 that the sample average GVE_{To}^{Ind} for medium-sized banks is 0.231 and for large banks it is 0.134. This suggests that the indirect efficiency spillovers for medium-sized banks are substantial and for large banks they are non-negligible, which provides support for our spatial efficiency methodology. Having summed the sample average GVE^{Dir} and GVE_{To}^{Ind}, the sample average GVE_{To}^{Tot} for medium-sized banks is 0.699 and for large banks it is 1.016. It is therefore apparent for large banks that the indirect efficiency spillovers are sufficiently large to push the sample average GVE_{To}^{Tot} slightly above the own GVE benchmark of 1. With regard to the average efficiencies by quartile in table 5 we highlight two features. Firstly, we can see that the 4th quartile average GVE_{To}^{Tot} for medium-sized and large banks are both well above the own GVE benchmark. Secondly, the difference between the 4th quartile average GVE_{To}^{Ind} for medium-sized and large banks is twice the corresponding difference for the 1st quartile. Finally, from the kernel densities of the GVE_{To}^{Tot} scores for medium-sized and large banks in figure 1 it is evident that the distribution of the GVE_{To}^{Tot} scores for large banks is positively skewed and not as smooth and less dispersed than we observe for medium-sized banks.

[14] Unlike for the whole sample, for the quartiles average GVE is not the product of average NIE and average NVE. This is because a bank will not necessarily be in the same quartile for NIE and NVE.

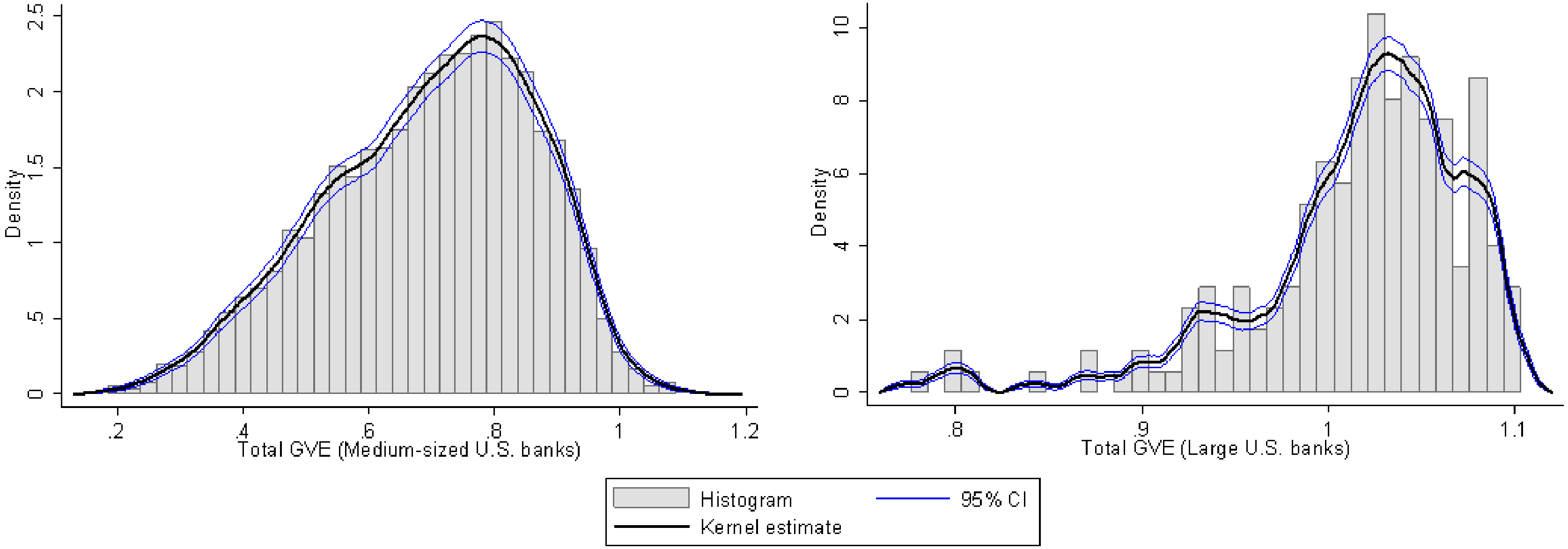


Figure 1: Kernel densiities of total gross cost efficiencies for medium-sized and large U.S. banks

Table 5: Summary direct, indirect and total gross cost efficiencies for medium-sized and large U.S. banks

	Medium-Sized Banks			Large Banks		
	Direct GVE	Indirect GVE	Total GVE	Direct GVE	Indirect GVE	Total GVE
4th quartile	0.870	0.307	1.138	0.964	0.143	1.103
3rd quartile	0.578	0.263	0.826	0.923	0.136	1.055
2nd quartile	0.486	0.234	0.719	0.892	0.135	1.027
1st quartile	0.368	0.205	0.582	0.860	0.133	0.994
Average	**0.468**	**0.231**	**0.699**	**0.883**	**0.134**	**1.016**

3.5 Spatial TFP Growth Decompositions

In figures 2 and 3 we present the average annual spatial TFP growth decompositions for large and medium-sized U.S. banks over the period $1992 - 2015$. In these figures the spatial TFP growth decompositions are expressed as indices as we exponentiate the growth rates. In particular, there are two aspects to each of these figures. The first is the decomposition of ΔTFP^{Tot} into ΔTFP^{Dir} and ΔTFP^{Ind} (panel A). The second is the decomposition of the direct, indirect and total TFP growth rates into direct, indirect and total TA, GVI, AE and RSE growth rates (panels B-D).

For large banks we can see from panel A in figure 2 that ΔTFP^{Tot} increased sharply since the financial crisis in 2008 due to a sharp increase in ΔTFP^{Dir}. Prior to the financial crisis there was a slower growing upward trend in ΔTFP^{Dir} for large banks and over the entire study period there is a steady declining trend in their ΔTFP^{Ind} with negative changes in the last few years of the sample. We can therefore infer from these results that, on average, over the study period a large bank's ΔTFP^{Tot} has become more dependent on the bank itself (i.e., ΔTFP^{Dir}) and less dependent on spatial spillovers (i.e., ΔTFP^{Ind}).

For large banks we identify from panels B-D in figure 2 three salient features of the components of ΔTFP^{Dir}, ΔTFP^{Ind} and ΔTFP^{Tot}. Firstly, we observe that over the study period a key driver of ΔTFP^{Dir} is ΔTA^{Dir} and a key driver of ΔTFP^{Ind} is ΔTA^{Ind}. This is because ΔTA^{Dir} consistently increases with progressively smaller negative changes in the first portion of our sample and ΔTA^{Ind} consistently decreases with progressively larger negative changes from $2005 - 06$ onwards. Although negative SAR dependence or negative spatial error autocorrelation are not frequently observed phenomena on the occasions where either is observed it is attributed to the effects of competition between neighbors (Kao and Bera, 2013). On this basis the declining positive ΔTA^{Ind} in the first portion of the sample suggests that there has been such a decline in technical diffusion across space that in the second portion of the sample there is increasing geographical technical competition owing to the increasing negative ΔTA^{Ind}.

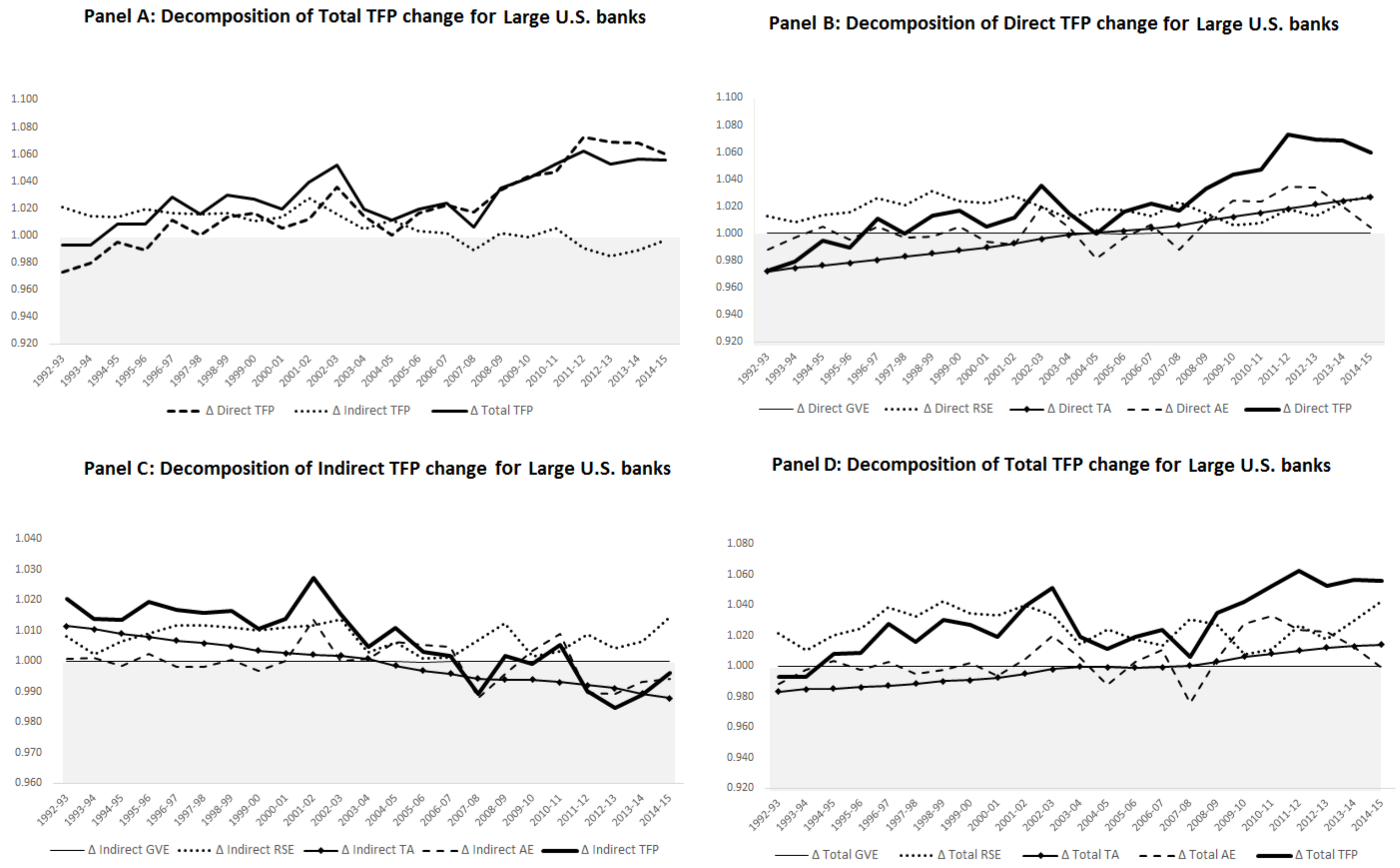


Figure 2: Spatial TFP growth decompositions for large U.S. banks

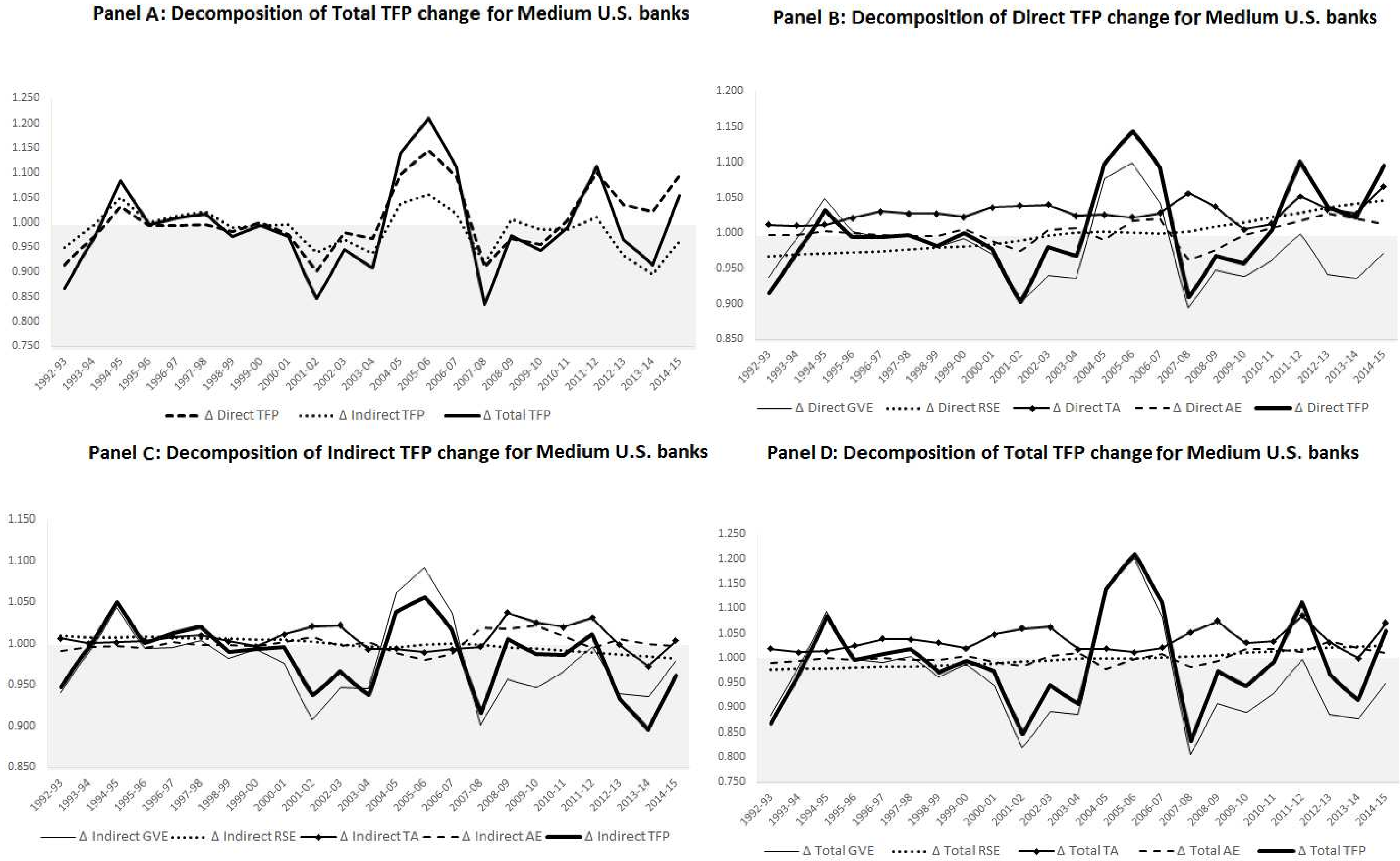


Figure 3: Spatial TFP growth decompositions for medium-sized U.S. banks

Secondly, ΔTA^{Tot} has an upward trend which reveals that ΔTA^{Dir} dominates ΔTA^{Ind} in the calculation of ΔTA^{Tot}. Thirdly, we can see that over the study period ΔTFP^{Dir} and ΔTFP^{Ind} closely track ΔAE^{Dir} and ΔAE^{Ind} and to a lesser extent ΔTFP^{Tot} tracks ΔAE^{Tot}. Taken together these findings emphasize the importance of direct allocative efficiency and indirect allocative efficiency spillovers for large U.S. banks. Also, in terms of direct allocative efficiency we can see that large banks reacted appropriately to the financial crisis as the crisis marked the beginning of a period of year-on-year increases in ΔAE^{Dir}.

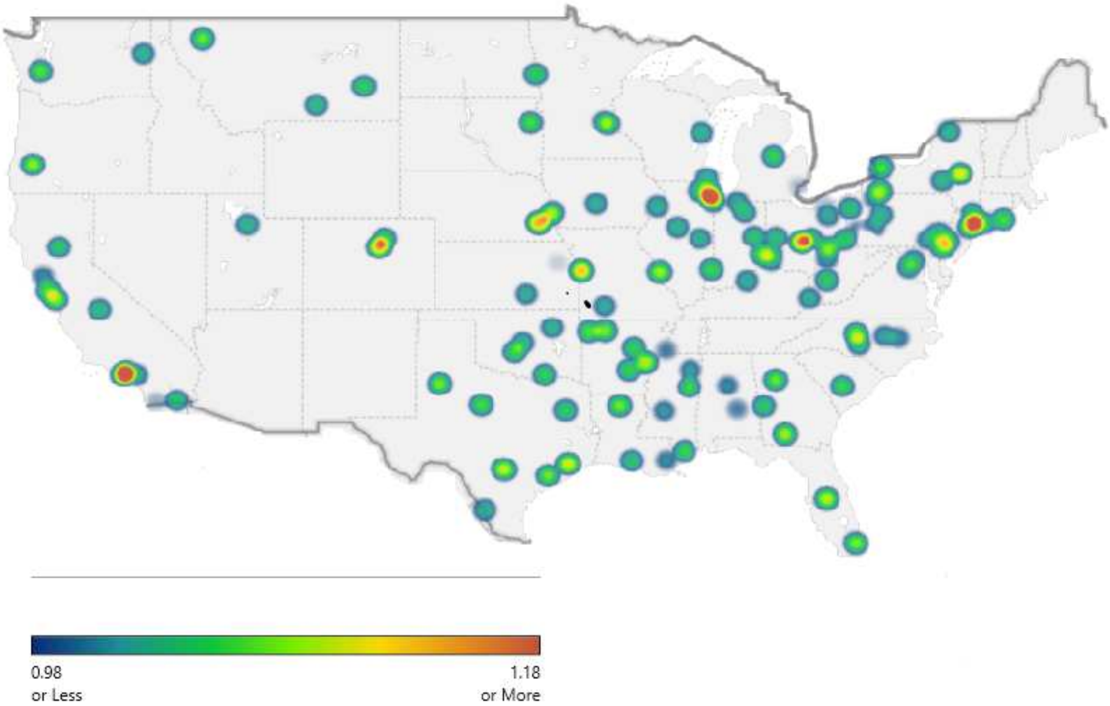
Panel A of figure 3 indicates the absence of trends in ΔTFP^{Dir}, ΔTFP^{Ind} and ΔTFP^{Tot} for medium-sized banks which is in stark contrast to what we observed above for large banks. It is evident, however, for medium-sized banks that ΔTFP^{Tot} closely tracks ΔTFP^{Dir} and ΔTFP^{Ind}, whereas for large banks we found that ΔTFP^{Tot} closely tracks ΔTFP^{Dir} but not ΔTFP^{Ind}. From panels B-D of figure 3 a noticeable feature of the components of ΔTFP^{Dir}, ΔTFP^{Ind} and ΔTFP^{Tot} for medium-sized banks is that they closely track ΔGVE^{Dir}, ΔGVE^{Ind} and ΔGVE^{Tot}, respectively. In contrast, as we pointed out above for large banks this type of relationship was between, in particular, the changes in direct and indirect TFP and the corresponding changes in allocative efficiency.

In figure 4 we present the geographical distributions of the average total TFP indices over the period $1992 - 2015$ for large banks in panel A and medium-sized banks in panel B. Interestingly there are some clear similarities between these geographical distributions for large and medium-sized banks and also some clear differences. The similarities that we highlight are the clusters of large and medium-sized banks with high average total TFP indices in the Chicago and Kansas areas and in two nearby but distinct areas in the vicinity of, as is expected, New York City. The differences relate to metropolitan areas where there are clusters of high average total TFP banks which are all large (Los Angeles; Denver; Omaha; Lincoln; and Columbus) or all medium-sized (e.g., San Antonio, Austin and Dallas-Fort Worth).

4 Concluding Remarks

In this paper we set out the methodology for a new spatial TFP growth decomposition. We also demonstrate the steps involved in the practical implementation of our decomposition via an empirical application. As a result our paper makes a substantial contribution to the literature on the spatial decomposition of TFP growth because although, as we highlight, there is a well-developed non-spatial literature on the decomposition of TFP growth and an evolving literature on spatial stochastic frontier modeling, there is just one short study by Glass *et al.* (2013) (denoted GKPF throughout this paper) on the

Panel A: Average TFP change for large U.S. banks



Panel B: Average TFP change for medium-sized U.S. banks

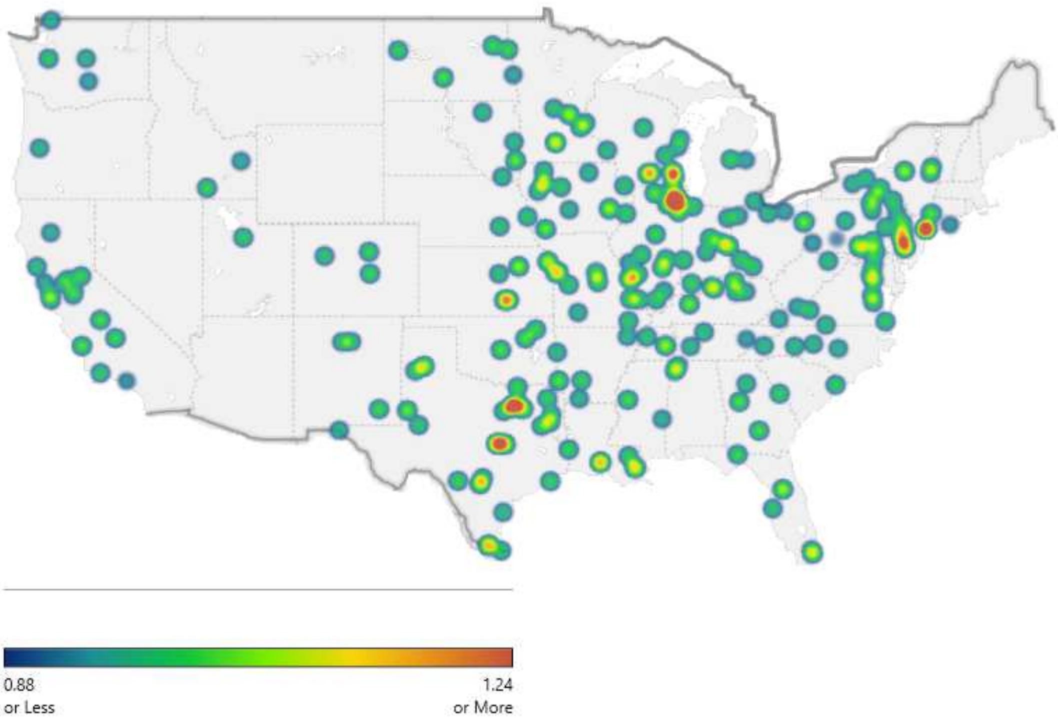


Figure 4: Geographical distribution of average TFP change over the period 1992-2015 for large and medium-sized U.S. banks

decomposition of TFP growth in the presence of spillovers. Specifically, our paper has extended GKPF in four varied ways, the first two of which are methodological. Firstly, we introduce a cost efficiency spillover change component to the GKPF decomposition. Secondly, we augment the GKPF decomposition with own and spillover allocative efficiency changes. Thirdly, our paper provides a much more detailed coverage of the spatial decomposition of TFP growth than the short communication in GKPF. Fourthly, whereas the empirical application in GKPF is traditional as they apply their decomposition using data for geographical areas (cities, regions, etc.), we apply our decomposition using firm level data, which suggests that spatial efficiency and productivity analysis can have an important future role in OR. With regard to our empirical application which focused on U.S. banks over the period $1992 - 2015$, among other things, for large banks we observe a steady decline in technical diffusion across space in the first half of the sample and increasing geographical technical competition in the second half of the study period.

References

Andrews, D. W. K. (2001): 'Testing when a parameter is on the boundary of the maintained hypothesis'. *Econometrica*, vol. 69, pp. 683–734.

Badunenko, O. and S. C. Kumbhakar (2016): 'When, where and how to estimate persistent and transient efficiency in stochastic frontier panel data models'. *European Journal of Operational Research*, vol. 255, pp. 272-287.

Badunenko, O. and S. C. Kumbhakar (2017): 'Economies of scale, technical change and persistent and time-varying cost efficiency in Indian banking: Do ownership, regulation and heterogeneity matter?'. *European Journal of Operational Research*, vol. 260, pp. 789-803.

Bauer, P. W. (1990): 'Decomposing TFP growth in the presence of cost inefficiency, nonconstant returns to scale, and technological progress'. *Journal of Productivity Analysis*, vol. 1, pp. 287-299.

Berger, A. N. and R. A. Roman (2016): 'Did saving Wall Street really save Main Street? The real effects of TARP on local economic conditions'. Forthcoming in the *Journal of Financial and Quantitative Analysis.* Available at SSRN: https://ssrn.com/abstract=2442070 or http://dx.doi.org/10.2139/ssrn.2442070

Bos, J. W. B., M. Koetter, J. W. Kolari and C. J. M. Kool (2009): 'Effects of heterogeneity on bank efficiency scores'. *European Journal of Operational Research*, vol. 195, pp. 251-261.

Christensen, L. R., D. W. Jorgenson and L. J. Lau (1973): 'Transcendental logarithmic production frontiers'. *Review of Economics and Statistics*, vol. 55, pp. 28-45.

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

Colombi, R., S. C. Kumbhakar, G. Martini and G. Vittadini (2014): 'Closed-skew normality in stochastic frontiers with individual effects and long/short run efficiency'. *Journal of Productivity Analysis*, vol. 42, pp. 123-136.

Cornwell, C., P. Schmidt and R. C. Sickles (1990): 'Production frontiers with cross-sectional and time-series variation in efficiency levels'. *Journal of Econometrics,* vol. 46, pp. 185-200.

Diewert, W. E. and K. J. Fox (2017): 'Decomposing productivity indexes into explanatory factors'. *European Journal of Operational Research*, vol. 256, pp. 275-291.

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.

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.

Filippini, M. and W. Greene (2016): 'Persistent and transient productive inefficiency: A maximum simulated likelihood approach'. *Journal of Productivity Analysis,* vol. 45, pp. 187-196.

Glass, A. J., A. Kenjegaliev and K. Kenjegalieva (2017): 'Influence of a global shock on an organization's input substitution: The case of the financial crisis and U.S. banks. Mimeo.

Glass, A. J., K. Kenjegalieva and J. Paez-Farrell (2013): 'Productivity growth decomposition using a spatial autoregressive frontier model'. *Economics Letters*, vol. 119, pp. 291-295.

Glass, A. J., K. Kenjegalieva and R. C. Sickles (2015): 'Returns to scale and curvature in the presence of spillovers: Evidence from European countries'. *Oxford Economic Papers*, vol. 68, pp. 40-63.

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

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

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

Greene, W. H. (2004): 'Distinguishing between heterogeneity and inefficiency: Stochastic frontier analysis of the World Health Organization's panel data on national health care systems'. *Health Economics*, vol. 13, pp. 959-980.

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

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.

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

Kelejian, H. H. and I. R. Prucha (1999): 'A generalized moments estimator for the autoregressive parameter in a spatial model '. *International Economic Review*, vol. 40, pp. 509-533.

Kumbhakar, S. C., G. Lien and J. B. Hardaker (2014): 'Technical efficiency in competing panel data models: A study of Norwegian grain farming'. *Journal of Productivity Analysis*, vol. 41, pp. 321-327.

Mukherjee, K., S. C. Ray and S. M. Miller (2001): 'Productivity growth in large US commercial banks: The initial post-deregulation experience'. *Journal of Banking and Finance*, vol. 25, pp. 913-939.

O'Donnell, C. J. (2016): 'Using information about technologies, markets and firm behaviour to decompose a proper productivity index'. *Journal of Econometrics*, vol. 190, pp. 328-340.

Orea, L. (2002): 'Parametric decomposition of a generalized Malmquist productivity index'. *Journal of Productivity Analysis*, vol. 18, pp. 5-22.

Orea, L., I. C. Álvarez and T. Jamasb (2016): 'Using a spatial econometric approach to mitigate omitted variables in stochastic frontier models: An application to Norwegian electricity distribution networks'. University of Cambridge Faculty of Economics Working Paper 1673.

Oude Lansink, A., S. Stefanou and T. Serra (2015): 'Primal and dual dynamic Luenberger productivity indicators'. *European Journal of Operational Research*, vol. 241, pp. 555-563.

Schmidt, P. and R. C. Sickles (1984): 'Production frontiers and panel data'. *Journal of Business and Economic Statistics*, vol. 2, pp. 367-374.

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

Sun, K., S. C. Kumbhakar and R. Tveras (2015): 'Productivity and efficiency estimation: A semiparametric stochastic cost frontier approach'. *European Journal of Operational Research*, vol. 245, pp. 194-202.

Tsionas, E. G. and S. C. Kumbhakar (2014): 'Firm heterogeneity, persistent and transient technical inefficiency: A generalized true random-effects model'. *Journal of Applied Econometrics*, vol. 29, pp. 110-132.

Tsionas, E. G. and P.G. Michaelides (2016): 'A spatial stochastic frontier model with spillovers: evidence for Italian regions'. *Scottish Journal of Political Economy*, vol. 63, pp. 243-257.