

NONPARAMETRIC, STOCHASTIC FRONTIER MODELS WITH MULTIPLE INPUTS AND OUTPUTS

Léopold Simar, Paul W. Wilson

REPRINT | 2022 / 40

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

<https://uclouvain.be/en/research-institutes/lidam/isba/publication.html>

Nonparametric, Stochastic Frontier Models with Multiple Inputs and Outputs

Léopold Simar & Paul W. Wilson

To cite this article: Léopold Simar & Paul W. Wilson (2022): Nonparametric, Stochastic Frontier Models with Multiple Inputs and Outputs, Journal of Business & Economic Statistics, DOI: [10.1080/07350015.2022.2110882](https://doi.org/10.1080/07350015.2022.2110882)

To link to this article: <https://doi.org/10.1080/07350015.2022.2110882>



View supplementary material [↗](#)



Published online: 03 Oct 2022.



Submit your article to this journal [↗](#)



Article views: 158



View related articles [↗](#)



View Crossmark data [↗](#)



Nonparametric, Stochastic Frontier Models with Multiple Inputs and Outputs

Léopold Simar^a  and Paul W. Wilson^{b,c} 

^aInstitut de Statistique, Biostatistique, et Sciences Actuarielles, Université Catholique de Louvain-la-Neuve, Louvain-la-Neuve, Belgium; ^bDepartment of Economics and School of Computing, Clemson University, Clemson, SC; ^cFederal Reserve Bank of St. Louis, St. Louis, MO

ABSTRACT

Stochastic frontier models along the lines of Aigner et al. are widely used to benchmark firms' performances in terms of efficiency. The models are typically fully parametric, with functional form specifications for the frontier as well as both the noise and the inefficiency processes. Studies such as Kumbhakar et al. have attempted to relax some of the restrictions in parametric models, but so far all such approaches are limited to a univariate response variable. Some (e.g., Simar and Zelenyuk; Kuosmanen and Johnson) have proposed nonparametric estimation of directional distance functions to handle multiple inputs and outputs, raising issues of endogeneity that are either ignored or addressed by imposing restrictive and implausible assumptions. This article extends nonparametric methods developed by Simar et al. and Hafner et al. to allow multiple inputs and outputs in an almost fully nonparametric framework while avoiding endogeneity problems. We discuss properties of the resulting estimators, and examine their finite-sample performance through Monte Carlo experiments. Practical implementation of the method is illustrated using data on U.S. commercial banks.

ARTICLE HISTORY

Received August 2021
Accepted July 2022

KEYWORDS

Efficiency; Nonparametric;
Stochastic frontier

1. Introduction and Background

Stochastic frontier models are widely used to analyze the productivity and efficiency of firms, and are attractive because they allow for noise in the data generating process (DGP). The models are typically fully parametric, with functional form specifications for the frontier as well as both the noise and the inefficiency processes along the lines of the composite-error models of Aigner et al. (1977), Battese and Corra (1977), and Meeusen and van den Broeck (1977). The generic model in this framework is

$$Y = \varphi(X) + \epsilon - \eta, \quad (1.1)$$

where Y is an output quantity, $X \in \mathbb{R}_+^P$ is a vector of input quantities, $\varphi(X)$ is the production frontier, $\eta \geq 0$ is stochastic inefficiency and $\epsilon \in \mathbb{R}$ is stochastic noise with $\mathbb{E}(\epsilon) = 0$. Conditionally on X , η and ϵ are independent and in some studies, both η and ϵ are allowed to be heteroscedastic. Note also that X could be replaced by (X, E) , with E denoting environmental factors that may influence the production process. In parametric applications, $\varphi(X)$ is almost always specified as either Cobb-Douglas or Translog. Coelli (1995) and Kumbhakar et al. (2020) observe that most applied articles specify a normal distribution for the two-sided noise term and a half-normal distribution for the one-sided inefficiency term, that is, $\epsilon \sim N(0, \sigma_\epsilon^2)$ and $\eta \sim N^+(0, \sigma_\eta^2)$.

This article presents a highly flexible approach for estimating production efficiency in the presence of multiple inputs and multiple outputs. We propose a flexible, almost fully nonpara-

metric model based on the directional distance measures of firms' efficiencies developed by Chambers et al. (1996, 1998), but we allow for noise. We then transform the model to facilitate estimation using the ideas of Simar et al. (2017) (hereafter, SVKZ) and Hafner et al. (2018) (hereafter, HMS) and to avoid endogeneity issues. As shown in the separate Appendix E.1, supplementary materials, our method can also accommodate radial, as opposed to directional, distances via a simple modification of the transformation we use below. The frontier is allowed to be fully nonparametric, and we impose only symmetry on the density of the noise term. We specify a specific functional form for the density of the inefficiency term, but we estimate its parameters locally. We show that estimates of marginal rates of substitution and transformation, as well as other features of the original, untransformed model, can be easily recovered from estimates of the transformed model. An empirical illustration is provided by estimating the efficiency of U.S. commercial banks before, during and after the late-2007 financial crisis.

Ours is not the first attempt to move away from the fully parametric version of (1.1) and the associated restrictive and unrealistic assumptions that are often rejected when tested; see Parmeter and Zelenyuk (2019), Sickles and Zelenyuk (2019), or Kumbhakar et al. (2020) for discussion. A fully nonparametric approach raises identification problems as discussed by Hall and Simar (2002), in addition to possible problems for inference. However, we go farther toward a nonparametric model than other proposed methods, while avoiding a number of problems in other proposals. Fan et al. (1996) and Kuosmanen

and Kortelainen (2012) allow the frontier function $\varphi(X)$ to be nonparametric, but the stochastic parts of their models are still parametric and homoscedastic. The model of Hall and Simar (2002) is fully nonparametric with symmetric noise, but asymptotic identification of the model requires the unrealistic, restrictive assumption that the variance of the noise ϵ goes to zero as the sample size n tends to infinity. Kumbhakar et al. (2007), Simar and Zelenyuk (2011), and Park et al. (2015) use nonparametric local maximum likelihood approaches. In these approaches, both the model for $\varphi(X)$ and the structure of the parameters of the local densities of η and ϵ are functional (nonparametric) and approximated by local polynomials, allowing unspecified heteroscedasticity for both stochastic components. Although this is attractive and flexible, estimation is difficult and computationally expensive.

Recently, SVKZ suggest a flexible model based also on local polynomials, but requiring only local least-squares methods for estimation. This model also allows for heteroscedasticity in both η and ϵ in a functional way, but is much faster to compute and is more robust since estimation is based on moment estimators and the full likelihood is not needed. In particular, only symmetry is required for the density of ϵ . We will adapt and use this approach below.

Other approaches are based on deconvolution techniques. Among these, Daouia et al. (2020) use a Tikhonov regularization technique to solve the ill-posed deconvolution problem, allowing $\varphi(X)$ and η to be fully nonparametric, but requiring the distribution of the noise to be fully known (e.g., normal with zero mean and known variance). Kneip et al. (2015) need only to specify the family (e.g., normal) of the symmetric density for the noise term ϵ , and prove that the model is identified and provide estimates of all the components of the model. A pseudo-likelihood approach is used for estimation, with the density of η approximated by an histogram. Another extension is provided by Florens et al. (2020), where the noise has an unspecified symmetric distribution and the frontier is nonparametric. Identification is obtained by specifying a flexible density for η based on Laguerre polynomials. Both the stochastic noise and inefficiency are allowed to be heteroscedastic and to depend on some extra environmental variables E , but computation is again difficult due to the required highly nonlinear optimization, and results are in many cases numerically unstable. Sun and Salim (2020) propose a semiparametric model where they estimate an input distance function while avoiding endogeneity problems, but their model involves somewhat more structure than our approach. Prokhorov et al. (2021) estimate a nonparametric stochastic production frontier model with endogenous inputs using a local limited information maximum likelihood method.

With the exception of Sun and Salim (2020), all of the above approaches are developed for a univariate response Y as in (1.1). Extensions to multivariate Y have also been proposed. Some are based on using polar coordinates in the space of the outputs Y , which is natural for radial efficiency measures. This is the approach of Simar (2007), who extends the univariate approach of Hall and Simar (2002), but the noise has to be of “small size,” and Simar and Zelenyuk (2011) who extend the local likelihood approach of Kumbhakar et al. (2007). The latter also provides a stochastic version of traditional envelopment estimators such as data envelopment analysis (DEA) and free-disposal hull (FDH)

estimators. In these approaches, the stochastic deviations from the frontier are radial. However, the statistical properties of these approaches have not been established, primarily because they suffer from endogeneity problems that are ignored in each of the articles mentioned above. In each case, after transformation to polar coordinates, the regressors are functions of the endogenous inputs (for the input orientation) or of the endogenous outputs (in the output orientation). As will be seen below, the transformation in our model avoids such issues, and we are able to establish desirable properties (e.g., consistency) of our estimators.

Other approaches are based on modeling directional distance functions. In parametric settings, this involves regressing one of the outputs on the inputs and on the other outputs divided by the dependent variable. This necessarily involves endogeneity issues that are either ignored or solved by adopting restrictive assumptions. For example, Cuesta and Orea (2002) ignore any possible endogeneity while citing Coelli and Perelman (1996), who argue that since ratios of outputs or inputs appear on the right-hand side of the estimated model, exogeneity can plausibly be assumed. Serra et al. (2011) similarly ignore any possible endogeneity. Malikov et al. (2016) estimate a parametric, quadratic, directional distance function which is homogeneous with respect to directions but not with respect to inputs nor outputs, and use a system of equations to deal with potential endogeneity.

Recently Kuosmanen and Johnson (2017) suggest a way to avoid endogeneity when estimating directional distance functions in a nonparametric approach, but their method requires restrictive assumptions on the DGP (as honestly admitted by the authors), including constant returns to scale, nonstochastic frontier points and homoscedasticity of the stochastic components. In addition, the variance of the noise must tend to zero asymptotically as $n \rightarrow \infty$ as in Hall and Simar (2002). Moreover, the properties of their proposed estimator, based on convex nonparametric least squares in a first stage, remain unknown.

In contrast to the various proposals for less-than-fully-parametric versions of the model in (1.1), our approach allows for both multiple inputs as well as multiple outputs, is computationally tractable, avoids endogeneity problems, and requires only symmetry for the density of the noise term ϵ . The article is organized as follows. The next section establishes notation and the basic statistical model as well as its transformation. In Section 3 we discuss estimation methods based on SVKZ. Section 4 presents empirical results from estimation of the efficiency of U.S. commercial banks, illustrating the practical usefulness of our new methods. Summary and conclusions are given in Section 5.

Some additional results are given in separate appendices. Appendix A, supplementary materials briefly presents evidence on the flexibility of our stochastic specification. Appendix B, supplementary materials shows how the model can be estimated using the HMS extension of SVKZ to deal with observations where residuals are skewed in the unexpected direction, and Appendix C, supplementary materials discusses some alternative approaches that might be used for this problem. Appendix D, supplementary materials discusses some

alternative approaches for estimation, Appendix E, supplementary materials discusses various extensions of our model to estimate hyperbolic or radial distances, introduce environmental variables, etc. Appendix F, supplementary materials describes our Monte Carlo experiments and results, providing evidence on the performance of our estimators and methods. Additional details on our empirical application in Section 4 are given in Appendix G, supplementary materials.

2. The Statistical Model

2.1. Notation

Let $X \in \mathbb{R}_+^p$ and $Y \in \mathbb{R}_+^q$ denote (stochastic) vectors of input and output quantities, respectively, and let $f(x, y)$ denote the joint density of (X, Y) with support on a strict subset of $\mathbb{R}_+^{p+q}$. Similarly, let $x \in \mathbb{R}_+^p$ and $y \in \mathbb{R}_+^q$ denote fixed, nonstochastic vectors of input and output quantities. In the absence of any noise or uncertainty, the production set

$$\Psi := \{(x, y) | x \text{ can produce } y\} \quad (2.1)$$

gives the set of feasible combinations of inputs and outputs, and the support of $f(x, y)$ is a subset of Ψ . In this case, $\Pr((X, Y) \in \Psi) = 1$. But in the presence of noise, the support of $f(x, y)$ may extend outside of Ψ and hence $\Pr((X, Y) \in \Psi) < 1$. Standard assumptions on Ψ in micro-economic theory of the firm include the following.

Assumption 2.1. Ψ is closed.

Assumption 2.1 ensures the existence of the efficient frontier, or technology given by

$$\Psi^\partial := \{(x, y) | (\alpha^{-1}x, \alpha y) \notin \Psi \text{ for all } \alpha > 1\}. \quad (2.2)$$

In cases where $q = 1$, the loci of points in Ψ^∂ comprise a function $\varphi(x)$, that is, the frontier production function.

Let $p \geq 1$, $q \geq 1$ and define $r := p + q$. Define a fixed, nonstochastic, $(r \times 1)$ direction vector $d = (-d'_x, d'_y)'$ where $d_x \in \mathbb{R}_+^p$ and $d_y \in \mathbb{R}_+^q$ (see the separate Appendix E.4, supplementary materials for discussion on the choice of directions). Then for a fixed point $(x, y) \in \mathbb{R}_+^{p+q}$, distance to the frontier in the direction d is measured by the directional distance function

$$\delta(x, y | d) := \sup\{\delta | (x - \delta d_x, y + \delta d_y) \in \Psi\} \quad (2.3)$$

proposed by Chambers et al. (1996, 1998) whenever $\delta(x, y | d)$ exists.

In the absence of noise, properties of nonparametric DEA and FDH estimators of the distance function defined in (2.3) are well-developed; see Kneip et al. (2008) and the sources cited therein or see Simar and Wilson (2013, 2015) for recent surveys. Except for the required assumption of no noise, DEA estimators (which assume strong disposability of inputs and outputs and convexity of Ψ) and FDH estimators (which only assume strong disposability of inputs and outputs) are very flexible, require no functional form assumptions, and easily handle multiple outputs and multiple inputs. DEA and FDH estimators rely on the assumption that the support of the stochastic variables

(X, Y) coincides with a subset of Ψ . For additional technical details, see Kneip et al. (2008).

Regardless of whether noise is present, the production set Ψ , its frontier Ψ^∂ , and the distances defined in (2.3) are unobserved and must be estimated from data. Some additional assumptions are needed to define a statistical model. We first introduce a basic model below in Section 2.2, and then transform this model in Section 2.3 to facilitate estimation. As will be seen, the basic model amounts to a system of equation describing the underlying process that generates the observed data, while the transformed model is used for estimation.

2.2. The Basic Model

We first assume that the underlying theoretical economic process generates exogenous, efficient, but latent, production plans.

Assumption 2.2. The Data Generating Process (DGP) generates random, optimal production plans that belongs to the efficient boundary Ψ^∂ via a well-defined probability mechanism, providing iid values $W_i^\partial := (X_i^\partial, Y_i^\partial)$, $i = 1, \dots, n$.

The optimal plans W_i^∂ (i.e., frontier points) are unobserved due to inefficiency and noise. Now consider a given, fixed direction vector d . As noted in the separate Appendix E.4, supplementary materials empirical researchers often choose a fixed direction vector d (e.g., see the discussion by Färe et al. 2008, p. 534), often with elements equal to means or medians of the corresponding variables. For our purposes, the point is to make d fixed, nonstochastic and the same for all firms (the latter is sometimes called the “egalitarian” evaluation in (2.3)). Let observed input–output pairs be denoted by $W_i := (X_i, Y_i)$.

Assumption 2.3. Given a direction vector d , the set of sample observations $\mathcal{S}_n = \{W_i\}_{i=1}^n$ is generated according the model $W_i = W_i^\partial + \xi_i d$, that is, component by component

$$\begin{bmatrix} X_i \\ Y_i \end{bmatrix} = \begin{bmatrix} X_i^\partial \\ Y_i^\partial \end{bmatrix} + \xi_i d, \quad (2.4)$$

where the ξ_i are, conditionally on W_i^∂ , independent, scalar random variables whose characteristics may depend on W_i^∂ .

Assumption 2.3 makes clear that the stochastic element ξ_i operates on frontier points in the direction d , and that changing the direction vector d results in a different model. See separate Appendix E.4, supplementary materials for additional discussion. Next, we decompose ξ_i into additive inefficiency and noise terms to complete the assumptions of our basic model.

Assumption 2.4. The stochastic term ξ_i in **Assumption 2.3** involves both an inefficiency component $\eta_i \in \mathbb{R}_+$ and a noise component $\epsilon_i \in \mathbb{R}$, that is

$$\xi_i := \epsilon_i - \eta_i, \quad (2.5)$$

where conditionally on W_i^∂ , ϵ_i and η_i are independent, with

$$\epsilon_i | W_i^\partial \sim \text{Sym}\left(0, \tilde{\sigma}_\epsilon(W_i^\partial)\right) \quad (2.6)$$

and

$$\eta_i | W_i^\partial \sim D_+(\tilde{\sigma}_\eta(W_i^\partial)), \quad (2.7)$$

where $\text{Sym}(0, a)$ denotes a symmetric (around zero), two-sided distribution on $\mathbb{R}$ with standard deviation $a \geq 0$, and $D_+(b)$ denotes a one-sided distribution on $\mathbb{R}_+$ belonging to some one-parameter scale family with parameter $b \geq 0$.

A density $f(\cdot)$ belongs to a one-parameter scale family if it can be written as $f(\cdot) = (1/\sigma)\tilde{f}(\cdot/\sigma)$ for some $\sigma > 0$, where $\tilde{f}(\cdot)$ is any density on $\mathbb{R}_+$. We will call $\tilde{f}$ the *reference* density of the family. Examples include the Half-Normal and Exponential distributions, and the Gamma and Weibull distributions with fixed shape parameters.

Both the inefficiency and the noise processes in (2.4) operate on both inputs and outputs, as is standard in the literature on directional distance functions. [Assumptions 2.1–2.4](#) ensure we have a random sample $\mathcal{S}_n = \{(X_i, Y_i)\}$ of size n with the latent efficient variables W_i^∂ playing the role of unobserved, exogenous variables. The components ϵ_i and η_i of the independent ξ_i may be heteroscedastic. These assumptions induce the joint density $f(x, y)$ of the observables (X, Y) .

As an aside, consider the following special case of our basic model. Suppose that $q = 1$ (univariate output) and that the chosen direction is output oriented with $d = (0'_p, 1)'$ where 0_p is a p -vector of zeros. Then (2.4) simplifies to

$$\begin{aligned} X_i &= X_i^\partial, \\ Y_i &= Y_i^\partial + \epsilon_i - \eta_i. \end{aligned} \quad (2.8)$$

Since $X_i = X_i^\partial$ is observed (because the corresponding elements of d are zero), and Y_i is univariate, we can in this case describe the frontier points by the production function $\mathbb{R}_+^p \mapsto \mathbb{R}_+^1$ such that $Y_i^\partial = Y^\partial(X_i)$. By conditioning on X_i , the model in (2.8) can be written as

$$Y_i = Y^\partial(X_i) + \epsilon_i - \eta_i, \quad (2.9)$$

where $\epsilon_i | X_i \sim \text{Sym}(0, \sigma_\epsilon(X_i))$ and $\eta_i | X_i \sim D_+(\sigma_\eta(X_i))$ since the parameters of the scale functions in (2.6)–(2.7) can now be expressed as functions of X_i . The model in (2.9) is identical to the output-oriented model suggested by Simar et al. (2017) for the case of a univariate output Y . Kumbhakar et al. (2007) use the same model with a specific localized distribution (i.e., Gaussian) for ϵ_i in order to derive the local likelihood function. Moreover, it is obvious that (2.9) nests the fully parametric models of Aigner et al. (1977), Battese and Corra (1977), and Meeusen and van den Broeck (1977) as a special case, but allows far more flexibility than the fully parametric models. In addition, (2.9) is a special case of our basic model in (2.4), which allows $q > 1$.

2.3. The Transformed Model

In order to estimate the frontier as well as features of the inefficiency process, we translate the problem to a new coordinate system to obtain a transformed version of the basic model defined by [Assumptions 2.1–2.4](#). For the r -dimensional direction vector d , let $\mathcal{V}(d)$ denote the $(r - 1)$ -dimensional linear

space of all vectors $\omega \in \mathbb{R}_+^r$ such that $\omega'd = 0$. Let $V_d = [v_1 \dots v_{r-1}]$ denote an arbitrary, but fixed orthonormal basis for $\mathcal{V}(d)$. Then V_d is an $r \times (r - 1)$ matrix, with $v_j'v_j = 1$, $v_j'v_k = 0$ for $j \neq k$, and $v_j'd = 0$. Consequently, $V_d'V_d = I_{r-1}$ where I_{r-1} denotes the $(r-1) \times (r-1)$ identity matrix. Moreover, V_d depends only on the given direction vector d . Using the matrix V_d , define the $(r \times r)$ rotation matrix

$$R_d = \begin{bmatrix} V_d' \\ d'/\|d\| \end{bmatrix} \quad (2.10)$$

whose transpose is $R_d' = [V_d \ d/\|d\|]$ and where $\|\cdot\|$ denotes the L_2 -norm. The matrix V_d is not unique, but this does not create any problem for what follows provided V_d is fixed after it is selected. Clearly, R_d is an orthogonal matrix. Hence, $R_d'R_d = I_r$, implying $R_d^{-1} = R_d'$. Moreover, by construction, $R_d d = [0'_{r-1} \ \|d\|]'$. It is important to remember that the rotation matrix R_d is specific to the given direction vector d as signified by the subscript “ d .” If the direction vector is changed, this will affect the rotation matrix as well as everything that follows. But as with V_d , R_d depends only upon the given direction vector d .

In the special case of a univariate output ($q = 1$) with output-oriented direction $d = (0'_p, 1)'$ considered above in (2.8), it is easy to show that the matrix R_d is given by the identity matrix of order r . In this case, no transformation of the basic model is needed, since the setup is the same as that of SVKZ. Otherwise, however, we can transform the observations $W_i = (X_i, Y_i) \in \mathcal{S}_n$ to a new coordinate system by applying the rotation matrix R_d in (2.10) to obtain

$$\begin{bmatrix} Z_i \\ U_i \end{bmatrix} = R_d W_i = R_d \begin{bmatrix} X_i \\ Y_i \end{bmatrix}, \quad (2.11)$$

where $Z_i \in \mathbb{R}^{r-1}$ and $U_i \in \mathbb{R}^1$ for each $i = 1, \dots, n$. Moreover, the independent, identically distributed (iid) sample $\mathcal{S}_n$ is transformed to an iid sample $\mathcal{S}_n(d) = \{(Z_i, U_i)\}_{i=1}^n$. Obviously, the (Z_i, U_i) depend on the direction vector d through the rotation matrix R_d , but we avoid further complicating the notation. Note that $U_i = d'W_i/\|d\|$ and hence is invariant to the ordering of the inputs and outputs since the ordering of the elements of the direction vector d necessarily corresponds to whatever order of the elements of X_i and Y_i is chosen.

From a geometrical point of view, we have a linear transformation from $w = (x, y) \in \mathbb{R}^r$ to $(z, u) \in \mathbb{R}^r$, where $[z' \ u']' = R_d w$. This one-to-one transformation can be inverted by writing $w = R_d' [z' \ u']'$. This is illustrated in the case $p = q = 1$ in [Figure 1](#) where the origin in the new (z, u) -space corresponds to the origin in the original (x, y) -space. The rotation transforms the production set Ψ in (x, y) -space to the set

$$\Gamma_d = \{(z, u) \in \mathbb{R}^r | R_d' [z' \ u']' \in \Psi\} \quad (2.12)$$

in (z, u) -space. Moreover, the density $f(x, y)$ is transformed to a new density $g(z, u) = g(u|z)g(z)$. Since u is in the direction d , and the $r - 1$ coordinates z are orthogonal to d , the frontier Ψ^∂ in (x, y) -space is transformed to the function $\phi(z): \mathbb{R}_+^{r-1} \mapsto \mathbb{R}^1$ in (z, u) -space such that

$$\phi(z) = \sup \{u | (z, u) \in \Gamma_d\}. \quad (2.13)$$

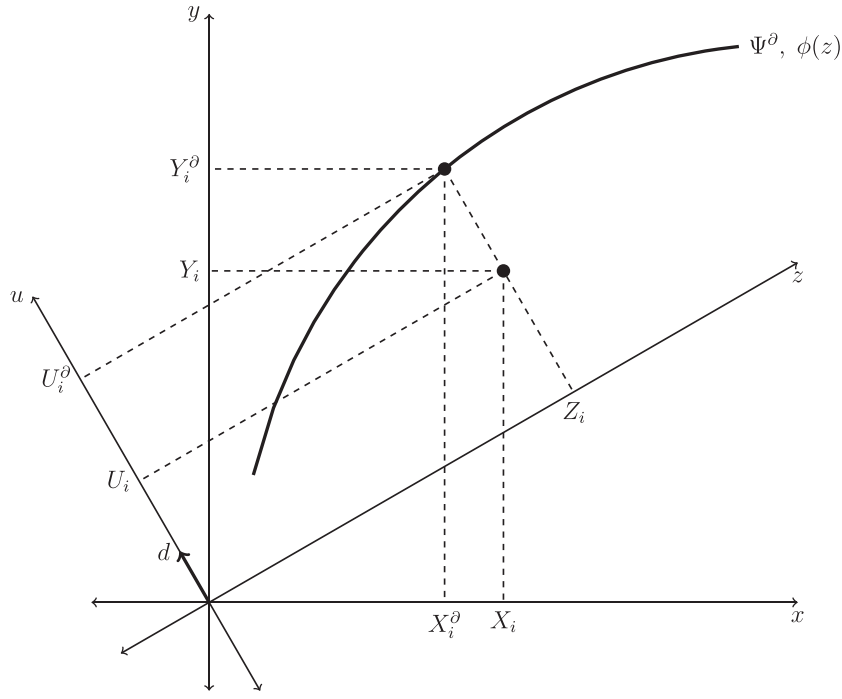


Figure 1. Transformation from (x, y) -Space to (z, u) -Space.

In (x, y) -space, the frontier is an $(r-1)$ -manifold, while in (z, u) -space it becomes a scalar-valued, $(r-1)$ -variate function $\phi(z)$. So we have

$$\Gamma_d = \{(z, u) \in \mathbb{R}^r | u \leq \phi(z)\}. \quad (2.14)$$

The frontier points $(X_i^\partial, Y_i^\partial)$ have coordinates (Z_i, U_i^∂) in (z, u) -space, and the observations (X_i, Y_i) have coordinates (Z_i, U_i) , as illustrated in Figure 1. For each $i = 1, \dots, n$, the random deviations from $(X_i^\partial, Y_i^\partial)$ along the direction d to observations (X_i, Y_i) correspond to random deviations from (Z_i, U_i^∂) to the observations (Z_i, U_i) in (z, u) -space. The heteroscedastic nature of ξ_i is now captured by the values of Z_i .

Pre-multiplying both sides of the basic model in (2.4) by R_d yields the transformed model

$$\begin{bmatrix} Z_i \\ U_i \end{bmatrix} = \begin{bmatrix} Z_i^\partial \\ U_i^\partial \end{bmatrix} + \xi_i \begin{bmatrix} 0_{r-1} \\ ||d|| \end{bmatrix}, \quad (2.15)$$

Then the transformed model in (2.15) can be written equivalently as

$$\begin{aligned} Z_i &= Z_i^\partial, \\ U_i &= U_i^\partial + ||d||\epsilon_i - ||d||\eta_i, \end{aligned} \quad (2.16)$$

where the equation for U_i is univariate. Note that (2.16) has structure analogous to that of the special case of our basic model described by (2.8). Since $Z_i = Z_i^\partial$ is observed, we can rewrite the frontier points in the (z, u) -space as $(Z_i, U_i^\partial) = (Z_i, \phi(Z_i))$. Therefore, by conditioning on Z_i , the second equation of (2.16) can be written as

$$U_i = \phi(Z_i) + ||d||\epsilon_i - ||d||\eta_i, \quad (2.17)$$

where the scale functions $\tilde{\sigma}_\epsilon(W_i^\partial)$ and $\tilde{\sigma}_\eta(W_i^\partial)$ in ξ_i can be expressed as functions of Z_i since W_i^∂ can be expressed as a

function of Z_i only. Hence, the heteroscedastic features defined in (2.6)–(2.7) of Assumption 2.4 can be written as

$$\epsilon_i | Z_i \sim \text{Sym}(0, \sigma_\epsilon(Z_i)) \quad (2.18)$$

and

$$\eta_i | Z_i \sim D_+(\sigma_\eta(Z_i)). \quad (2.19)$$

It is important to note that

$$\begin{bmatrix} Z_i^\partial \\ U_i^\partial \end{bmatrix} = R_d \begin{bmatrix} X_i^\partial \\ Y_i^\partial \end{bmatrix} = R_d W_i^\partial. \quad (2.20)$$

The first line of (2.16) results from the facts that (i) R_d depends only upon the fixed d , and (ii) $R_d d = [0_{r-1} \ 1]'$ by construction as discussed above. Consequently, by Assumption 2.2 and the first line of (2.16), $Z_i = Z_i^\partial$ and therefore Z_i is exogenous since W_i^∂ , and hence Z_i^∂ , are exogenous. In the transformed model (2.17), only the scalar left-hand side variable U_i depends on the stochastic terms ϵ_i and η_i . In addition, U_i depends on the exogenous Z_i , and hence U_i is endogenous. This greatly simplifies estimation, and stands in sharp contrast to most of the literature where researchers work in the original (x, y) -space where all of the elements of X_i and Y_i are potentially endogenous. Of course this depends on a fixed, exogenously-chosen direction vector, but this is in fact what most empirical researchers adopt. See the separate Appendix E.4, supplementary materials for additional discussion regarding the choice of direction vector d .

In order to identify the frontier function $\phi(Z_i)$ in (2.13), a functional form for the distribution D_+ in (2.19) must be specified. Various possibilities are mentioned above following Assumption 2.4; for purposes of illustration, we adopt the following.

Assumption 2.5. The distribution of η , conditional on Z , is half-normal with scale parameter $\sigma_\eta(Z)$, that is,

$$\eta|Z \sim N^+(0, \sigma_\eta^2(Z)). \quad (2.21)$$

In addition, for purposes of estimation using local-linear methods along the lines of SVKZ, two additional assumptions are needed.

Assumption 2.6. For any orthonormal basis V_d of $\mathcal{V}(d)$, the function $\phi(z)$ defined in (2.13) is differentiable at all points z in its domain.

Assumption 2.7. For all $i \in \{1, 2, \dots\}$, the conditional moments $\mathbb{E}(|U_i|^{2+\nu}|Z_i) \mathbb{E}(|(\xi_i + \mu_\eta(Z_i))^3|^{2+\nu^*}|Z_i)$ exist for some $\nu > 0$ and $\nu^* > 0$.

For a given Z_i , there is a unique U_i^∂ and W_i^∂ in the direction d . By [Assumption 2.4](#) and (2.19), $\mathbb{E}(\eta_i^\partial|Z_i) = k_j \sigma_\eta^j(Z_i)$ for all $j \geq 1$, provided k_j , the j th moment of the reference density of the family $\tilde{f}$ exists (see the remarks following [Assumption 2.4](#)). Moreover, [Assumption 2.4](#) ensures that the conditional mean efficiency function is given by

$$\mu_\eta(Z_i) := \mathbb{E}(\eta_i|Z_i) = k_1 \sigma_\eta(Z_i). \quad (2.22)$$

Under [Assumption 2.5](#), $k_1 = \sqrt{2/\pi}$. Replacing [Assumption 2.5](#) with a different distributional assumption will change the value of k_1 . [Assumption 2.6](#) imposes some smoothness on the frontier, in all directions, while [Assumption 2.7](#) imposes some minor restrictions on the tail behavior of the distributions of ϵ_i and η_i in (2.18)–(2.19). Both assumptions are needed to ensure consistency of the local-linear estimators used below in [Section 3](#).

In [Section 3](#), we show how the unknown functions $\phi(Z_i)$, $\sigma_\epsilon^2(Z_i)$ and $\sigma_\eta(Z_i)$ can be easily estimated using the transformed model in (2.17) using standard nonparametric regression estimators (e.g., local polynomial or orthogonal series estimators). In particular the “ σ functions” in (2.18)–(2.19) will be functional parameters, enhancing the flexibility of the model. Estimates of the conditional mean efficiency function in (2.22) are obtained directly from the transformed model. Estimates of features of the basic model, such as marginal rates of substitution or transformation, marginal products, or derivatives of conditional mean efficiency with respect to inputs or outputs, can be recovered easily from estimates of the transformed model due to the fact that the transformation from (x, y) -space to (z, u) -space is one-to-one as shown below in [Section 2.4](#).

2.4. Links between the Two Spaces

Recall that the direction d and the chosen matrix V_d are fixed as discussed earlier in [Section 2.3](#). In applied economic studies, researchers are often interested in marginal effects such as marginal productivity, marginal rates of substitution, etc. As demonstrated below, it is possible to recover partial derivatives in the original (x, y) -space from derivatives obtained in the new, transformed (z, u) -space introduced in [Section 2.3](#).

First, the locus of the frontier function in the transformed model is given by the set of points $(z, \phi(z))$, where $z \in \mathbb{R}^{r-1}$.

From this the frontier surface in (x, y) -space (an $(r - 1)$ -manifold) can be obtained from

$$w^\partial = \begin{bmatrix} x^\partial \\ y^\partial \end{bmatrix} = R'_d \begin{bmatrix} z \\ \phi(z) \end{bmatrix} = \begin{bmatrix} V_d & \frac{d}{\|d\|} \end{bmatrix} \begin{bmatrix} z \\ \phi(z) \end{bmatrix}, \quad (2.23)$$

which can be viewed as a mapping $F: \mathbb{R}^{r-1} \mapsto \mathbb{R}^r$, where $w^\partial = F(z)$ is a manifold *parameterized* by F . The Jacobian of the transformation in (2.23) can be written as the $r \times (r - 1)$ matrix

$$J_\phi(z) = \left\{ \frac{\partial w_i^\partial}{\partial z_j} \right\} = V_d + \frac{d}{\|d\|} \frac{\partial \phi(z)}{\partial z'}, \quad (2.24)$$

where $i = 1, \dots, r$ and $j = 1, \dots, (r - 1)$. Existence of the Jacobian in (2.24) is ensured by [Assumption 2.6](#).

Now consider a fixed point $(z_0, \phi(z_0))$ on the frontier. Then from (2.23) the image on the frontier manifold in w -space is $w_0^\partial = F(z_0)$. Standard results from real analysis (e.g., [Lang 1964](#), pp. 164–174) ensure that under [Assumption 2.6](#),

$$F(z_0 + \vartheta) = F(z_0) + J_\phi(z_0)\vartheta + \|\vartheta\| G(\vartheta), \quad (2.25)$$

where ϑ is a small vector in $\mathbb{R}^{r-1}$ and $\lim_{\|\vartheta\| \rightarrow 0} G(\vartheta) = 0$. Therefore, $J_\phi(z_0)$ defines a linear map from $\mathbb{R}^{r-1}$ to $\mathbb{R}^r$ which can be viewed as an approximation of F near the point z_0 , up to an error term whose magnitude is small. This linear map is said to be a tangent hyperplane to F at the point z_0 . Since F is differentiable at z_0 by [Assumption 2.6](#), this hyperplane is unique due to Theorem 1 of [Lang \(1964, p. 166\)](#).

Clearly, $w = F(z_0 + \vartheta)$ in (x, y) -space. Therefore, the equation of the unique hyperplane tangent to the surface $F(z)$ at the point z_0 is given by

$$w = w_0^\partial + J_\phi(z_0)\vartheta. \quad (2.26)$$

This hyperplane can be written in explicit form by the linear equation

$$c'(w - w_0^\partial) = c'J_\phi(z_0)\vartheta = 0 \quad (2.27)$$

where $c \in \mathbb{R}^r$. The vector c is identified up to a multiplicative constant by the equation $c'J_\phi(z_0) = 0'_{r-1}$, or equivalently,

$$J'_\phi(z_0)c = 0_{r-1}. \quad (2.28)$$

In other words, c is determined (up to a scalar constant) by a basis of the null space of the matrix $J'_\phi(z_0)$.

Using the equation of the tangent hyperplane in (2.27), it is obvious that any component w_ℓ of w can be expressed as a function of the other elements w_k , $k = 1, \dots, r$, $k \neq \ell$. Hence, partial derivatives in (x, y) -space can be written as

$$\left\{ \frac{\partial w_\ell}{\partial w_k} \right\}_{w=w_0^\partial} = -\frac{c_k}{c_\ell} \quad (2.29)$$

provided $c_\ell \neq 0$. As shown below in [Section 3](#), the result in (2.29) can be used to estimate marginal products $\partial Y_\ell / \partial X_k \forall k \in \{1, \dots, p\}$ and $\ell \in \{1, \dots, q\}$, marginal rates of substitution $\partial X_\ell / \partial X_k \forall k, \ell \in \{1, \dots, p\}$, $k \neq \ell$, and marginal rates of transformation $\partial Y_\ell / \partial Y_k \forall k, \ell \in \{1, \dots, q\}$, $k \neq \ell$. If $c_\ell = 0$, then the derivative in (2.29) is undefined. However, in this case,

$$\left\{ \frac{\partial w_k}{\partial w_\ell} \right\}_{w=w_0^\partial} = 0 \quad (2.30)$$

provided $c_k \neq 0$, indicating that the variable w_k has no effect on the component w_k at the particular frontier point of interest.

Another interesting feature of the transformed model is the function $\mu_\eta(z)$ defined in (2.22). It is clear that the vector $\partial\mu_\eta(w^\partial)/\partial w^\partial$ can be recovered from knowledge of the vector $\partial\mu_\eta(z)/\partial z$. Since $z = V'_d w^\partial$, the Jacobian of the mapping from w^∂ to z is

$$J_\mu = \left[\frac{\partial z_i}{\partial w^\partial_j} \right] = V'_d, \quad (2.31)$$

with $i = 1, \dots, r-1$ and $j = 1, \dots, r$. Applying the chain rule yields

$$\frac{\partial\mu_\eta(w^\partial)}{\partial w^\partial} = V_d \frac{\partial\mu_\eta(z)}{\partial z}, \quad (2.32)$$

which is an r -vector of derivatives of mean inefficiency with respect to the components of w at a point w^∂ .

Details on estimation are given in the next section.

3. Estimation and Inference

3.1. Estimation Methodology and Basic Properties

As discussed above in Section 2.3, the transformed model given by (2.17)–(2.19) involves a scalar response variable U_i , a link function $\phi(Z_i)$ and a composite error term $\xi_i = \epsilon_i - \eta_i$. The distributions of $\epsilon_i \in \mathbb{R}$ and $\eta_i \in \mathbb{R}_+$, conditional on Z_i , are given in (2.18)–(2.19). These are necessarily related to the direction vector d through the dependence on Z_i . It is easy to prove that the model is identified once the reference family is chosen (as in Assumption 2.5) using the sufficient condition for identification given by Florens et al. (2020).

We use the moments-based method of SVKZ and its extension described by HMS to estimate our transformed model. To conserve space, we discuss here how the SVKZ method can be used; a discussion of the HMS method appears in the separate Appendix B, supplementary materials. In addition, alternative methods that might be used are discussed briefly in the separate Appendix D, supplementary materials. However, our stochastic specification is quite flexible since it is localized (i.e., the specification may change with the value of z), and the convolution of a one-parameter scale family density with a symmetric density is also quite flexible (see separate Appendix A, supplementary materials for further discussion and illustration). Another advantage of our estimation approach is its numerical easiness and relatively low computational requirements, since only local least-squares methods are used and most of the relevant information about the model can be recovered by assuming only symmetry of ϵ and selecting a specific one-parameter scale family for the distribution of efficiency.

We first show how the transformed model given by (2.17)–(2.19) can be estimated. The estimation method can be viewed as a nonparametric version of the modified OLS procedure suggested by Richmond (1974) that is sometimes used in fully parametric settings (e.g., see Lovell 1993 for discussion). Our approach allows the link function $\phi(z)$ to remain nonparametric and uses only information of the first three local moments of the random terms. Consequently, the method used here is less sensitive to error in specifications of the local densities of the

stochastic terms than some of the approaches discussed in the separate Appendix D, supplementary materials.

To explain the procedure, note that $\mu_\eta(Z_i)$ defined in (2.22) must be nonnegative for all $i = 1, \dots, n$. Define $\varepsilon_i := \xi_i + \mu_\eta(Z_i)$ and $r_1(Z_i) := \phi(Z_i) - |d|\mu_\eta(Z_i)$. Clearly, $\mathbb{E}(\varepsilon_i|Z_i) = 0$. Then necessarily

$$U_i = r_1(Z_i) + |d|\varepsilon_i \quad (3.1)$$

and $\mathbb{E}(U_i|Z_i) = r_1(Z_i) = \phi(Z_i) - |d|\mu_\eta(Z_i)$. The conditional mean function $r_1(Z_i)$ can be viewed as an average production function that can be estimated using standard nonparametric regression estimators (e.g., the local-linear estimator).

The assumption of symmetry of $\epsilon_i|Z_i$ ensures that

$$\mathbb{E}(\varepsilon|z) = 0, \quad (3.2)$$

$$\mathbb{E}(\varepsilon^2|z) = \text{var}(\epsilon|z) + \text{var}(\eta|z) > 0 \quad (3.3)$$

and

$$\mathbb{E}(\varepsilon^3|z) = -\mathbb{E}\left[(\eta - \mu_\eta(z))^3|z\right]. \quad (3.4)$$

Now let $r_j(z) := \mathbb{E}(\varepsilon^j|z)$ for $j \in \{2, 3\}$. Under Assumptions 2.4 and 2.5, it is straightforward to show that

$$\mu_\eta(z) = \mathbb{E}(\eta|z) = (2/\pi)^{1/2}\sigma_\eta(z), \quad (3.5)$$

$$r_2(z) = \sigma_\epsilon^2(z) + (\pi - 2)\pi^{-1}\sigma_\eta^2(z) \quad (3.6)$$

and

$$r_3(z) = (2/\pi)^{1/2}(\pi - 4)\pi^{-1}\sigma_\eta^3(z) \leq 0. \quad (3.7)$$

From (3.7) is clear that the scale function $\sigma_\eta(z)$ is identified by the third (local) moment $r_3(z)$, and this in turn provides identification of $\mu_\eta(z)$.

Estimation is straightforward. First, (3.1) is estimated by local linear least squares (LLS), yielding estimates $\hat{r}_1(z)$ and $\hat{\nabla}r_1(z)$ of $r_1(z)$ and $\nabla r_1(z)$, respectively. The properties of these estimators are well-known; for example, see Fan and Gijbels (1996) or Li and Racine (2007). Second, compute the residuals $\hat{\varepsilon}_i = |d|^{-1}(U_i - \hat{r}_1(Z_i))$ for $i = 1, \dots, n$, and again using LLS, regress $\hat{\varepsilon}_i^2$ and $\hat{\varepsilon}_i^3$ on the Z_i , $i = 1, \dots, n$ to obtain corresponding estimates $\hat{r}_j(Z_i)$ and $\hat{\nabla}r_j(Z_i)$ of $r_j(Z_i)$ and $\nabla r_j(Z_i)$, $j \in \{2, 3\}$. As noted in the Appendix of SVKZ, reasoning along the lines of Masry (1996) establishes that $\sup_z |\hat{r}_1(z) - r_1(z)| = O_p((nh^{r-1})^{-1/2} \log n + h^2)$. With least-squares cross-validation for the bandwidths this is clearly $o_p(n^{-1/(r+3)})$, and hence $\hat{r}_1(z)$ converges uniformly. SVKZ then prove, under the additional assumption that the conditional moments of ε of order up to 7 are finite, and selecting the bandwidths for the regression of $\hat{\varepsilon}_i^2$ and of $\hat{\varepsilon}_i^3$ on z such that $nh^{r-1}(\log n)^{-4} \rightarrow \infty$ and $nh^{r+3} \rightarrow 0$ as $n \rightarrow \infty$, we have

$$(nh^{r-1})^{1/2}(\hat{r}_j(z) - r_j(z)) \xrightarrow{\mathcal{L}} N(0, s_j^2(z)) \quad (3.8)$$

for $j \in \{2, 3\}$ and where the expression the variances are given in SVKZ for the local constant regression case. In practice the variances have to be evaluated by bootstrap techniques. SVKZ note that, as usual, for avoiding the asymptotic bias in (3.8), the bandwidths h must be of slightly smaller order, for example, $h = O(n^{-(1+\nu)/(r+3)})$ for some $0 < \nu < 4/(r-1)$. Since the

bandwidths for the r components of the Z_i have the same order, then the $\hat{r}_j(z)$, $j \in \{2, 3\}$ converge at rate $(nh^{p+q-1})^{1/2}$ where the bandwidth h is written without a subscript since all are of the same order. It is well-known that the optimal order of h for local-linear regression is $h = O(n^{-1/(p+q+3)})$, which leads to an optimal rate $n^{2/(p+q+3)}$ for the $\hat{r}_j(z)$. Here, as elsewhere with nonparametric regression, there is no escape from the curse of dimensionality, but there is minimal risk of specification error.

Substituting $\hat{r}_3(z)$ into (3.7) and rearranging terms yields the estimator

$$\hat{\sigma}_\eta(z) = \max \left\{ 0, \left[(\pi/2)^{1/2} \pi (\pi - 4)^{-1} \hat{r}_3(z) \right]^{1/3} \right\} \geq 0 \quad (3.9)$$

of the scale function $\sigma_\eta(z)$. The max operator is used to ensure that the estimate is nonnegative, as in Olson et al. (1980); note that the HMS estimator discussed in the separate Appendix B, supplementary materials avoids this truncation. Substituting $\hat{\sigma}_\eta(z)$ into (3.5) yields an estimator of mean conditional inefficiency $\mu_\eta(z)$, that is,

$$\hat{\mu}_\eta(z) = (2/\pi)^{1/2} \hat{\sigma}_\eta(z). \quad (3.10)$$

Similarly, an estimator

$$\hat{\sigma}_\epsilon^2(z) = \max \left\{ 0, \hat{r}_2(z) - (\pi - 2)\pi^{-1} \hat{\sigma}_\eta^2(z) \right\} \geq 0 \quad (3.11)$$

of the variance of ϵ is obtained from (3.6), although this is not needed if one is only interested in estimates of the frontier and of mean conditional inefficiency. Finally, an estimator of the frontier function in the transformed space is given by

$$\hat{\phi}(z) = \hat{r}_1(z) + ||d|| \hat{\mu}_\eta(z). \quad (3.12)$$

In addition, the partial derivative estimates $\hat{\nabla} r_j(z)$, $j = 1, \dots, 3$ can be used to recover estimates $\hat{\nabla} \mu_\eta(z)$ and $\hat{\nabla} \phi(z)$ of the partial derivatives of $\mu_\eta(z)$ and $\phi(z)$ with respect to the r elements of z .

Properties of the estimators $\hat{\sigma}_\eta(z)$, $\hat{\mu}_\eta(z)$, $\hat{\sigma}_\epsilon^2(z)$, and $\hat{\phi}(z)$ are obviously inherited from the asymptotic properties of $\hat{r}_j(z)$, $j \in \{1, 2, 3\}$ discussed above. The fact that truncation is used to define $\hat{\sigma}_\eta(z)$ $\hat{\sigma}_\epsilon^2(z)$ has no effect on their asymptotic properties since under the assumptions of the model, $\Pr(\hat{r}_3(z) \geq 0) \rightarrow 0$ as $n \rightarrow \infty$. Then it follows that asymptotically, $\hat{\mu}_\eta(z)$ is a continuous function of $\hat{r}_3(z)$. Thus, a standard delta method establishes asymptotic normality of $\hat{\mu}_\eta(z)$, with the same rate of convergence as $\hat{r}_3(z)$, that is, $(nh^{r-1})^{1/2}$. Theorem 2.7 of Li and Racine (2007) ensures that the LLLS estimators $\nabla r_j(z)$, $j \in \{1, 2, 3\}$ are similarly asymptotically normally distributed and also converge at rate $(nh^{r+1})^{1/2}$. The same reasoning then establishes that $\hat{\nabla} \mu_\eta(z)$ and $\hat{\nabla} \phi(z)$ are also asymptotically normally distributed and converge at rate $(nh^{r+1})^{1/2}$.

One can also derive predictions of individual efficiencies for each observation, using a local version of the Jondrow et al. (1982) approach. However, this requires specification of a particular density in the family $\text{Sym}(0, \sigma_\epsilon(z))$ (e.g., normality). Then the frontier estimates $\hat{\phi}(Z_i)$ can be used to compute residuals $\hat{\xi}_i = ||d||^{-1} (U_i - \hat{\phi}(Z_i))$, which are in turn used to compute $\hat{\eta}_i = \hat{\mathbb{E}}(\eta_i | \hat{\xi}_i, Z_i)$ along the lines of Jondrow et al. (1982). However, this should be used with caution. As observed by SVKZ, the resulting $\hat{\eta}_i$ are predicted values based on a *single, particular* realization $\hat{\xi}_i$. In fact, the $\hat{\eta}_i$ are not predictions, but

rather plug-in estimates of the conditional mean $\mathbb{E}(\eta_i | \xi_i, Z_i)$ as discussed by Simar and Wilson (2010) and Wheat et al. (2014). Even so, confidence intervals for $\hat{\eta}_i$ are likely to be quite wide since the estimate is based on a single observation. Moreover, if one is interested in predictions, one should estimate prediction intervals as discussed by Simar and Wilson (2010) instead of confidence intervals for the conditional mean.

As noted above, the derivatives $\partial \mu_\eta(z)/\partial z$ and $\partial \phi(z)/\partial z$ can be estimated easily using local-linear (or local-quadratic) estimation of the transformed model. By inverting the transformation in (2.11), the points $(Z_i, \hat{\phi}(Z_i))$ can be translated to estimates of the frontier surface in the original (x, y) coordinate system. This leads to

$$\hat{W}_i^\partial = \begin{bmatrix} X_i^\partial \\ Y_i^\partial \end{bmatrix} = R_d' \begin{bmatrix} Z_i \\ \hat{\phi}(Z_i) \end{bmatrix}, \quad (3.13)$$

where the matrix R_d defined earlier is known and nonstochastic. Similarly, estimates of points on the average production function $r_1(z)$ can be obtained in the original space via

$$\hat{W}_i^{\text{aver}} = \begin{bmatrix} X_i^{\text{aver}} \\ Y_i^{\text{aver}} \end{bmatrix} = R_d' \begin{bmatrix} Z_i \\ \hat{r}_1(Z_i) \end{bmatrix}. \quad (3.14)$$

Since these transformations are nonstochastic, it is clear that the asymptotic properties of the estimators in the transformed (z, u) -space remain valid in the original (x, y) -space with the same rate of convergence (i.e., $n^{2/(p+q+3)}$).

The same is true for the other quantities of interest derived in Section 2.4. In particular, the derivatives of expected efficiency with respect to inputs and outputs given in (2.32) can be estimated by noting first that differentiation in (3.5) yields

$$\frac{\partial \mu_\eta(z)}{\partial z} = \sqrt{\frac{2}{\pi}} \frac{\partial \sigma_\eta(z)}{\partial z}. \quad (3.15)$$

The derivative $\frac{\partial \sigma_\eta(z)}{\partial z}$ can be easily estimated, and repeated substitution leads to an estimate of $\partial \mu_\eta(w^\partial)/\partial w^\partial$ in (2.32).

Estimates of the marginal effects given by (2.29) can also be computed, giving estimates of marginal products, marginal rates of substitution and marginal rates of transformation. LLLS estimation of the average production function $r_1(z)$ provides estimates of $\partial r_1(Z_i)/\partial Z_i$, and from the discussion above we have estimates of $\partial \mu_\eta(Z_i)/\partial Z_i$. Adding these give estimates of $\partial \phi(Z_i)/\partial Z_i$ due to (3.12), and these can be used to compute estimates $\hat{J}_\phi(Z_i)$ of the Jacobian matrix $J_\phi(Z_i)$ given in 2.24. Substituting $\hat{J}_\phi(Z_i)$ for $J_\phi(Z_i)$ in (2.28) and solving for the vector c allows estimates of the marginal effects given in (2.29) to be computed.

3.2. Further Details

Several extensions of the ideas developed above are possible. To conserve space, we give brief discussions of these in the separate Appendix E, supplementary materials, with systematic analysis of the ideas left for future work. In Section E.1, we show how the methods we have described above can be adapted to estimate radial input, output and hyperbolic distances. In Section E.2, we discuss how exogenous environmental variables can be introduced into the analysis. This allows investigation

of the effect of such variables on the production frontier Ψ^0 as well as on the distribution of inefficiency. Section E.3 discusses how the SVKZ and HMS estimators can be extended to impose free disposability or convexity on the production set Ψ , thereby providing stochastic analogs of FDH or DEA estimators. Finally, Section E.4 discusses the consequences of allowing the choice of direction d to be data-driven.

4. Efficiency of U.S. Banks

We use the methods developed above to estimate technical efficiency and scale elasticities of U.S. commercial banks operating during first-quarter 2001 (2001Q1) through fourth-quarter 2020 (2020Q4). We use quarterly data from the U.S. Federal Deposit Insurance Corporation Statements of Income and Condition (Call Reports). Our data cover the periods before, during and after the financial crisis of 2007–2009, and thus allow us to examine how banks responded to the crisis. Following prior work (e.g., Wheelock and Wilson 2018), and using subscript i to index observations $1, \dots, n$, we define five inputs, that is, purchased funds (X_{i1}), labor (X_{i2}), physical capital (X_{i3}), equity (X_{i4}) and risk (X_{i5}). Labor is measured in terms of full-time equivalents, and physical capital is measured by the book value of premises and fixed assets as is typical in banking studies. All dollar amounts are measured in thousands of constant 2012 U.S. dollars.

Consistent with Wheelock and Wilson (2018) and other studies, we view both equity and risk as quasi-fixed inputs. Equity is another source of financial capital, but as noted by Akhavein et al. (1997) and Berger and Mester (1997) is very difficult and costly to adjust substantially in the short run. Risk is measured in terms of nonperforming loans as in Wheelock and Wilson (2018). Others, including Malikov et al. (2016) and Simper et al. (2017), view nonperforming loans as an undesirable output. Since we hold risk constant, it can also be viewed as a “netput,” and the results will suggest whether it behaves as an input or an output as discussed below.

We define three outputs, including total loans (Y_{i1}), securities (Y_{i2}) and off-balance sheet items (Y_{i3}). Off-balance sheet items are measured by total noninterest income and reflects fiduciary activities, foreign exchange and securities trading, data processing services and other activities.

After deleting observations with missing or implausible values as well as credit-card, bankers’ cash management, nonbank, pure internet, cooperative and wholesale banks, depository trust companies and industrial loan companies, we have 486,461 observations, with number of observations per quarter ranging from 8001 in 2001Q1 to 4261 in 2020Q4, reflecting the on-going consolidation in the U.S. banking industry through mergers and acquisitions. Summary statistics for these variables are given in Table G.1 appearing in the separate Appendix G, supplementary materials. Table G.1 indicates that marginal distributions of the input and output variables are skewed to the right, reflecting the skewness of the distribution of bank sizes measured by total assets.

As others (e.g., Wheelock and Wilson 2001, 2012, 2018) find, there is substantial collinearity among the specified inputs and outputs. We exploit this by reducing the dimensionality of our problem along the lines of Daraio and Simar (2007) and

Wilson (2018). First, we standardize each of our eight variables by dividing by their standard deviations (this amounts to merely changing units of measurement). Let n denote the sample size, and consider the $(n \times 3)$ matrix X whose columns consist of the observations X_{i1} , X_{i2} , and X_{i3} , and the $(n \times 2)$ matrix Y whose columns consist of the observations Y_{i2} and Y_{i3} , $i = 1, \dots, n$. For the moment matrix $X'X$, the ratio of its largest eigenvalue to the sum of its eigenvalues is 0.9609. Similarly, for the moment matrix $Y'Y$ the ratio of its largest eigenvalue to the sum of its eigenvalues is 0.9661.

These values suggest that most of the information in X and Y are contained in their first principle components, and the simulation results of Wilson (2018) suggest that in the absence of noise, DEA and FDH estimates of efficiency are likely to have smaller mean-square error using reduced-dimensional data as opposed to the full, eight-dimensional data in our problem. Consequently, we replace the first three inputs (i.e., purchased funds, labor and physical capital) with the first principle component $X_1^* = X\lambda_1$ of X , where λ_1 denotes the (3×1) eigenvector of $X'X$ corresponding to its largest eigenvalue. We similarly replace the second and third outputs (i.e., securities and off-balance sheet items) with their first principle component, denoted by Y_2^* . We retain the standardized versions of equity and nonperforming loans, denoted by X_2^* and X_3^* and the standardized version of total loans, denoted by Y_1^* . Thus, after dimension reduction, we have $p = 3$ inputs and $q = 2$ outputs. Summary statistics for these variables are also given in Table G.1.

We set the second and third elements of the direction vector to zero in order to hold equity and risk constant, and we set the first, fourth and fifth elements equal to the medians of X_{i1}^* , Y_{i1}^* and Y_{i2}^* , ensuring that our results are independent of units of measurement. We construct the $(r \times r)$ rotation matrix R_d (where $r = 5$ after reducing dimensionality) as described in Section 2.3, and use this to translate each observation (X_i^*, Y_i^*) to (Z_i, U_i) , where $X_i^* = [X_{i1}^* \ X_{i2}^* \ X_{i3}^*]$, $Y_i^* = [Y_{i1}^* \ Y_{i2}^*]$, Z_i is $1 \times (r - 1)$ and U_i is scalar. Summary statistics for the transformed data in (z, u) -space are given in Table G.1 appearing in the separate Appendix G, supplementary materials.

In addition to Z_i , we include in our regressions both time and firm effects. For estimation of $r_1(\cdot)$ in (3.1), we regress U_i on (Z_i, T_i, L_i) , $i = 1, \dots, n$, where $T_i \in \{1, 2, \dots, 80\}$ corresponding to 2001Q1, 2001Q2, ..., 2020Q4 (respectively), and L_i is a unique numeric label for the firm represented by observation i . Local linear regression requires an $(n \times n)$ diagonal matrix of kernel weights for each observation $i = 1, \dots, n$. We use a Gaussian product kernel for the four continuous elements of Z_i . For estimating $r_1(\cdot)$ at observation i , the j th diagonal element of our weighting matrix is given by

$$\psi_j = \left[\prod_{k=1}^4 h_k^{-1} K\left(\frac{Z_{ijk} - Z_{ik}}{h_k}\right) \right] h_5^{|T_j - T_i|} h_6^{I(L_j \neq L_i)} \quad (4.1)$$

where Z_{ik} denotes the k th element of Z_i , $K(\cdot)$ is the Gaussian kernel function, $I(\cdot)$ denotes the indicator function and h_k , $k \in \{1, \dots, 6\}$ are bandwidths with $h_k \in [0, 1]$ for $k \in \{5, 6\}$. The rate of convergence in the nonparametric regressions depends on the number of continuous right-hand side variables (four in our case), and is not affected by the discrete time and firm effects (see Li and Racine 2007 or Henderson and Parmeter 2015 for discussion).

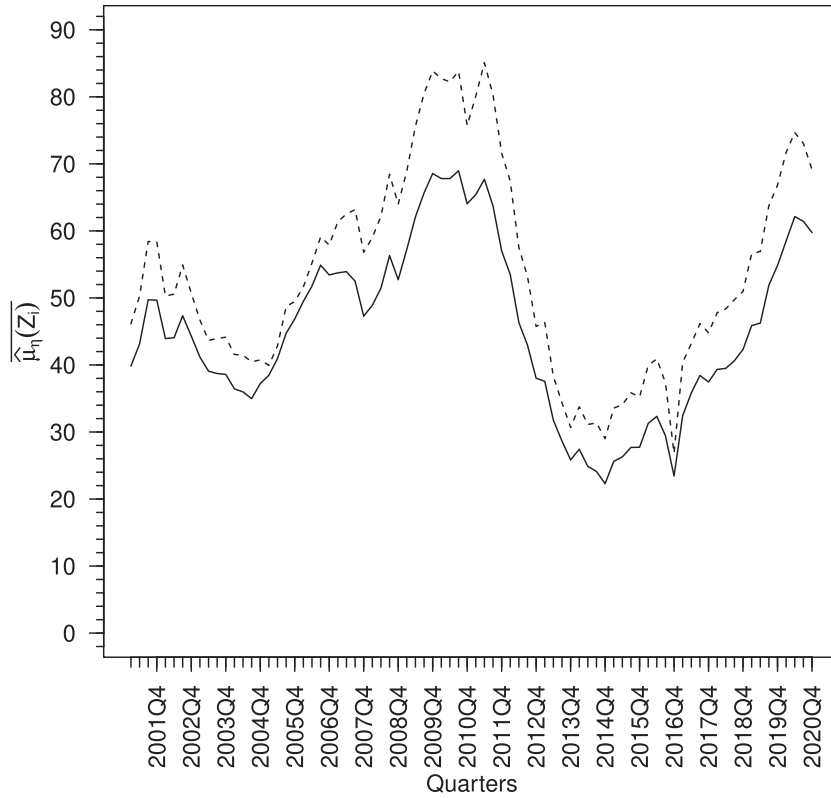


Figure 2. Sample Average and Median Values of $\hat{\mu}_\eta(Z_i)$, SVKZ Estimates, 2001Q1–2020Q4.
NOTE: Averages are indicated by the solid curve, while medians are represented by the dashed curve.

Bandwidths are optimized via leave-one-out cross-validation. To avoid possible numerical difficulties, we constrain the first four bandwidths (corresponding to elements of Z_i) by imposing a lower bound $0.1 (\hat{\sigma}_k n^{-1/(4+r^*)})$ on h_k , $k \in \{1, 2, 3, 4\}$ where $\hat{\sigma}_k$ is the sample standard deviation of Z_{ik} , $i = 1, \dots, n$ and $r^* = 4$ is the number of continuous variables used in each regression. In addition, we impose an upper bound on each h_k , $k \in \{1, 2, 3, 4\}$ given by $3 [\max_i(Z_{ik}) - \min_i(Z_{ik})]$. Table G.2 in the separate Appendix G, supplementary materials shows the optimized values of the bandwidths along with the corresponding lower and upper bounds. For each of the three estimated regressions, regression, the optimized bandwidths are well within the intervals between corresponding lower and upper bounds.

After regressing U_i on Z_i , and the squared as well as the cubed residuals from this regression on Z_i , we estimate $\hat{\sigma}_\eta(Z_i)$ and $\hat{\mu}_\eta(Z_i)$ for each observation in our sample. Figure 2 shows the (sample) average and median values of the estimates $\hat{\mu}_\eta(Z_i)$ across the 80 quarters represented in our sample. The estimates show an increase in inefficiency just before and during the recession of 2007–2009. Both average and median inefficiency are seen to decline after reaching a peak in 2011Q2, reaching a trough in 2014Q4. The estimates indicate that both mean and median inefficiency began to increase again starting in 2015, coincident with increases in the U.S. Federal Funds rate by the Federal Reserve System.

Specific values of $\hat{\mu}_\eta(Z_i)$ are difficult to interpret due to the fact that in the model, η_i is multiplied by the norm $\|d\|$ of the direction vector. The nonzero elements of the direction vector correspond to medians of X_1^* , Y_1^* , and Y_2^* , and hence

the direction vector is $d = [-0.01313 \quad 0.0 \quad 0.0 \quad 0.004934 \quad 0.002622]$ and its norm is $\|d\| = 0.01427$. Increasing (decreasing) the length of the direction vector necessarily decreases (increases) η_i and its conditional expectation. In any case, the values indicated by Figure 2 should be viewed while keeping in mind the length of the direction vector.

As a check on our results—as well as the ability of our methods to provide meaningful estimates—we also estimate marginal rates of substitution, marginal rates of transformation and marginal products in the space containing the untransformed, reduced-dimensional inputs X_j^* , $j \in \{1, 2, 3\}$ and outputs Y_k^* , $k \in \{1, 2\}$. In addition, we estimate scale elasticities with respect to the variable input X_1^* by computing

$$\hat{S}_i = \frac{\partial Y_{i1}^*}{\partial X_{i1}^*} \cdot \frac{X_{i1}^*}{Y_{i1}^*} + \frac{\partial Y_{i2}^*}{\partial X_{i1}^*} \cdot \frac{X_{i1}^*}{Y_{i2}^*} \quad (4.2)$$

for each observations $i = 1, \dots, n$. All derivatives are estimated using (2.29) as described in Section 2.4. In order to conserve space, we report summary statistics for these estimates in Table G.3 in the separate Appendix G, supplementary materials. Interestingly, all of the estimates of scale elasticity in Table G.3 are greater than 1, suggesting that all banks in the sample face increasing returns to scale. This is consistent with results obtained by Wheelock and Wilson (2018) and may explain why even the largest U.S. banks continue to grow larger.

It is important to note that the derivative estimates are for a particular point on the frontier, and that our use of local estimation methods is likely to result in noisy estimates at each point. Nonetheless, the results in Table G.3 reveal that among the

486,461 observations in our sample, the median and mean values of estimates of (i) the marginal rate of substitution between X_1^* and X_2^* , (ii) the marginal products with respect to X_1^* and X_2^* , (iii) the marginal rate of transformation between Y_1^* and Y_2^* and (iv) scale elasticity have the expected signs. One should expect marginal rates of substitution and transformation to be negative, and marginal products and scale efficiency to be positive. The right-most column in Table G.3 gives the percentage of observations yielding negative estimates, and with the exception of derivatives involving X_3^* , these range from 86.1% to 99.8%.

The results for marginal rates of substitution between X_3^* and the other inputs, as well as for the marginal products of Y_1^* and Y_2^* with respect to X_3^* are surprising, with most observations—between 0.8% and 14.6%—yielding unexpected signs. While we initially viewed risk as measured by nonperforming loans, the results for derivatives involving X_3^* in Table G.3 suggest that this variable behaves as an output. However, recalling the discussion above, we treat this variable as fixed, and the corresponding direction is set to zero in the direction vector d . Consequently, as noted above, X_3^* is allowed to act as a “netput,” that is, either as an input or an output, and the evidence we obtain suggests that it is better viewed as an undesirable output. Our use of a zero direction for this variable means that our initial view of the variable as a quasi-fixed input has no bearing on the estimation when we change our view of the variable as an undesirable output. Treating the variable as a fixed, undesirable output merely amounts to changing the label on the variable from X_3^* to Y_3^* .

We examine potential heteroscedasticity in the inefficiency process by estimating derivatives $\partial \mu_\eta(W_i^\partial)/\partial W_i^\partial$ given by (2.32). Table G.4 in the separate Appendix G, supplementary materials gives the percentages of observations where derivatives of the conditional expectation $\mu_\eta(\cdot)$ with respect to X_j^* , $j \in \{1, 2, 3\}$ and Y_k^* , $k \in \{1, 2\}$ that are less than, equal to or greater than zero. The estimated derivatives equal zero in about 39.7% of all cases, but otherwise show varying proportions that are less than or greater than zero. The results in Table G.4 provide evidence of heteroscedasticity in the distribution of the one-sided inefficiency process. Our model explicitly allows for this in Assumption 2.4. There are many published examples where parametric, stochastic frontier models specified with a fixed, constant shape parameter for the one-sided inefficiency distribution; Few studies that estimate parametric stochastic frontier models allow for or test for heteroscedasticity in the inefficiency distribution.

As a final note on our application, $\hat{\sigma}_\epsilon(Z_i) = 0$ for 307,456 observations, or for about 63.2% of the sample. Thus, for almost two-thirds of the sample, we find no evidence of noise. As discussed in the separate Appendix C, Section C.1, supplementary materials, one could impose observation-specific bounds when estimating $r_2(x)$ in order to ensure $\hat{\sigma}_\epsilon^2(z)$ is greater than zero. But, as also noted in Section C.1, this is computationally prohibitive given our large sample size. As noted in separate Appendix G.2, supplementary materials, using the HMS estimators results in 485,664 cases (99.8% of the sample) where $\hat{\sigma}_\epsilon(Z_i) = 0$. Given that we use accounting data on banks, and that the data are audited both by the banks as well as regulators, it may not be unreasonable to find little evidence of noise.

5. Summary and Conclusions

The methods proposed in this article provide a very flexible, almost fully nonparametric approach for evaluating the performance of firms in terms of their efficiency. For purposes of estimating the production frontier, we need only some mild assumptions on continuity and smoothness of the frontier, symmetry of the density of the noise process, and tail behavior of the noise and inefficiency distributions. No functional form assumptions are needed—not for the frontier, nor for the distributions of noise and inefficiency. In addition, our method accommodates both multiple inputs as well as multiple outputs while avoiding endogeneity problems. As discussed in Section 1, Other flexible methods that have been proposed to date either do not allow for noise, do not permit multiple inputs and outputs, involve significant computational difficulties, or involve endogeneity problems. Of course, as with most nonparametric methods, ours incurs the curse of dimensionality, but as illustrated in Section 4, dimension reduction methods can mitigate this curse in many cases. Our empirical application on U.S. commercial banks demonstrates that our method is computationally tractable with real data.

There have been for many years two disparate strands in the literature on frontier efficiency estimation. One strand consists of articles describing deterministic estimators such as DEA and FDH estimators that require no functional form assumptions, and allow for multiple inputs as well as multiple outputs, but which do not allow for noise. The other strand consists of articles that are often fully parametric, or in recent years, partially parametric, that permit noise, but which require functional form assumptions and do not allow multiple inputs and outputs. Our article provides a bridge between these two worlds.

Supplementary Materials

The Supplementary Appendices give additional details on the flexibility of the stochastic specification, skewness in the “wrong” direction and alternative approaches for dealing with it, and alternative approaches for estimation. Some extensions are also discussed, including extensions to radial distances, introduction of exogenous environmental variables and stochastic versions of FDH/DEA-type estimators. The Supplementary Appendices also provide Monte Carlo evidence as well as additional results on the Efficiency of U.S. Banks.

Acknowledgments

The views expressed in this article are those of the authors and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Any remaining errors are solely our responsibility.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

We are grateful for (i) the Palmetto cluster operated by the Cyber Infrastructure Technology Integration group at Clemson University, and (ii) high-performance computing equipment purchased with funds from U.S. National Science Foundation MRI award no. 1725573.

ORCID

Paul W. Wilson  <http://orcid.org/0000-0002-9865-033X>
 Léopold Simar  <http://orcid.org/0000-0003-0791-8490>

References

- Aigner, D., Lovell, C. A. K., and Schmidt, P. (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics*, 6, 21–37. [1,4]
- Akhavein, J. D., Berger, A. N., and Humphrey, D. B. (1997), "The Effects of Megamergers on Efficiency and Prices: Evidence from a Bank Profit Function," *Review of Industrial Organization*, 12, 95–139. [9]
- Battese, G. E., and Corra, G. S. (1977), "Estimation of a Production Frontier Model: With Application to the Pastoral Zone Off Eastern Australia," *Australian Journal of Agricultural Economics*, 21, 169–179. [1,4]
- Berger, A. N., and Mester, L. J. (1997), "Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?" *Journal of Banking and Finance*, 21, 895–947. [9]
- Chambers, R. G., Chung, Y., and Färe, R. (1996), "Benefit and Distance Functions," *Journal of Economic Theory*, 70, 407–419. [1,3]
- (1998), "Profit, Directional Distance Functions, and Nerlovian Efficiency," *Journal of Optimization Theory and Applications*, 98, 351–364. [1,3]
- Coelli, T. (1995), "Estimators and Hypothesis Tests for a Stochastic Frontier Function: A Monte Carlo Analysis," *Journal of Productivity Analysis*, 6, 247–268. [1]
- Coelli, T., and Perelman, S. (1996), "Efficiency Measurement, Multiple-Output Technologies and Distance Functions: With Application to European Railways," CREPP working paper 96/05, University of Liège, Belgium. [2]
- Cuesta, R. A., and Orea, L. (2002), "Mergers and Technical Efficiency in Spanish Savings Banks: A Stochastic Distance Function Approach," *Journal of Banking and Finance*, 26, 2231–2247. [2]
- Daouia, A., Florens, J. P., and Simar, L. (2020), "Robust Frontier Estimation from Noisy Data: A Tikhonov Regularization Approach," *Econometrics and Statistics*, 14, 1–23. [2]
- Daraio, C., and Simar, L. (2007), *Advanced Robust and Nonparametric Methods in Efficiency Analysis*, New York: Springer. [9]
- Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, London: Chapman and Hall. [7]
- Fan, Y., Li, Q., and Weersink, A. (1996), "Semiparametric Estimation of Stochastic Production Frontier Models," *Journal of Business and Economic Statistics*, 14, 460–468. [1]
- Färe, R., Grosskopf, S., and Margaritis, D. (2008), "Productivity and Efficiency: Malmquist and More," in *The Measurement of Productive Efficiency* (2nd ed.), eds. H. Fried, C. A. K. Lovell, and S. Schmidt, pp. 522–621, Oxford: Oxford University Press. [3]
- Florens, J. P., Simar, L., and Van Keilegom, I. (2020), "Estimation of the Boundary of a Variable Observed with Symmetric Error," *Journal of the American Statistical Association*, 115, 425–441. [2,7]
- Hafner, C. M., Manner, H., and Simar, L. (2018), "The 'Wrong Skewness' Problems in Stochastic Frontier Models: A New Approach," *Econometric Reviews*, 37, 380–400. [1]
- Hall, P., and Simar, L. (2002), "Estimating a Change-point, Boundary or Frontier in the Presence of Observation Error," *Journal of the American Statistical Association*, 97, 523–534. [1,2]
- Henderson, D. J., and Parmeter, C. F. (2015), *Applied Nonparametric Econometrics*, New York: Cambridge University Press. [9]
- Jondrow, J., Lovell, C. A. K., Materov, I. S., and Schmidt, P. (1982), "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Model," *Journal of Econometrics*, 19, 233–238. [8]
- Kneip, A., Simar, L., and Van Keilegom, I. (2015), "Frontier Estimation in the Presence of Measurement Error with Unknown Variance," *Journal of Econometrics*, 184, 379–393. [2]
- Kneip, A., Simar, L., and Wilson, P. W. (2008), "Asymptotics and Consistent Bootstraps for DEA Estimators in Non-parametric Frontier Models," *Econometric Theory*, 24, 1663–1697. [3]
- Kumbhakar, S. C., Park, B. U., Simar, L., and Tsionas, E. G. (2007), "Non-parametric Stochastic Frontiers: A Local Likelihood Approach," *Journal of Econometrics*, 137, 1–27. [2,4]
- Kumbhakar, S. C., Parmeter, C. F., and Zelenyuk, V. (2020), "Stochastic Frontier Analysis: Foundations and Advances i," in *Handbook of Production Economics*, eds. S. C. Ray, R. Chambers, and S. Kumbhakar, pp. 1–40, Singapore: Springer. [1]
- Kuosmanen, T., and Johnson, A. (2017), "Modeling Joint Production of Multiple Outputs in StoNED: Directional Distance Function Approach," *European Journal of Operational Research*, 262, 792–801. [2]
- Kuosmanen, T., and Kortelainen, M. (2012), "Stochastic Non-smooth Envelopment of Data: Semi-parametric Frontier Estimation Subject to Shape Constraints," *Journal of Productivity Analysis*, 38, 11–28. [2]
- Lang, S. (1964), *A Second Course in Calculus*, Reading, MA: Addison-Wesley Publishing Company, Inc. [6]
- Li, Q., and Racine, J. (2007), *Nonparametric Econometrics*, Princeton, NJ: Princeton University Press. [7,8,9]
- Lovell, C. A. K. (1993), "Production Frontiers and Productive Efficiency," in *The Measurement of Productive Efficiency: Techniques and Applications*, eds. H. Fried, C. A. K. Lovell, and S. Schmidt, pp. 3–67, Oxford: Oxford University Press, Inc. [7]
- Malikov, E., Kumbhakar, S. C., and Tsionas, M. G. (2016), "A Cost System Approach to the Stochastic Directional Technology Distance Functions with Undesirable Outputs: The Case of US Banks in 2011–2010," *Journal of Applied Econometrics*, 31, 1407–1429. [2,9]
- Masry, E. (1996), "Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates," *Journal of Time Series Analysis*, 17, 571–599. [7]
- Meeusen, W., and van den Broeck, J. (1977), "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error," *International Economic Review*, 18, 435–444. [1,4]
- Olson, J. A., Schmidt, P., and Waldman, D. M. (1980), "A Monte Carlo Study of Estimators of Stochastic Frontier Production Functions," *Journal of Econometrics*, 13, 67–82. [8]
- Park, B. U., Simar, L., and Zelenyuk, V. (2015), "Categorical Data in Local Maximum Likelihood: Theory and Application to Productivity Analysis," *Journal of Productivity Analysis*, 43, 199–214. [2]
- Parmeter, C. F., and Zelenyuk, V. (2019), "Combining the Virtues of Stochastic Frontier and Data Envelopment Analysis," *Operations Research*, 67, 1628–1658. [1]
- Prokhorov, A., Tran, K. C., and Tsionas, M. G. (2021), "Estimation of Semi- and Nonparametric Stochastic Frontier Models with Endogenous Regressors," *Empirical Economics*, 60, 3043–3068. [2]
- Richmond, J. (1974), "Estimating the Efficiency of Production," *International Economic Review*, 15, 515–521. [7]
- Serra, T., Lansink, A. O., and Stefanou, S. E. (2011), "Measurement of Dynamic Efficiency: A Directional Distance Function Parametric Approach," *American Journal of Agricultural Economics*, 93, 756–767. [2]
- Sickles, R. C., and Zelenyuk, V. (2019), *Measurement of Productivity and Efficiency*, Cambridge, UK: Cambridge University Press. [1]
- Simar, L. (2007), "How to Improve the Performances of DEA/FDH Estimators in the Presence of Noise," *Journal of Productivity Analysis*, 28, 183–201. [2]
- Simar, L., Van Keilegom, I., and Zelenyuk, V. (2017), "Nonparametric least squares methods for stochastic frontier models," *Journal of Productivity Analysis*, 47, 189–204. [1,4]
- Simar, L., and Wilson, P. W. (2010), "Estimation and Inference in Cross-Sectional, Stochastic Frontier Models," *Econometric Reviews*, 29, 62–98. [8]
- (2013), "Estimation and Inference in Nonparametric Frontier Models: Recent Developments and Perspectives," *Foundations and Trends in Econometrics*, 5, 183–337. [3]
- (2015), "Statistical Approaches for Non-parametric Frontier Models: A Guided Tour," *International Statistical Review*, 83, 77–110. [3]
- Simar, L., and Zelenyuk, V. (2011), "Stochastic FDH/DEA Estimators for Frontier Analysis," *Journal of Productivity Analysis*, 36, 1–20. [2]

- Simper, R., Hall, M. J. B., Liu, W., Zelenyuk, V., and Zhou, Z. (2017), "How Relevant is the Choice of Risk Management Control Variable to Non-parametric Bank Profit Efficiency Analysis? The Case of South Korean Banks," *Annals of Operations Research*, 250, 105–127. [9]
- Sun, K., and Salim, R. (2020), "A Semiparametric Stochastic Input Distance Frontier Model with Application to the Indonesian Banking Industry," *Journal of Productivity Analysis*, 54, 139–156. [2]
- Wheat, P., Greene, W., and Smith, A. (2014), "Understanding Prediction Intervals for Firm Specific Inefficiency Scores from Parametric Stochastic Frontier Models," *Journal of Productivity Analysis*, 42, 55–65. [8]
- Wheelock, D. C., and Wilson, P. W. (2001), "New Evidence on Returns to Scale and Product Mix Among U.S. Commercial Banks," *Journal of Monetary Economics*, 47, 653–674. [9]
- (2012), "Do Large Banks have Lower Costs? New Estimates of Returns to Scale for U.S. Banks," *Journal of Money, Credit, and Banking*, 44, 171–199. [9]
- (2018), "The Evolution of Scale-Economies in U.S. Banking," *Journal of Applied Econometrics*, 33, 16–28. [9,10]
- Wilson, P. W. (2018), "Dimension Reduction in Nonparametric Models of Production," *European Journal of Operational Research*, 267, 349–367. [9]