

# “When, Where and How” of Efficiency Estimation: Improved procedures

Mike G. Tsionas\*

May 23, 2016

## Abstract

The issues of functional form, distributions of the error components and endogeneity are for the most part still open in stochastic frontier models. The same is true when it comes to imposition of restrictions of monotonicity and curvature, making efficiency estimation an elusive goal. In this paper we attempt to consider these problems simultaneously and offer practical solutions to the problems raised by Stone (2002) and addressed in Badunenko, Henderson and Kumbhakar (2012). We provide major extensions to smoothly mixing regressions and fractional polynomial approximations for both the functional form of the frontier and the structure of inefficiency. Endogeneity is handled, simultaneously, using copulas. We provide detailed computational experiments and an application to US banks. To explore the posteriors of the new models we rely heavily on Sequential Monte Carlo techniques.

**Key words:** Stochastic Frontiers; Smoothly Mixing Regressions; Fractional Polynomial Approximations; Efficiency Estimation; Bayesian inference; Sequential Monte Carlo.

---

\*Lancaster University Management School, LA1 4YX, UK, m.tsionas@lancaster.ac.uk & Athens University of Economics and Business, 76 Patission Str., Athens 10434, Greece.

# 1 Introduction

Stone (2002) has raised several important points in connection to efficiency estimation, including the choice of functional forms, the distributions of statistical error term and the one-sided error term that represents inefficiency etc. Badunenko, Henderson and Kumbhakar (2012, BHK henceforth) used simulations to compare the performance of (i) a two-stage semi-parametric stochastic frontier (SSF) estimator due to Fan et al. (1996), and (ii) the non-parametric bias-corrected DEA estimator of Kneip et al. (2008).

One important outcome from their simulation studies is that, in realistic situations, “both estimators do a fair job at identifying the best and the worst performing decision making units. However, each does a relatively poor job at identifying the median performer.” Another important conclusion for policy making was the following:

“For a practical example, we note that [...] identifying a benchmark firm(s) is often important to regulators. A benchmark firm(s) (say top 5%) is often used to decide the penalty (or carrot) for the bottom firms in yardstick competition. In such a case, it is important to accurately estimate the top 5% and bottom 5% of firms. We note that identification of both the top 5% and bottom 5% of firms is feasible in scenarios s1, s2 and s4. In general, FLW would produce more reliable results for the bottom 5% (except for s2) and KSW would generally produce more reliable results for the upper 5%.”<sup>1</sup>

Of course, BHK used an SSF alternative to the bootstrapped DEA approach. This partly solves the functional form assumption although not in a comprehensive manner. There are two problems: (i) With many inputs and small samples, the estimator would not be reliable. (ii) The estimator still uses a normal-half-normal distribution in the second stage to estimate technical efficiencies. (iii) There are other problems with actual data sets that neither Stone (2002) nor BHK considered. (iv) The recommendation that one estimator is best for the upper 5% and another for the bottom 5% of firms is, at best, problematic as the overall statistical properties are hard to analyze.

The questions that Stone (2002) raised and which were partly answered by the simulation studies in BHK are still at large. Functional forms and error term distributions in stochastic frontier analysis are still problems that we need to consider jointly. Another problem that plagues applications is the potential endogeneity of inputs: Inputs are not experimentally designed nor they are exogenous to production; they are decided by firms. Nothing really implies that outputs are exogenous and inputs are endogenous. In a cost minimization framework, for example, the firm decides the quantities of inputs, given input prices and outputs. Under this behavioral consideration, inputs would be endogenous and outputs would be exogenous. Many public firms would belong to this category. Under

---

<sup>1</sup>Here, s1 to s4 refer to different scenarios in their simulations. Suppose  $\sigma_v^2$  and  $\sigma_u^2$  denote the variances of two sided error term ( $v$ ) and one sided error term ( $u$ ) respectively and  $\lambda = \sigma_u/\sigma_v$ . “In scenario s1 ( $\sigma_v = \sigma_u = 0.01, \lambda = 1.0$ ), both terms are relatively small. In other words, the data are measured with relatively little error and the units are relatively efficient. In scenario s2 ( $\sigma_v = 0.01$  and  $\sigma_u = 0.05, \lambda = 5.0$ ), the data have relatively little noise, but the units under consideration are relatively inefficient. In scenario s3 ( $\sigma_v = 0.05$  and  $\sigma_u = 0.01, \lambda = 0.2$ ), the data are relatively noisy and the units are relatively efficient. The fourth scenario s4 ( $\sigma_v = \sigma_u = 0.05, \lambda = 1.0$ ) is redundant as  $\lambda = 1.0$  as in s1. However, we show this case to emphasize that the results of the experiment depend upon the ratio of  $\sigma_u$  to  $\sigma_v$  and not their absolute values.”

profit maximization, both inputs and outputs are endogenous as they are both subject to choice by the firm. The endogeneity problem has proven to be quite hard to attack. Measurement error in inputs is also a problem and, for practical purposes, it means that observed inputs are correlated with the error term(s) so we have an endogeneity problem.

In stochastic frontier models the basic model is  $y = f(\mathbf{x}) \exp(v - u)$  where  $y$  is output,  $\mathbf{x}$  is a vector of inputs,  $v$  is statistical noise usually assumed to be normally distributed,  $\mathcal{N}(0, \sigma_v^2)$ , and  $u$  is a positive error term usually assumed to follow a half-normal distribution<sup>2</sup>,  $u \sim \mathcal{N}_+(0, \sigma_u^2)$ . The ratio  $\lambda = \frac{\sigma_u}{\sigma_v}$  is known as signal-to-noise ratio and indicates how important is inefficiency ( $u$ ) relative to noise ( $v$ ). The simulation study in BHK concluded that with a low value of  $\lambda$  it does not make much sense to proceed and measure inefficiency as this is inherently hard in this situation. When  $\lambda \geq 1$ , both methods they used turned out to be, for the most part at least, similar in performance subject to the considerations stated above.

When the value of  $\lambda$  is small, the data are not very informative about the presence of an inefficiency signal. However, this could be attributed, at least partially, to the potential correlation between  $\mathbf{x}$  and  $v$  and / or  $u$ . Ignoring this may seriously bias efficiency scores while formally accounting for it, it can lead to better separation of inefficiency from noise.

The distributional assumptions problems seem to be resolved only in the context of bootstrapped-DEA as this approach does not need any such particular assumptions. In the two-stage SSF estimator due to Fan et al. (1996), the first stage estimates a term that includes  $f(\mathbf{x})$  and thus it dispenses away with the functional form assumption (at least, on the surface) but it does assume a normal-half-normal specification in the second stage to estimate technical inefficiencies. From that point of view the approach is not completely satisfactory. Another problem is that monotonicity and concavity are not imposed (e.g. Pu, Parmeter and Racine, 2013) on  $f(\mathbf{x})$  and, therefore, it is not clear that a production function is estimated indeed. The bootstrapped-DEA method, again, seems to have an advantage here.

Why not use the bootstrapped-DEA method then? BHK have shown that adopting this practice indiscriminately may lead to serious problems as, generally, the semi-parametric stochastic frontier estimator seems to perform better! Moreover, under conditions of endogeneity, measurement errors etc, bootstrapped-DEA cannot offer a panacea. The semi-parametric stochastic frontier estimator is, clearly, not a panacea either in this situation. Another consideration is that with data sets running in the thousands of observations, applying the the bootstrapped-DEA method is extremely computationally intensive whereas applying the SSF estimator is much easier. Therefore, we believe that *there are reasons to stick with stochastic frontiers but functional form issues, issues related to distributional assumptions and endogeneity issues must be resolved simultaneously, under the same umbrella of a common, unifying flexible stochastic frontier model*. We should, nevertheless, mention that there have important developments in DEA. These are summarized in Daraio and Simar (2007). Two-stage models that allow for explanatory variables

---

<sup>2</sup>Many other distributions have been used in the literature like the exponential, the *gamma*, the Weibull etc.

are treated fully in Simar and Wilson (2007). See also Daraio, Bonaccorsi and Simar (2015) and Daraio and Simar (2014) who have used a directional distance function with estimated input - output - specific directions, see also Simar and Vanhems (2012) and Simar, Vanhems and Wilson (2012).

## 2 Model

### 2.1 The approach in Badunenko, Henderson and Kumbhakar (2012)

Badunenko, Henderson and Kumbhakar (2012) compare two promising estimators of technical efficiency in the cross-sectional case. They compare the non-parametric kernel estimator of Fan et al. (1996) with the non-parametric bias-corrected DEA estimator of Kneip et al. (2008).

The model of Fan et al. (1996) requires a parametric second stage (and hence it is semiparametric), it is more robust than the initial stochastic frontier model by Aigner et al. (1977) and Meeusen and van den Broeck (1977)<sup>3</sup>. The method of Kneip et al. (2008) introduces statistical inference via bootstrapping has been shown to complement well the standard DEA model found in Charnes et al. (1978).

Suppose  $\mathbf{x}_i \in \mathfrak{R}^p$  is a vector of inputs and  $y_i$  denotes output for firm  $i = 1, \dots, n$ . In the model of Kneip et al. (2008) we have:

$$y_i = f(\mathbf{x}_i) + v_i - u_i, \quad i = 1, \dots, n, \quad (1)$$

where  $f : \mathfrak{R}^p \rightarrow \mathfrak{R}$  is an unknown smooth estimation,  $v_i$  is a two-sided error term and  $u_i \geq 0$  is a one-sided error representing inefficiency. As  $E(y_i|\mathbf{x}_i) = f(\mathbf{x}_i) - E(u_i) \neq f(\mathbf{x}_i)$ , a non-parametric estimator of the regression function  $E(y_i|\mathbf{x}_i)$  is not  $f(\mathbf{x}_i)$  itself. It becomes necessary to adopt distributional assumptions on both  $v$  and  $u$  to construct the log-likelihood function to be maximized to obtain the efficiency scores. Under the assumption that  $v_i|\mathbf{x}_i \sim N(0, \sigma_v^2)$  and, independently,  $u_i|\mathbf{x}_i \sim N_+(0, \sigma_u^2)$ , the log-likelihood function is:

$$l(\lambda) = -n \ln \hat{\sigma} + \sum_{i=1}^n \ln \Phi \left( -\frac{\lambda \hat{\varepsilon}_i}{\hat{\sigma}} \right) - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n \hat{\varepsilon}_i^2, \quad (2)$$

where  $\hat{\varepsilon}_i = y_i - \hat{E}(y_i|\mathbf{x}_i) - \mu(\lambda, \hat{\sigma})$ ,  $\mu(\lambda, \hat{\sigma}) = \lambda \sigma \sqrt{2} \{ \pi(1 + \lambda^2)^{1/2} \}$ , and

$$\hat{\sigma}^2 = \frac{n^{-1} \sum_{i=1}^n \{y_i - \hat{E}(y_i|\mathbf{x}_i)\}^2}{\{1 - \frac{2\lambda^2}{\pi(1+\lambda^2)}\}}. \quad (3)$$

The log-likelihood function in (2) is a function of the single parameter  $\lambda = \frac{\sigma_u}{\sigma_v}$  given the estimator  $\hat{\sigma}^2$  of  $\sigma^2 = \sigma_v^2 + \sigma_u^2$ . Given an estimate  $\hat{\lambda}$  the estimate of  $\hat{\sigma}^2$  can be recovered from (3). The point estimator of inefficiency via Jondrow et al. (1982) can be obtained as:

$$\hat{u}_i = \hat{\mu}_i^* + \hat{\sigma}^* \frac{\phi(\hat{\mu}_i^*/\hat{\sigma}^*)}{\Phi(\hat{\mu}_i^*/\hat{\sigma}^*)}, \quad i = 1, \dots, n, \quad (4)$$

---

<sup>3</sup>For a review, see Kumbhakar and Lovell (2003).

where  $\hat{\mu}_i^* = -\hat{\varepsilon}_i \hat{\sigma}_u^2 / (\hat{\sigma}_v^2 + \hat{\sigma}_u^2)$ ,  $\hat{\sigma}^* = \hat{\sigma}_v \hat{\sigma}_u / (\hat{\sigma}_v^2 + \hat{\sigma}_u^2)^{1/2}$ , and  $\phi$ ,  $\Phi$  denote the standard normal pdf and cdf respectively. Badunenko, Henderson and Kumbhakar (2012) use a local linear estimator to formulate the log-likelihood in (1).

## 2.2 In search of a new model

In this paper we are not concerned with DEA. The challenges in stochastic frontier modeling are the following:

i) To *estimate the unknown function*  $f(\mathbf{x})$  in (1). The function must satisfy certain theoretical properties. It must be increasing and concave for all  $\mathbf{x} \in \mathbb{R}_+^n$ .

ii) To *dispense with the distributional assumptions* on  $v_i$  and  $u_i$ . In this context it would be useful if the distribution of  $u_i$  depends flexibly on a given vector of predetermined environmental variables  $\mathbf{z}_i \in \mathbb{R}^s$ . This vector may include  $\mathbf{x}_i$  but in this case the following point applies more forcefully.

iii.1) To *account for the endogeneity of the regressors*  $\mathbf{x}_i$ . We should mention that what is endogenous depends on behavioral assumptions. Under cost minimization  $\mathbf{x}$  is endogenous but  $y$  is not. Under profit maximization both are endogenous. Under revenue maximization only  $y$  is endogenous.

iii.2) If prices are available, the *endogeneity problem* can be solved by appending the first-order-conditions from the cost minimization, profit maximization or revenue maximization problem. If, as is most common, prices are not available, endogeneity must be addressed in a different way.

## 3 New models

### 3.1 Functional form

There is considerable evidence that mixture-of-normals models tend to perform substantially better compared to other alternatives such as the Dirichlet process prior. Geweke and Kean (2007) propose the smoothly mixing regression (SMR):

$$p(y|\mathbf{x}) = \sum_{g=1}^G \omega_g(\mathbf{x}) f_N(\beta'_g \mathbf{x}, \sigma_g^2), \quad (5)$$

where  $\mathbf{x} \in \mathbb{R}^k$  is a multivariate covariate,  $f_N(\mu, \sigma^2)$  denotes the density of a  $N(\mu, \sigma^2)$  distribution and the weights satisfy  $\omega_g(\mathbf{x}) \geq 0$ ,  $\sum_{g=1}^G \omega_g(\mathbf{x}) = 1$ . Villani, Kohn and Nott (2012) and Villani, Kohn and Giordani (2009) extend the model as follows:

$$p(y|\mathbf{x}) = \sum_{g=1}^G \omega_g(\mathbf{x}) f_N(\beta'_g v_g(\mathbf{x}), \exp(\delta'_g w_g(\mathbf{x}))). \quad (6)$$

Typically, it is assumed that:

$$\omega_g(\mathbf{x}) = \frac{\exp(\gamma'_g \mathbf{x})}{\sum_{s=1}^G \exp(\gamma'_s \mathbf{x})}, \quad (7)$$

where  $v_g(\mathbf{x})$  and  $w_g(\mathbf{x})$  are transformations of the regressors, such as splines.

Geweke and Petrella (2012) introduce an alternative set of fractional polynomial approximations and show that the set of fractional polynomial approximations is dense on a Sobolev space of functions on a compact set.<sup>4</sup> Moreover, imposing regularity conditions directly on the fractional polynomials produces pseudo-true approximations that converge rapidly to production functions having no exact representation as fractional polynomials. Fractional polynomial approximations (with  $k$  terms) are defined as:

$$f(x_1, \dots, x_n) = \sum_{i=1}^k a_i \prod_{l=1}^n x_l^{s_{il}}, \quad (8)$$

which is assumed positive on  $\mathfrak{R}_{++}^n$ . Such fractional polynomial approximations can approximate the unknown function and its first two derivatives on a closed compact subset of  $\mathfrak{R}_{++}^n$ . In compact notation fractional polynomials are denoted by:

$$p(\mathbf{x}; \mathbf{a}) = \sum_{i=1}^k a_i \mathbf{x}^{\mathbf{j}^{(i)}}, \quad (9)$$

where  $\mathbf{x} = [x_1, \dots, x_n]'$ ,  $\mathbf{j}^{(i)} = [j_1^{(i)}, \dots, j_n^{(i)}]'$  is a sequence of multi-indices. Therefore,  $\mathbf{x}^{\mathbf{j}^{(i)}}$  is a ‘‘Cobb-Douglas-type’’ function whose exponents are given by the multi-index  $\mathbf{j}^{(i)}$ . A twice differentiable function  $f(\mathbf{x})$  is called strictly regular on  $C \subseteq \mathfrak{R}_{++}^n$  if  $f(\mathbf{x}) > 0$ ,  $\partial f(\mathbf{x})/\partial \mathbf{x} > 0$  and  $\partial^2 f(\mathbf{x})/\partial \mathbf{x} \partial \mathbf{x}'$  is negative definite, for all  $\mathbf{x} \in C$ . For strictly regular functions, Geweke and Petrella (2012) have proved that fractional polynomials can approximate to any degree any function and its first two derivatives.

Suppose we have the data  $\mathbf{x}_t \in \mathfrak{R}^n$ ,  $t = 1, \dots, T$ . The fractional approximation can be written as:

$$p(\mathbf{x}_t; \mathbf{a}) = \sum_{i=1}^k a_i \prod_{j=1}^n x_{tj}^{s_{ji}}, \quad (10)$$

with  $s_{ji} \equiv s_j^{(i)}$ . We can represent this function more compactly as:

$$p(\mathbf{x}_t; \mathbf{a}) = \begin{bmatrix} x_{t1}^{s_{11}} & x_{t2}^{s_{21}} & \dots & x_{tn}^{s_{n1}} \\ x_{t1}^{s_{12}} & x_{t2}^{s_{22}} & \dots & x_{tn}^{s_{n2}} \\ & \dots & \dots & \\ x_{t1}^{s_{1k}} & x_{t2}^{s_{2k}} & \dots & x_{tn}^{s_{nk}} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix} = \mathbf{z}'_t \mathbf{a}. \quad (11)$$

Define the  $T \times k$  matrix  $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_T]'$  so that

$$[p(\mathbf{x}_1; \mathbf{a}), p(\mathbf{x}_2; \mathbf{a}), \dots, p(\mathbf{x}_T; \mathbf{a})]' = \mathbf{Z} \mathbf{a}, \quad (12)$$

---

<sup>4</sup>Fractional polynomials are also used in Sauerbrei and Royston (1999).

with the typical element being  $z_{ti} = \prod_{l=1}^n x_{tl}^{s_{tl}}$ . It can be shown that the derivatives are given as follows:

$$\partial p(\mathbf{x}_t; \mathbf{a}) / \partial \mathbf{x}_t = \mathbf{W}'_t \mathbf{a} \quad (13)$$

where  $\mathbf{W}_t = [w_{tij}]$ ,  $w_{tij} = s_{ji} x_{tj}^{-1} z_{ti}$ . The  $n \times n$  Hessian matrix is:

$$\partial^2 p(\mathbf{x}_t; \mathbf{a}) / \partial \mathbf{x}_t \partial \mathbf{x}'_t = \sum_{i=1}^k a_i \mathbf{C}_{ti} = (\mathbf{a}' \otimes \mathbf{I}_n) \begin{bmatrix} \mathbf{C}_{t1} \\ \vdots \\ \mathbf{C}_{tk} \end{bmatrix}, \quad (14)$$

where  $\mathbf{C}_{ti} = [c_{tijr}]$ ,  $c_{tijr} = x_{tj}^{-1} x_{tr}^{-1} s_{ji} s_{ri} z_{ti} - \delta_{jr} s_{ji} x_{tj}^{-2} z_{ti}$  and  $\delta_{jr}$  represents the Kronecker  $\delta$ . If we define the  $nk \times nT$  matrix  $\mathbf{C} = [\mathbf{C}_{ti}, t = 1, \dots, T, i = 1, \dots, k]$  then we have:

$$[\nabla^2 p(\mathbf{x}; \mathbf{a}), \dots, \nabla^2 p(\mathbf{x}_T; \mathbf{a})] = (\mathbf{a}' \otimes \mathbf{I}_n) \mathbf{C}. \quad (15)$$

Positivity and monotonicity correspond to the restrictions:

$$\mathbf{Za} > \mathbf{O}, \mathbf{Wa} > \mathbf{O}. \quad (16)$$

Strict concavity can be checked using the eigenvalues of matrices  $\mathbf{C}_{ti}$  which are nonlinear functions of  $\mathbf{a}$ . Geweke and Petrella (2012) advocate checking the conditions at a number of points in  $G$  rather than checking or enforcing them at all data points.

With base  $b = \frac{1}{2}$  a cost function, for example, in three input prices, can be represented as

$$a_1 + a_2 x_1 + a_3 x_1^{1/2} x_2^{1/2} + a_4 x_1^{1/2} x_3^{1/2} + a_5 x_1^{1/2} + a_6 x_2 + a_7 x_2^{1/2} x_3^{1/2} + a_8 x_2^{1/2} + a_9 x_3 + a_{10} x_3^{1/2}, \quad (17)$$

when homogeneity is not imposed. Notice that with  $b = \frac{1}{2}$  when homogeneity is imposed this results in the generalized Leontief functional form. Similarly, polynomials in bases  $b = 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$  can be defined. With homogeneity imposed the cost function is:

$$a_1 x_1 + a_2 x_1^{1/2} x_2^{1/2} + a_3 x_1^{1/2} x_3^{1/2} + a_4 x_2 + a_5 x_2^{1/2} x_3^{1/2} + a_6 x_3.$$

In their application to the 158 observations of Christensen and Greene (1976), Geweke and Petrella (2012) impose regularity at 1,000 points randomly generated on a hyper-rectangle bounded below by 90% of minimum price and bounded above by 110% of maximum price. A significant result of the study was that base  $b = \frac{1}{2}$  polynomials provided the best fit assessed by the Bayes factor so the generalized Leontief functional form is strongly favored by

the data (as Bayes factor were always in excess of 5.5).

### 3.2 Smoothly mixing regressions and fractional polynomial approximations

The simplicity of fractional polynomials of base  $b = \frac{1}{2}$  makes them ideal candidates for incorporation into the SMR framework. We can represent the conditional distribution of a dependent variable  $y$  given  $\mathbf{x} \in \mathfrak{R}^n$  as:

$$p(y|\mathbf{x}) = \sum_{g=1}^G \omega_g(\mathbf{x}; \mathbf{c}) f_N(v_g(\mathbf{x}; \mathbf{a}), \exp(w_g(\mathbf{x}; \mathbf{b}))), \quad (18)$$

where

$$v_g(\mathbf{x}; \mathbf{a}) = \sum_{i=1}^k a_{gi} \prod_{j=1}^n x_{tj}^{s_{ji}}, \quad (19)$$

$$w_g(\mathbf{x}_t; \mathbf{b}) = \sum_{i=1}^k b_{gi} \prod_{j=1}^n x_{tj}^{s_{ji}}, \quad (20)$$

assuming that both fractional polynomials have the same number of terms  $k$ . The parametrization of the variances is made for convenience so that one does not have to check positivity. The polynomials imply vectors of parameters  $\mathbf{a}_g = [a_{g1}, \dots, a_{gn}]'$  and  $\mathbf{b}_g = [b_{g1}, \dots, b_{gn}]'$  for each  $g = 1, \dots, G$ . Then,  $\mathbf{a} = [\mathbf{a}'_1, \dots, \mathbf{a}'_G]'$ ,  $\mathbf{b} = [\mathbf{b}'_1, \dots, \mathbf{b}'_G]'$  and  $\mathbf{c} = [\mathbf{c}'_1, \dots, \mathbf{c}'_G]$ .

### 3.3 Major extension

Given the ability of fractional polynomial expansions to approximate arbitrary regular functions and their derivatives it seems plausible that they can be generalized to the non-fractional case as follows:

$$p(\mathbf{x}_t; \boldsymbol{\alpha}) = \sum_{i=1}^k \alpha_i \prod_{j=1}^n x_{tj}^{\beta_{ji}}, \quad (21)$$

where  $\alpha_i$  are parameters as the  $a_i$  s but the  $\beta_{ji}$  s are parameters as well ( $\beta_{ji} \geq 0, \sum_{j=1}^n \beta_{ji} = 1$ ). The functional form can be written as:

$$p(\mathbf{x}_t; \boldsymbol{\alpha}) = \alpha_1 \left( x_{t1}^{\beta_{11}} x_{t2}^{\beta_{21}} \dots x_{tn}^{\beta_{n1}} \right) + \alpha_2 \left( x_{t1}^{\beta_{12}} x_{t2}^{\beta_{22}} \dots x_{tn}^{\beta_{n2}} \right) + \dots + \alpha_k \left( x_{t1}^{\beta_{1k}} x_{t2}^{\beta_{2k}} \dots x_{tn}^{\beta_{nk}} \right) \quad (22)$$

For identification we can impose the restrictions:

$$\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_k. \quad (23)$$

When the  $\beta_{ji}$  s are parameters we can economize on the order of a given fractional polynomial expansion.



Homogeneity of degree one can be imposed easily using:

$$\sum_{i=1}^k \alpha_i = 1. \quad (24)$$

Positivity reduces to  $\alpha_i > 0, i = 1, \dots, k$  and the curvature restrictions can be imposed using the methodology of the previous section.

An important exercise is to test whether the proposed functional form outperforms a given fractional polynomial. It is, indeed, possible to do so when the number of terms  $k$  increases. At the same time the proposed functional form satisfies automatically the positivity, homogeneity and positive-derivative properties provided we impose the restrictions  $\beta_{ji} \geq 0, j = 1, \dots, n, i = 1, \dots, k$ . These restrictions are independent of the data so it remains to impose only curvature restrictions. These restrictions, however, are automatically satisfied provided  $\beta_{ji} < 1, j = 1, \dots, n, i = 1, \dots, k$  since the functional form is a sum of Cobb-Douglas type functions.

Embedded into the SMR framework, effectively we can generalize the conditional distribution  $y|\mathbf{x}$  to a semi-parametric form that is capable of approximating arbitrary distributions as the number of groups,  $G$ , increases. The SMR allows for arbitrary heteroskedasticity, skewness and kurtosis without sacrificing simplicity as we deal with a mixture-of-normals distribution.

### 3.4 Technical inefficiency

We can allow for technical inefficiency by assuming:

$$y_t = \sum_{i=1}^k \alpha_i \prod_{j=1}^n x_{tj}^{\beta_{ji}} + \varepsilon_t + u_t, \quad (25)$$

where  $u_t \geq 0$  is the one-sided error term representing technical inefficiency. Flexible distributions for the two error components are as follows:

$$p(\varepsilon_t|\mathbf{x}_t) = \sum_{g=1}^{G_\varepsilon} \omega_g^{(\varepsilon)}(\mathbf{x}_t) f_N\left(0, \exp(w_g^{(\varepsilon)}(\mathbf{x}))\right), \quad (26)$$

and, independently,

$$p(u_t|\mathbf{x}_t) = \sum_{g=1}^{G_u} \omega_g^{(u)}(\mathbf{x}_t) f_{N+}\left(0, \exp(w_g^{(u)}(\mathbf{x}_t))\right). \quad (27)$$

Here<sup>5</sup>,  $\omega_g^{(\varepsilon)}$  and  $\omega_g^{(u)}$  represent fractional or Cobb-Douglas type approximations ( $g = 1, \dots, G$ ). We should notice that standard assumptions in the SFM literature, responsible for its inflexibility, such as independence or orthogonality between the error components and the regressors, are removed here. Here,  $f_{N+}(\mu, \sigma^2)$  denotes the density of the

---

<sup>5</sup>Clearly, flexibility in modeling the distributions of the two error components also accounts for endogeneity which is an additional advantage of smoothly mixing regressions.

half-normal distribution. The assumption can be generalized to the case of the truncated normal distribution:

$$p(u_t|\mathbf{x}_t) = \sum_{g=1}^{G_u} \omega_{u,g}(\mathbf{x}_t) f_{N+} \left( v_g^{(u)}(\mathbf{x}_t), \exp(w_g^{(u)}(\mathbf{x}_t)) \right). \quad (28)$$

There is an alternative to (21) which is, perhaps, more faithful to the SMR concept. The alternative is to assume:

$$y_t|\mathbf{x}_t, u_t, I_t = g \sim \mathcal{N} \left( v_g^{(\varepsilon)}(\mathbf{x}_t) + u_t, \exp(w_g^{(\varepsilon)}(\mathbf{x}_t)) \right), \quad (29)$$

where  $I_t$  is an index that represents the group to which the observation belongs and

$$p(I_t = g|\mathbf{x}_t) = \omega_g(\mathbf{x}) = \frac{\exp(\gamma'_g \mathbf{x})}{\sum_{s=1}^G \exp(\gamma'_s \mathbf{x})}, \quad g = 1, \dots, G. \quad (30)$$

The marginal distribution of  $u_t$  is given by (27).

In (29)  $v_g^{(\varepsilon)}(\mathbf{x}_t)$  is a real or fractional polynomial expansion which is group-specific whereas in (21) a single polynomial expansion is used to approximate the functional form. The distributional assumptions are relaxed via (29) and (30).

## 4 Panel data and Endogeneity

### 4.1 Panel data

With panel data, unobserved heterogeneity can be introduced via the assumption:

$$y_{it}|\mathbf{x}_{it}, u_{it}, I_{it}, \lambda_i = g \sim \mathcal{N} \left( v_g^{(\varepsilon)}(\mathbf{x}_{it}) + \lambda_i + u_{it}, \exp(w_g^{(\varepsilon)}(\mathbf{x}_{it})) \right), \quad (31)$$

( $i = 1, \dots, N, t = 1, \dots, T$ ) where  $\lambda_i$  denotes the individual effects or

$$y_{it} = v_g^{(\varepsilon)}(\mathbf{x}_{it}) + \lambda_i + u_{it} + \varepsilon_{it}, \quad (32)$$

conditional on  $\mathbf{x}_{it}, \lambda_i, u_{it}$  and  $I_{it} = g \in \{1, \dots, G\}$ , along with:

$$\varepsilon_{it}|\mathbf{x}_{it} \sim \mathcal{N} \left( 0, \exp(w_g^{(\varepsilon)}(\mathbf{x}_{it})) \right). \quad (33)$$

Modeling the individual effects relies on a substantial generalization of Mundlak's (1978) device:

$$p(\lambda_i) = \sum_{g=1}^{G_\lambda} \omega_{u,g}(\bar{\mathbf{x}}_i) f_{N+} \left( 0, \exp(w_g^{(\lambda)}(\bar{\mathbf{x}}_i)) \right), \quad (34)$$

where  $\bar{\mathbf{x}}_i = T^{-1} \sum_{t=1}^T \mathbf{x}_{it}$  represents the average value of the covariates and  $\exp(w_g^{(\lambda)}(\bar{\mathbf{x}}_i))$  is a (flexible) function representing the log of their variance. In this formulation it is not necessary to assume independence between the individual effects and the regressors.

## 4.2 Endogeneity

Often the assumption that the regressors  $\mathbf{x}_{it}$  satisfy a strong exogeneity assumption cannot be maintained. We have used this assumption previously as flexibility in modeling the distributions of the two error components also accounts for endogeneity which is an additional advantage of SMR. Suppose an  $M \times 1$  vector  $\mathbf{d}_{it}$  may be available which can be considered exogenous (for example, time trend). Endogeneity can be handled, alternatively, by assuming a panel vector autoregression (PVAR) of the form:

$$\mathbf{x}_{it} = \mathbf{a}_i + \mathbf{B}\mathbf{x}_{i,t-1} + \mathbf{\Gamma}\mathbf{d}_{it} + \boldsymbol{\xi}_{it}, \quad (35)$$

where  $\boldsymbol{\xi}_{it} \sim \mathcal{N}(\mathbf{O}, \boldsymbol{\Sigma}_\xi)$ ,  $\mathbf{B}$  and  $\mathbf{\Gamma}$  are matrices  $n \times n$  and  $n \times M$  respectively, and the individual effects

$$\mathbf{a}_i \sim \mathcal{N}_n(\mathbf{O}, \boldsymbol{\Sigma}_a), \quad (36)$$

independently of  $\boldsymbol{\xi}_{it}$ . The assumption can be generalized to an SMR specification for the  $\mathbf{a}_i$  s as functions of  $\bar{\mathbf{x}}_i$  and, possibly,  $\mathbf{d}_i = T^{-1} \sum_{t=1}^T \mathbf{d}_{it}$ . It is possible to extend the PVAR in (35) to an SMR context but this extension does not seem very promising. The important issue is to allow for correlation between  $\boldsymbol{\xi}_{it}$  and  $\varepsilon_{it}$  in (32). This can be done by assuming that:

$$\begin{bmatrix} \varepsilon_{it} \\ \boldsymbol{\xi}_{it} \end{bmatrix} \sim \mathcal{N}_{n+1}(\mathbf{O}, \boldsymbol{\Sigma}_{it}). \quad (37)$$

The dependence between  $\varepsilon_{it}$  and  $\boldsymbol{\xi}_{it}$  can be used to model endogeneity as an alternative to SMR, see (29)-(30). The assumption has been introduced by Kutlu (2010) and Tran and Tsionas (2013) although with a fixed  $\boldsymbol{\Sigma}_{it}$ .

For the elements of  $\boldsymbol{\Sigma}_{it}$  we assume that  $\text{var}(\varepsilon_{it})$  is given by (33), the diagonal elements of  $\text{cov}(\boldsymbol{\xi}_{it})$  are fixed parameters or depend on  $\mathbf{x}_{i,t-1}$  and possibly  $\mathbf{d}_{it}$  via flexible functional forms, and non-diagonal elements are fixed. Alternatively, consider the Cholesky factorization:

$$\boldsymbol{\Sigma}_{it} = \mathbf{H}_{it}\mathbf{H}'_{it}, \quad (38)$$

where  $\mathbf{H}_{it}$  is a lower diagonal  $n \times n$  matrix. Denote  $\mathbf{h}_{it} = \text{vec}(\mathbf{H}_{it}) = [h_{it,1}, \dots, h_{it,m}]'$  which is a vector of dimension  $m = \frac{n(n+1)}{2}$ . We assume

$$h_{it,j} = w^{(j)}(\mathbf{x}_{i,t-1}, \mathbf{d}_{it}), j = 1, \dots, m \quad (39)$$

where  $w^{(j)}()$  represents a flexible functional form. The parameters of the specification are unrestricted and allow for a quite flexible specification of the joint distribution of  $\varepsilon_{it}$  and  $\boldsymbol{\xi}_{it}$ . The assumption can be further generalized as follows:

$$p(\varepsilon_{it}, \boldsymbol{\xi}_{it}) \equiv p(\boldsymbol{\zeta}_{it}) = \sum_{g=1}^{G_{\boldsymbol{\zeta}}} \omega^{(g)}(\mathbf{x}_{it}) f_{N,n+1}(\mathbf{O}, \boldsymbol{\Sigma}_{it}^{(g)}), \quad (40)$$

where  $f_{N,n+1}(\mathbf{O}, \boldsymbol{\Sigma}_{it})$  denotes the  $(n+1)$ -variate normal distribution with mean zero and covariance matrix  $\boldsymbol{\Sigma}_{it}$ , whose specification is given above. As in SMR the weights  $\omega^{(g)}$  depend on the regressors  $\mathbf{x}_{it}$ . Due to the explicit parametrization of  $\mathbf{h}_{it}$  in (39) the use of a multivariate normal mixture for the joint multivariate distribution is not as demanding as when  $\boldsymbol{\Sigma}^{(g)}$  is left unrestricted and, therefore, it can be used even when the number of variables ( $n$ ) is relatively large.

### 4.3 Endogeneity through copulas

In this section suppose  $\Phi(z)$  denotes the standard normal cdf and  $\Phi_{k+1,\Sigma}(\mathbf{z})$  denotes the cdf of a  $(k+1)$ -dimensional normal distribution with correlation matrix  $\Sigma$ .

Given the marginal distributions  $F_{\varepsilon}(\varepsilon_{it})$  and  $F_j(x_{j,it})$ ,  $j = 1, \dots, k$  and the densities  $f_{\varepsilon}(\varepsilon_{it})$  and  $f_j(x_{j,it})$ ,  $j = 1, \dots, k$ , the joint distribution of  $\varepsilon_{it}$  and  $\mathbf{x}_{it}$  can be represented by a copula function  $\mathcal{C}(F_{\varepsilon}, \mathbf{F})$  where  $\mathbf{F}(\mathbf{x}) = [F_1(x_1), \dots, F_k(x_k)]'$ ,  $\mathbf{x} \in \mathfrak{R}^k$ . The copula is defined in  $[0, 1]^{k+1}$  by:

$$\mathcal{C}(\xi_1, \dots, \xi_{k+1}) = P(F_{\varepsilon}(\varepsilon_{it}) \leq \xi_1, F_1(x_{1,it}) \leq \xi_2, \dots, F_k(x_{k,it}) \leq \xi_{k+1}), \quad (41)$$

so that the copula is itself a cdf. Moreover,  $U_j = F_j(x)$  and  $U_{\varepsilon} = F_{\varepsilon}(\varepsilon)$  have uniform distributions. If  $c(\xi_1, \dots, \xi_{k+1})$  denotes the pdf associated with the copula, then by Sklar's (1959) theorem, we have:

$$f(\varepsilon, x_1, \dots, x_k) = c(\varepsilon, x_1, \dots, x_k) \cdot f_{\varepsilon}(\varepsilon) \cdot \prod_{j=1}^k f_j(x_j). \quad (42)$$

One commonly used copula function is the Gaussian copula<sup>6</sup>. The  $(k+1)$ -dimensional CDF with correlation matrix  $\Sigma$  is given by  $c(w; \Sigma) = \Phi_{k+1,\Sigma}(\Phi^{-1}(U_{\varepsilon}), \Phi^{-1}(U_1), \dots, \Phi^{-1}(U_k))$ , where  $w := (U_{\varepsilon}, U_1, \dots, U_k) = (F_{\varepsilon}(\varepsilon), F_1(x_1), \dots, F_k(x_k))$ .

The density copula is:

$$c(w; \Sigma) = \det(\Sigma)^{-1/2} \exp \left\{ -\frac{1}{2} [\Phi^{-1}(U_{\varepsilon}), \Phi^{-1}(U_1), \dots, \Phi^{-1}(U_k)]' (\Sigma^{-1} - I_{k+1}) [\Phi^{-1}(U_{\varepsilon}), \Phi^{-1}(U_1), \dots, \Phi^{-1}(U_k)] \right\} \quad (43)$$

<sup>6</sup>Relative to other copulas, the Gaussian copula is generally robust for most application and has many desirable properties (Danaher and Smith, 2011).

and the joint distribution is given by

$$\log f(\varepsilon, x_1, \dots, x_k) = \log c(w; \Sigma) + \log f_\varepsilon(\varepsilon) + \sum_{j=1}^k \log f_j(x_j). \quad (44)$$

Clearly, it is easy to formulate and evaluate (44) as  $c(w; \Sigma)$  can be computed easily and marginals  $f_\varepsilon(\varepsilon)$  and  $f_1(x_1), \dots, f_k(x_k)$  are assumed available. Regarding the marginal density  $f_\varepsilon(\varepsilon)$  enough has been written in the previous sections. For the marginals of the regressors, we approximate them using:

$$\tilde{F}_j(x_j) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{I}(x_{j,it} \leq x_j), \quad j = 1, \dots, k, \quad (45)$$

and we use the scaling factor  $n+1$  to avoid difficulties arising from the potential unboundedness of  $\log c(w; \Sigma)$  as some of arguments tend to one. Additionally, as  $\sum_{j=1}^k \log f_j(x_j)$  does not depend on the parameters, it can be omitted from (44). To obtain  $F_\varepsilon(\varepsilon) = \int_{-\infty}^{\varepsilon} f_\varepsilon(e) de$  we, generally, need to use univariate numerical integration through, for example, Gaussian quadrature. We use 40-point quadrature using the Gauss-Kronrod rule as implemented in IMSL.

Based on (44) we can define the log-likelihood function:

$$\log f(\varepsilon, \mathbf{x}_1, \dots, \mathbf{x}_k) = \sum_{i,t} \log c(w_{it}; \Sigma) + \sum_{i,t} \log f_\varepsilon(\varepsilon_{it}), \quad (46)$$

ignoring a constant term,  $\varepsilon_{it} := y_{it} - f(\mathbf{x}_{it}; \beta)$  and  $w := (F_\varepsilon(\varepsilon_{it}), F_1(x_{1,it}), \dots, F_k(x_{k,it}))$ . The log-likelihood can be maximized with respect to  $\theta = [\beta', \sigma, \lambda]'$  and the different elements of  $\Sigma$ .

## 5 Computational experiment I

In this computational experiment we consider artificial data with  $N = 1,500$  and  $N = 5,000$  observations with  $T = 10$  time periods (and therefore  $n = 150$  or  $n = 500$ ). Of course,  $N = nT$ . The sample sizes are typical of what is used in applied econometrics. We have three relative input prices and three outputs, and a cost function is assumed along with share equations derived from Shephard's lemma. The cost function is of the form  $C = C(w_1, w_2, w_3; y_1, y_2, y_3, t)$ . We consider two basic data generating processes:

- i) A mixture of Translog models with  $C$  components.
- ii) A mixture of Quadratic models with  $C$  components.

All models satisfy the theoretical regularity conditions. For the generalized Leontief this is straightforward while for the translog models we follow the procedure set forth by Perelman and Santin (undated). We vary  $C$  from 1 to 5 so effectively we have ten data generating processes in cases (i) and (ii). Mixture models are generated so that each one has equal probability. We set  $\sigma_v = \sigma_u = 0.3$  which is also typical. Individual effects are included in all models and are generated from a normal distribution with zero mean and standard deviation 0.1.

Our interest focuses on estimating (i) technical change, (ii) technical inefficiency, (iii) input elasticities and returns to scale. These measures depend on the parametrizations used to make the functional forms regular for each observation in the samples.

To determine the best model when using a fractional polynomial approximation (FPA) or an approximation using the additive Cobb-Douglas (ACD) specification we use the modified log predictive scores methodology. The cross-validated log score (Gelfand et al, 1992) is:

$$\sum_{i=1}^N \log p(y_i | \mathbf{Y}_{-i}) \cong \sum_{i=1}^N \log \left( S^{-1} \sum_{s=1}^S \log p(y_i | \boldsymbol{\theta}^{(s)}, \mathbf{Y}_{-i}) \right). \quad (47)$$

The modified cross-validated log score (Geweke and Keane, 2007) relies on keeping the first  $N_1 < N$  observations for estimation and keeping the remaining for cross-validation. This is repeated  $R$  times and we take the mean:

$$R^{-1} \sum_{r=1}^R \left( \sum_{i=N_1+1}^N \log p \left( y_i^{(r)} | y_{-i}^{(r)}, \mathbf{Y}_{N_1}^{(r)} \right) \right), \quad (48)$$

where  $\mathbf{Y}^{(r)}$  is a random selection from  $\mathbf{Y}$  (the data) and averaging over the draws  $\boldsymbol{\theta}^{(s)}$  is also performed to compute the score:

$$R^{-1} \sum_{r=1}^R \left( \sum_{i=N_1+1}^N \log p \left( y_i^{(r)} | y_{-i}^{(r)}, \mathbf{Y}_{N_1}^{(r)} \right) \right) \cong R^{-1} \sum_{r=1}^R \left( S^{-1} \sum_{s=1}^S \sum_{i=N_1+1}^N \log p \left( y_i^{(r)} | y_{-i}^{(r)}, \mathbf{Y}_{N_1}^{(r)}, \boldsymbol{\theta}^{(s)} \right) \right). \quad (49)$$

The difference is that the Gelfand et al (1992) variant is more computationally demanding as  $N$  posterior simulators are needed while the Geweke and Keane (2007) variant uses only  $R \ll N$  posterior MCMC simulators so it can be computationally more efficient. Models with higher log predictive score perform better in terms of cross-validated prediction. In this work we take  $R = 50$  and  $N_1 = 1,000$  when  $N = 1,500$  and  $N_1 = 4,500$  when  $N = 5,000$  so that 500 observations are always left out for cross-validation.

Posterior simulation via MCMC is performed, for both FPA and ACD, using a variant of Sequential Monte Carlo as described in the Appendix. All posterior simulators rely on a transient phase of 50,000 simulations followed by another 100,000 draws. Convergence is monitored using Geweke's (1992) convergence diagnostics and numerical standard errors (NSE) as well as relative numerical efficiency (RNE) are computed as well, based on AR(10) approximations to compute the spectral density of the process at the origin. Latent technical efficiency is explicitly integrated out of the posterior to reduce the amount of autocorrelation inherent in MCMC computations.

The theoretical restrictions are imposed on a grid rather than on each observation. Since we have three relative prices and three outputs, following Geweke and Petrella (2012) we choose a random grid of size  $aN$  where  $a = \frac{1}{2}$ , 1 or 2. When, for example,  $N = 5,000$  this means that the theoretical restrictions are imposed at 2,500, 5,000 or 10,000 points using rejection. In all cases the results were qualitatively and quantitatively the same when  $a = 1$  or

$a = 2$  implying that the imposition of restrictions at  $N/2$  points was more than adequate. Relative to Geweke and Petrella (2012) we added the following check to facilitate the imposition of theoretical restrictions: The restrictions are imposed at the point of means and then also at points on a ball whose radius was  $r = \frac{1}{2}$  and  $r = \frac{3}{2}$  relative to the mean. We count the number of violations in each specification and if it exceeds  $\frac{1}{3}$  of the sample size ( $N$ ) we add more points. We ended up with a fixed specification where the theoretical restrictions are exactly imposed at points with radius  $r = 1$  (the mean) and also  $r = \pm \frac{3}{k}$  for  $k = 2, 4, 5$  and  $r = \pm \frac{1}{k}$  for  $k = 2, 3, 4, 5$  plus the 90% and 110% minimum and maximum of the data: The random grid was set up so that 90% and 110% of the minimum and maximum respectively of the data, are included in the random grid. We use standard uniform random numbers to construct the grid instead of quasi-random numbers as in Geweke and Petrella (2012). We did not experiment with quasi-random numbers as the algorithms were quite successful after imposing the constraints at points around the means as described above. To facilitate the imposition of theoretical restrictions the data are divided by their means so that in logs they have zero mean (in translog) or unit mean (in the Quadratic specification).

As a basis of comparison of take the simple translog specification when  $N = 1,500$ . The results for  $N = 5,000$  are quite similar qualitatively and are not reported in the interest of space.

**Table 1. Model Comparison**

	Fractional Polynomial Approximation					Additive Cobb-Douglas					$k = 6$
	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	
Translog	0.000	-1.252	-3.454	-4.821	-4.901	0.000	-4.216	-3.217	-6.218	-3.216	-3.891
Mixture of Translog, $C = 2$	0.000	-3.742	-5.8231	-6.771	-7.004	0.000	-3.256	-1.517	-2.872	-3.217	-3.779
Mixture of Translog, $C = 3$	0.000	-2.476	-3.821	-4.202	-5.891	0.000	-4.289	-2.777	-3.252	-4.220	-6.242
Mixture of Translog, $C = 4$	0.000	-1.717	-3.672	-4.406	-5.202	0.000	-3.781	-2.882	-4.847	-5.201	-8.782
Mixture of Translog, $C = 5$	0.000	-2.233	-1.785	-3.229	-5.227	0.000	-4.472	-5.212	-3.217	-6.220	-7.961
Quadratic	0.000	1.217	2.822	3.217	3.474	0.000	-2.122	-4.181	-4.557	-5.228	-8.278
Mixture of Quadratic, $C = 2$	0.000	-0.893	-1.217	-1.517	-3.891	0.000	-2.771	-4.891	-5.289	-7.007	-11.651
Mixture of Quadratic, $C = 3$	0.000	-1.891	-1.343	-3.117	-3.828	0.000	-7.289	-3.219	-11.232	-18.298	-20.235
Mixture of Quadratic, $C = 4$	0.000	-3.774	-2.338	-4.555	-6.229	0.000	-7.298	-5.224	-6.282	-8.272	-11.289
Mixture of Quadratic, $C = 5$	0.000	-5.217	-3.202	-5.178	-8.272	0.000	-9.216	-6.781	-8.991	-10.765	-13.738

Notes: Although we have panel data, we do not determine the sample of size  $N_1$  by retaining the firm structure. Therefore, we randomize as if the data were iid. We feel this provides a more stringent cross-validation comparison. Otherwise, we would have to select a sample of size  $n_1 < n$  which retains all  $T$  observations for a particular firm. This could bias the results for technical change reported below in Table 2.

To make the comparison easier we normalize the log predictive score to 0 and we take deviation of other LPS from the LPS at zero.

From Table 1 it turns out that low-order fractional polynomials but also low order ACD models perform quite

well in terms of cross-validation. Perhaps surprisingly, the ACD model does not need more than three components to approximate complicated DGPs like the Translog or Quadratic specification with large number of mixing components. This shows that the ACD can be a valid competitor to the more complicated Fractional Polynomial Approximation.

Of course, the critical issue is whether functions of interest like technical change and technical inefficiency can be approximated equally well by the FPA and the ACD approximations. For the best models, selected using the modified cross-validated LPS of Geweke and Keane (2007) relevant information is presented in Table 2, where we examine the performance of both FPA and ACD.

**Table 2. Model Comparison in terms of functions of interest**

(rank correlation coefficients, median across all MCMC draws finally retained, skipping every other 10th draw)

	Technical change	Inefficiency	$\varepsilon_{w1}$	$\varepsilon_{w2}$	$\varepsilon_{w3}$	$\varepsilon_{y1}$	$\varepsilon_{y2}$	$\varepsilon_{y3}$
Translog	0.982	0.995	0.988	0.993	0.985	0.981	0.995	0.997
Mixture of Translog, $C = 2$	0.985	0.995	0.985	0.995	0.991	0.986	0.997	0.997
Mixture of Translog, $C = 3$	0.983	0.991	0.982	0.985	0.984	0.982	0.983	0.950
Mixture of Translog, $C = 4$	0.980	0.991	0.981	0.983	0.981	0.982	0.982	0.944
Mixture of Translog, $C = 5$	0.977	0.988	0.985	0.975	0.976	0.974	0.978	0.938
Quadratic	0.999	0.998	0.998	0.998	0.999	0.997	0.996	0.998
Mixture of Quadratic, $C = 2$	0.997	0.996	0.998	0.995	0.994	0.995	0.994	0.993
Mixture of Quadratic, $C = 3$	0.994	0.994	0.991	0.992	0.993	0.995	0.992	0.993
Mixture of Quadratic, $C = 4$	0.987	0.988	0.990	0.982	0.987	0.983	0.981	0.992
Mixture of Quadratic, $C = 5$	0.984	0.983	0.988	0.982	0.984	0.980	0.973	0.985

Notes:  $\varepsilon_{wj} = \frac{\partial C}{\partial w_j} \frac{w_j}{C}$  and  $\varepsilon_{yj} = \frac{\partial C}{\partial y_j} \frac{y_j}{C}$ . For each MCMC draw these are series which can be computed for both the FPA and ACD approximations. The rank correlation is then computed, monitored for each draw (omitting every other tenth) and then the medians are reported. A measure of inverse returns to scale is  $e_{cy} = \sum_j \varepsilon_{yj}$ , that is  $RTS = e_{cy}^{-1}$ . Technical change is  $TC = \frac{1}{C} \frac{\partial C}{\partial t}$ . Technical inefficiency for each draw is computed from its posterior conditional distribution after the draws for regression parameters and scale parameters become available, using well-known expressions. Averaging across parameter draws produces technical inefficiency estimates which are then compared for the FPA and ACD approximations using rank correlation coefficients.

From Table 2, it is evident that estimated functions of interest are quite close in terms of rank correlation coefficients and therefore the good behavior of ACD, reported in Table 1, produces estimated functions of interest that are quite close to those corresponding to the FPA. Given the hardness of imposing the theoretical restrictions in FPA, the excellent behavior of ACD should, undoubtedly, encourage its use in applied econometrics.

An interesting question is whether Finite-Mixture-of-Normals-Models (FMNM) can also be used profitably as approximations. FMNM based on Cobb-Douglas regression models with different scale parameters can also



approximate, in theory, arbitrary functional forms (Norets and Pelenis, 2011). Although Norets and Pelenis (2011) employ a more general approach to extract conditional distributions from a general distribution provided by a FMNM, a simpler approach is to use FMNM of stochastic frontier models in the usual sense often used in the literature.

**Table 3. Results from FMNM**

(rank correlation coefficients, median across all MCMC draws finally retained, skipping every other 10th draw)

	Technical change	Inefficiency	Optimal order of FMNM
Translog	0.855	0.671	5
Mixture of Translog, $C = 2$	0.651	0.555	7
Mixture of Translog, $C = 3$	0.454	0.403	8
Mixture of Translog, $C = 4$	0.301	0.322	9
Mixture of Translog, $C = 5$	0.271	0.255	9
Quadratic	0.345	0.353	7
Mixture of Quadratic, $C = 2$	0.212	0.301	9
Mixture of Quadratic, $C = 3$	0.117	0.274	11
Mixture of Quadratic, $C = 4$	0.101	0.216	12
Mixture of Quadratic, $C = 5$	0.007	0.189	14

Notes: Reported results are for sample size  $N=5,000$ . Results for  $N=1,500$  were qualitatively similar.

From Table 3 it turns out that the approximation properties of FMNM are worse when compared to ACD and FPA. We have computed but do not report, in the interest of space, rank correlations between true and estimated elasticities and returns to scale (RTS) to obtain a similar result. The rank correlation between true and estimated elasticities and RTS range from, approximately, 0.3 to 0.6.

## 6 Computational experiments II

In this section we consider the computational experiments in Badunenko, Henderson and Kumbhakar (2012). They consider two simple production functions: (1) Cobb-Douglas (CD)  $y = x_1^\alpha x_2^{1-\alpha} \exp(v - u)$  and (2) a constant elasticity of substitution (CES)  $y = [\beta x_1^\rho + (1 - \beta)x_2^\rho]^{-1/\rho} \exp(v - u)$ . They set  $\alpha = \frac{1}{3}$ ,  $\beta = \frac{2}{3}$  and  $\rho = \frac{1}{2}$ . For the error term we assume  $v \sim N(0, \sigma_v^2)$  and for inefficiency we have  $u \sim N_+(0, \sigma_u^2)$ . Three sample sizes are analyzed (total number of observations is 50, 100 and 200). To generate  $x_1$  and  $x_2$  they assume that they are uniformly drawn in the interval  $[1, 2]$  iid and independently of each other. With respect to the parameters of noise and inefficiency they have three scenarios:

In scenario S1 ( $\sigma_v = \sigma_u = 0.01, \lambda = 1.0$ ), both terms are relatively small. In other words, the data are measured with relatively little error and the units are relatively efficient. In scenario S2 ( $\sigma_v = 0.01, \sigma_u = 0.05, \lambda = 5.0$ ), the data have relatively little noise, but the units under consideration are relatively inefficient. In scenario S3 ( $\sigma_v = 0.05, \sigma_u = 0.01, \lambda = 0.2$ ), the data are relatively noisy and the the firms are relatively efficient. In scenario S4 ( $\sigma_v = 0.05, \sigma_u = 0.05, \lambda = 1.0$ ), the data are relatively noisy and the the firms are relatively inefficient. This scenario is actually redundant as the results mostly depend on  $\lambda$  and not the individual values of  $\sigma_v$  and  $\sigma_u$ . However, the case is interesting since we can examine whether the results depend on actual values of  $\lambda$  and / or the magnitude of noise and inefficiency.

All experiments consist of 2000 Monte Carlo trials. Data sets where the residuals have wrong (positive) skewness are discarded. We compare our results only with the frontier estimator (FLW) as the DEA estimator (KSW) has been found to have disappointing performance under noise -even moderate- in the study of Badunenko, Henderson and Kumbhakar (2012).

We use the following metrics, as in Badunenko, Henderson and Kumbhakar (2012):

$$\text{Bias} = n^{-1} \sum_{i=1}^n (\widehat{TE}_i - TE_i)$$

$$\text{RMSE} = \{n^{-1} \sum_{i=1}^n (\widehat{TE}_i - TE_i)\}^{1/2}$$

$$\text{Upward Bias} = n^{-1} \sum_{i=1}^n I(\widehat{TE}_i > TE_i)$$

$$\text{Kendall's } \tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$$

where  $\widehat{TE}_i$  is estimated technical efficiency,  $TE_i = \exp(-u_i)$  is actual technical efficiency,  $n_c, n_d$  represent the number of concordant pairs and the number of discordant pairs in the data set (efficiency ranks) respectively.

From Tables 4 and 5 it turns out in cases of Cobb-Douglas and CES that the methods developed in this study improve drastically over the procedures in FLW. Coverage is much better in many cases, biases and RMSEs are significantly lower, the upward bias is mitigated and the correlations between actual and predicted efficiency are much higher.

**Table 4. Finite sample performance of the efficiency estimates: Cobb-Douglas production technology**

	ECA, 95% <sup>(a)</sup>		Bias		RMSE		Upward bias <sup>(c)</sup>		Correlation <sup>(d)</sup>	
	FLW <sup>(b)</sup>	this study	FLW	this study	FLW	this study	FLW	this study	FLW	this study
	Scenario S1 ( $\sigma_v = \sigma_u = 0.01, \lambda = 1.0$ )									
n=50	0.92	0.92	-0.0034	-0.0018	0.0044	0.0038	0.320	0.503	0.43	0.58
n=100	0.95	0.95	-0.0030	-0.0017	0.0039	0.0035	0.320	0.503	0.45	0.66
n=200	0.96	0.95	-0.0031	-0.0013	0.0591	0.0302	0.315	0.501	0.46	0.79
	Scenario S2 ( $\sigma_v = 0.01, \sigma_u = 0.05, \lambda = 5.0$ )									
n=50	0.92	0.93	-0.0107	-0.007	0.0133	0.008	0.28	0.504	0.76	0.80
n=100	0.94	0.95	-0.0127	-0.005	0.0132	0.007	0.21	0.504	0.78	0.82
n=200	0.94	0.95	-0.0148	-0.003	0.0148	0.006	0.16	0.502	0.80	0.85
	Scenario S3 ( $\sigma_v = 0.05, \sigma_u = 0.01, \lambda = 0.2$ )									
n=50	0.78	0.85	-0.0379	-0.0181	0.0379	0.0122	0.020	0.503	0.12	0.08
n=100	0.85	0.89	-0.0331	-0.0122	0.0331	0.0120	0.020	0.505	0.12	0.07
n=200	0.88	0.93	-0.0278	-0.0115	0.0278	0.0117	0.025	0.505	0.12	0.06
	Scenario S4 ( $\sigma_v = 0.05, \sigma_u = 0.05, \lambda = 1.0$ )									
n=50	0.92	0.95	-0.0134	-0.007	0.020	0.015	0.34	0.507	0.46	0.51
n=100	0.95	0.95	-0.0092	-0.006	0.016	0.008	0.36	0.503	0.47	0.55
n=200	0.96	0.95	-0.0059	-0.004	0.012	0.007	0.40	0.502	0.47	0.59

Notes:

Cobb-Douglas production function:  $y_i = x_{i1}^{1/2} x_{i2}^{2/3} \exp(v_i - u_i)$ ,  $v_i \sim N(0, \sigma_v^2)$ ,  $u \sim N_+(0, \sigma_u^2)$ ,  $\lambda = \frac{\sigma_u}{\sigma_v}$ .

(a) Empirical Coverage Accuracy is the share of true technical efficiencies that are within bounds of predicted 95% confidence interval for estimated technical efficiency. Reported in this table is the median of such shares across all Monte Carlo simulations;

(b) FLW represents the SSF estimator and the results are taken from Badunenko, Henderson and Kumbhakar (2012).

(c) Upward Bias is the share of predicted technical efficiencies strictly larger than the true ones. The desired value of upward bias is 0.5. The values less (greater) than 0.5 indicates systematic underestimation (overestimation) of technical efficiencies. Reported in the table is the median of such shares across all Monte Carlo simulations;

(d) Kendall correlation coefficient between predicted and true technical efficiencies. Reported in the table is the median of such coefficients across all Monte Carlo simulations.

**Table 5. Finite sample performance of the efficiency estimates: CES**

	ECA, 95% <sup>(a)</sup>		Bias		RMSE		Upward bias <sup>(c)</sup>		Correlation <sup>(d)</sup>	
	FLW <sup>(b)</sup>	this study	FLW	this study	FLW	this study	FLW	this study	FLW	this study
	S1 ( $\sigma_v = \sigma_u = 0.01, \lambda = 1.0$ )									
n=50	0.92	0.95	-0.0045	-0.0012	0.0052	0.0032	0.28	0.503	0.45	0.55
n=100	0.95	0.95	-0.0040	-0.0008	0.0045	0.0025	0.28	0.505	0.46	0.59
n=200	0.96	0.96	-0.0040	-0.0006	0.0650	0.0012	0.27	0.503	0.47	0.62
	S2 ( $\sigma_v = 0.01, \sigma_u = 0.05, \lambda = 5.0$ )									
n=50	0.92	0.95	-0.0120	-0.0001	0.0142	0.008	0.22	0.511	0.78	0.85
n=100	0.92	0.95	-0.0148	<-0.0001	0.0151	0.006	0.17	0.502	0.80	0.91
n=200	0.92	0.94	-0.0168	<-0.0001	0.0168	0.004	0.13	0.503	0.81	0.95
	S3 ( $\sigma_v = 0.05, \sigma_u = 0.01, \lambda = 0.2$ )									
n=50	0.78	0.95	-0.0397	-0.0122	0.0397	0.0215	0.020	0.504	0.11	0.38
n=100	0.85	0.95	-0.0334	-0.007	0.0334	0.0122	0.020	0.503	0.11	0.51
n=200	0.88	0.95	-0.0280	-0.005	0.0280	0.009	0.025	0.502	0.12	0.65
	S4 ( $\sigma_v = 0.05, \sigma_u = 0.05, \lambda = 1.0$ )									
n=50	0.92	0.94	-0.0157	-0.0034	0.0209	0.017	0.32	0.511	0.47	0.55
n=100	0.95	0.95	-0.0105	-0.0021	0.0167	0.007	0.35	0.507	0.47	0.69
n=200	0.96	0.95	-0.0068	-0.0015	0.0133	0.004	0.38	0.503	0.47	0.76

Notes:

CES production function:  $y_i = [\frac{2}{3}x_{i1}^{1/2} + \frac{1}{3}x_{i2}^{1/2}]^{1/2} \exp(v_i - u_i)$ ,  $v_i \sim N(0, \sigma_v^2)$ ,  $u \sim N_+(0, \sigma_u^2)$ ,  $\lambda = \frac{\sigma_u}{\sigma_v}$ .

(a) Empirical Coverage Accuracy is the share of true technical efficiencies that are within bounds of predicted 95% confidence interval for estimated technical efficiency. Reported in this table is the median of such shares across all Monte Carlo simulations;

(b) FLW represents the SSF estimator and the results are taken from Badunenko, Henderson and Kumbhakar (2012).

(c) Upward Bias is the share of predicted technical efficiencies strictly larger than the true ones. The desired value of upward bias is 0.5. The values less (greater) than 0.5 indicates systematic underestimation (overestimation) of technical efficiencies. Reported in the table is the median of such shares across all Monte Carlo simulations;

(d) Kendall correlation coefficient between predicted and true technical efficiencies. Reported in the table is the median of such coefficients across all Monte Carlo simulations.

In tables 4 and 5 we have shown that “on average” our improved procedure is arguably better than FLW. One interesting question is what would happen at low and upper ends of the efficiency distribution. We provide the results in Tables 4a and 5a for the 5% and 10% lower and upper ends of the efficiency distribution. Specifically, we proceed as follows: Given that an observation has efficiency in the lower 5% of the efficiency distribution *in the data*, we first examine whether it belongs to the lower 5% of the efficiency distribution *in a given simulation*. If not we record the correlation as zero; otherwise, we compute the actual correlation.

From the results in Tables 4a and 5a it turns out that, in most cases, the correlations at the lower and upper 5% and 10% ends of the efficiency distribution are lower than the “average” but fairly close to the “average” value.

Therefore, the performance of the improved procedure does not deteriorate and can be used safely in most cases.

**Table 4a. Finite sample performance of the efficiency estimates: Cobb-Douglas production technology (Correlations at lower and upper ends of the efficiency distribution)**

	lower 5%	lower 10%	upper 5%	upper 10%
	S1 ( $\sigma_v = \sigma_u = 0.01, \lambda = 1.0$ )			
n=50	0.48	0.51	0.50	0.52
n=100	0.60	0.62	0.61	0.62
n=200	0.72	0.75	0.72	0.75
	S2 ( $\sigma_v = 0.01, \sigma_u = 0.05, \lambda = 5.0$ )			
n=50	0.72	0.77	0.75	0.78
n=100	0.73	0.79	0.78	0.79
n=200	0.78	0.81	0.80	0.82
	S3 ( $\sigma_v = 0.05, \sigma_u = 0.01, \lambda = 0.2$ )			
n=50	0.05	0.07	0.04	0.05
n=100	0.06	0.07	0.06	0.07
n=200	0.05	0.07	0.05	0.07
	S4 ( $\sigma_v = 0.05, \sigma_u = 0.05, \lambda = 1.0$ )			
n=50	0.48	0.50	0.50	0.51
n=100	0.52	0.54	0.50	0.53
n=200	0.45	0.57	0.49	0.55

Notes: See notes to Table 4.

**Table 5a. Finite sample performance of the efficiency estimates: CES production technology (Correlations at lower and upper ends of the efficiency distribution)**

	lower 5%	lower 10%	upper 5%	upper 10%
S1 ( $\sigma_v = \sigma_u = 0.01, \lambda = 1.0$ )				
n=50	0.52	0.54	0.52	0.54
n=100	0.57	0.59	0.57	0.59
n=200	0.59	0.61	0.60	0.62
S2 ( $\sigma_v = 0.01, \sigma_u = 0.05, \lambda = 5.0$ )				
n=50	0.81	0.84	0.81	0.83
n=100	0.87	0.90	0.87	0.91
n=200	0.92	0.94	0.92	0.95
S3 ( $\sigma_v = 0.05, \sigma_u = 0.01, \lambda = 0.2$ )				
n=50	0.35	0.38	0.34	0.37
n=100	0.48	0.51	0.47	0.50
n=200	0.60	0.64	0.61	0.65
S4 ( $\sigma_v = 0.05, \sigma_u = 0.05, \lambda = 1.0$ )				
n=50	0.50	0.55	0.51	0.54
n=100	0.65	0.68	0.64	0.67
n=200	0.73	0.75	0.72	0.75

Notes: See notes to Table 5.

## 7 Data

We apply the new techniques to the US banking data of Malikov et al (2015) on which we rely heavily for the following description. Our data on commercial banks come from Call Reports available from the Federal Reserve Bank of Chicago and include all FDIC-insured commercial banks with reported data for 2001:Q1–2010:Q4. We focus on a selected subsample of relatively homogeneous large banks, namely those with total assets in excess of \$1 billion dollars (in 2005 US dollars) in the first 3 years of observation. We further exclude Internet banks, commercial banks conducting primarily credit card activities and banks chartered outside the continental USA. After cleaning the data we have an unbalanced panel with 2397 bank–year observations for 285 banks. We deflate all nominal stock variables to 2005 US dollars using the consumer price index (for all urban consumers).

We follow the commonly used ‘intermediation approach’ of Sealey and Lindley (1977) to model the bank’s production technology. According to this approach a bank’s balance sheet is assumed to capture the essential structure of its core business. Liabilities, together with physical capital and labor, are taken as inputs to the bank’s

production process, whereas assets (other than physical) are considered as outputs. Liabilities include core deposits and purchased funds; assets include loans and trading securities. We define the following outputs of the bank's production process: consumer loans ( $y_1$ ), real estate loans ( $y_2$ ), commercial and industrial loans ( $y_3$ ) and securities ( $y_4$ ). These output categories are essentially the same as those in Berger and Mester (1997, 2003). Following Hughes and Mester (1998, 2013), we further include off-balance-sheet income ( $y_5$ ) as an additional output. We also include bank's total non-performing loans ( $b$ ).

The variable inputs are labor, i.e. the number of full-time equivalent employees ( $x_1$ ), physical capital ( $x_2$ ), purchased funds ( $x_3$ ), interest-bearing transaction accounts ( $x_4$ ) and non-transaction accounts ( $x_5$ ). We also include financial capital (equity,  $e$ ) as an additional input to the production technology. We follow Berger and Mester (1997, 2003) and Feng and Serletis (2009) and assume that equity is a quasi-fixed input. The treatment of equity capital as an input to banking production technology is consistent with Hughes and Mester's (1993, 1998) argument that banks may use it as a source of funds and thus as potential protection against losses. We compute the prices of variable inputs ( $w_1$ – $w_5$ ) by dividing total expenses on each input by the corresponding input quantity. Table I in Malikov, Kumbhakar and Tsionas (2015) presents summary statistics of the data we use.<sup>7</sup>

## 8 Empirical results

We will consider several models to examine whether the new techniques perform better and / or provide a better description of the data.

Model I: a translog cost function model which depends on input prices, outputs, equity, non-performing loans and a time trend. The model does not allow for inefficiency. This model is used as a benchmark.

Model II: An FPA model with normal and half-normal distributions for the two error components of the model.

Model III: An FPA model with the following modification. The log variance of the two-sided error term is itself an FPA model and the log of the scale parameter of the half-normal distribution for the one-sided error term is also an FPA model.

Models IV, V and VI: As models I, II and III with the modification that we use ACD instead of FPA.

Model VII: A general SMR (or GSMR) model for the cost function. The probabilities and logs of the scale parameters of the normal and half-normal distributions for the two error components of the model are functions of logs of input prices, outputs, equity, non-performing loans and, finally, a time trend.

Additionally, we will also look closely into the models developed for panel data in sections 4.1, 4.2 and 4.3.

---

<sup>7</sup>The data are available in <http://qed.econ.queensu.ca/jae/datasets/malikov001/>

**Table 6. Model comparison for US banking data**

Notes: Model I: a translog cost function model which depends on input prices, outputs, equity, non-performing loans and a time trend. The model does not allow for inefficiency. This model is used as a benchmark.

Model II: An FPA model with normal and half-normal distributions for the two error components of the model.

Model III: An FPA model with the following modifications: The log variance of the two-sided error term is itself an FPA model and the log of the scale parameter of the half-normal distribution for the one-sided error term is also an FPA model.

Models IV, V and VI: As models I, II and III with the modification that we use ACD instead of FPA.

Model VII: A general SMR (or GSMR) model for the cost function. The probabilities and logs of the scale parameters of the normal and half-normal distributions for the two error components of the model are functions of logs of input prices, outputs, equity, non-performing loans and, finally, a time trend.

model	normalized LPS	model	order of Additive Cobb-Douglas									
translog	0.000		1	2	3	4	5	6	7	8	9	10
I	4.713	IV	0.027	0.514	1.102	1.514	2.217	3.345	4.919	4.102	3.213	2.131
II	17.091	V	1.072	3.210	5.103	7.154	9.782	14.151	8.201	7.332	4.202	2.550
III	27.616	VI	0.785	2.115	4.122	15.543	33.817	22.761	11.454	7.215	3.210	1.105
IV	4.919		base of Fractional Polynomial Approximation									
V	14.151	model	1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{7}$	$\frac{1}{8}$	$\frac{1}{9}$	$\frac{1}{10}$
VI	33.817	II	0.764	4.355	8.201	17.091	12.021	8.872	5.413	3.202	1.210	0.551
VII	29.715	III	8.337	15.201	27.616	13.201	8.045	4.317	1.414	0.881	0.224	0.038

From the results in the left of Table 6 it turns out that an ACD specification is better, followed by model VII (the full SMR which uses bases  $\frac{1}{3}$  for the cost function and  $\frac{1}{2}$  for the logs of scale parameters of the two error components; the weight function also turns out to be of base  $\frac{1}{2}$ ).<sup>8</sup>The right panel of Table 6 provides the normalized LPS for (i) the order of the ACD in the upper panel, and (ii) the base of FPA in the lower panel.

Given these results, it seems safe to proceed conditional on the choices of the particular order of ACD or base of the FPA for models IV, V, VI and II, III. Before doing so, it would be desirable to test the different models based on the alternative assumptions proposed in section 4, viz. panel data (section 4.1), endogeneity (section 4.2) and copulas (section 4.3). To summarize, in section 4.1 we introduce one-sided individual effects  $\lambda_i$  which are handled in a flexible way, see (33) and (34). In section 4.2 we capture the correlation between  $\mathbf{x}_{it}$  and  $\varepsilon_{it}$  using a panel VAR and a normal joint distribution where the elements of the Cholesky factor of the covariance matrix are flexible functions, see (35)-(37) and (38)-(39). Finally, in section 4.3 we have a copula dependence model, see (44) or (46). To save space we report results only for the best models in each of the three alternatives and we report the results in Table 7.

<sup>8</sup>For model VII we used a full comparison between FPA for the cost function, the two scale parameters and the weights in bases  $\frac{1}{k}$ ,  $k = 1, \dots, 10$ . This involved a choice among 10,000 different models so we give only the final best choice.



**Table 7. Alternative assumptions about panel data and endogeneity**

Results for models IV,V, VI and I, II are the same as in Table 6 to allow direct comparison. Models x.1, x.2 x.3 correspond to x=IV, V, V and I,II. Models of the form x.1 are as in section 4.1 where we introduce one-sided individual effects  $\lambda_i$  which are handled in a flexible way, see (33) and (34). Models of the form x.2 are as in section 4.2 where we capture the correlation between  $\mathbf{x}_{it}$  and  $\varepsilon_{it}$  using a panel VAR and a normal joint distribution where the elements of the Cholesky factor of the covariance matrix are flexible functions, see (35)-(37) and (38)-(39). Models of the form x.3 are as in section 4.3 where we have a copula dependence model, see (44) or (46). In models x.1 and x.2 the orders of the various flexible functions are optimized and the best results are provided. All model orders run from 1 to 10 and they are combined with different orders of the Additive Cobb-Douglas or the Fractional Polynomial Approximation for the functional form of the cost function. Precise model orders are available on request.

model	order of Additive Cobb-Douglas									
	1	2	3	4	5	6	7	8	9	10
IV	0.027	0.514	1.102	1.514	2.217	3.345	4.919	4.102	3.213	2.131
IV.1	0.011	0.124	0.332	0.444	0.512	0.633	0.782	0.981	1.106	1.225
IV.2	0.017	0.225	0.832	0.913	1.414	1.505	1.617	1.771	1.803	1.981
IV.3	0.021	0.043	0.055	0.077	0.082	0.093	0.225	0.301	0.514	0.623
V	1.072	3.210	5.103	7.154	9.782	14.151	8.201	7.332	4.202	2.550
V.1	0.561	0.602	0.731	0.855	1.051	1.118	1.245	1.302	1.515	1.777
V.2	1.202	1.281	1.300	1.452	1.551	1.617	1.713	1.802	1.903	2.001
V.2	0.038	0.061	0.082	0.171	0.185	0.202	0.302	0.425	0.552	0.633
VI	0.785	2.115	4.122	15.543	33.817	22.761	11.454	7.215	3.210	1.105
VI.1	0.034	0.082	0.171	0.253	0.348	1.414	1.761	1.822	1.873	1.903
VI.2	0.542	0.616	2.313	3.156	5.105	4.180	3.210	3.191	2.774	0.723
VI.3	0.021	0.032	0.043	0.050	0.072	0.084	0.078	0.063	0.051	0.040
	base of Fractional Polynomial Approximation									
model	1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{7}$	$\frac{1}{8}$	$\frac{1}{9}$	$\frac{1}{10}$
II	0.764	4.355	8.201	17.091	12.021	8.872	5.413	3.202	1.210	0.551
II.1	0.121	0.255	0.313	0.520	0.615	0.503	0.442	0.316	0.217	0.201
II.2	0.551	2.332	3.444	5.670	6.154	5.332	4.120	2.561	0.781	0.120
II.3	0.072	0.103	0.217	0.353	0.410	0.332	0.298	0.271	0.265	0.120
III	8.337	15.201	27.616	13.201	8.045	4.317	1.414	0.881	0.224	0.038
III.1	0.056	0.072	0.091	1.312	2.510	1.651	1.312	0.873	0.641	0.553
III.2	2.330	4.551	5.005	7.217	6.271	5.440	4.212	3.155	2.364	2.120
III.3	0.041	0.055	0.073	0.092	1.210	0.871	0.773	0.555	0.430	0.212

From the results in Table 7 it turns out that, at least in this empirical application, it is not necessary to resort to the flexible panel specification (section 4.1), panel VAR-based-endogeneity as in section 4.2 or copulas (section 4.3). In fact, copulas seem to perform worst. The flexible panel specification performs worst relative to the benchmark

models (II, III, IV or II,III). Assuming cost inefficiency to be flexible but time-invariant is quite restrictive and a panel VAR (*even with a quite flexible for the elements of the Cholesky factor of the covariance matrix  $\Sigma_{it}$* ) does not seem sufficient to handle the joint distribution of  $\mathbf{x}_{it}$  and  $\varepsilon_{it}$  relative to the benchmark SMR specification. This shows that the SMR and its substantial generalizations in this study, are quite capable of representing actual and possibly complicated features of the data, including endogeneity in particular. In fact, the dominance of models III and VI (with VI being the best) also implies that there is considerable *value added from modeling endogeneity*: The LPS of III and VI are 27.616 and 33.817, respectively, over a simple translog model. From Table 6, models that do not account for endogeneity (like model II) sometimes imply a sizable LPS (like 17.091) which is still 10 times smaller compared to the LPS of model III. This is, of course, an overwhelming difference implying that *endogeneity is critical in the performance of models in this empirical application*.

The copula specification seems to perform worst relative to the other two models. The bad performance of copulas is evidently as result of the more flexible specifications in SMR. The specification in section 4.2 (a panel VAR with a flexible parametrization for the elements of the Cholesky decomposition of the covariance matrix,  $\Sigma_{it}$ ) performs best suggesting that explicit modeling of  $\Sigma_{it}$  is critical. However, none of the models in section 4 performs better than the models presented in Table 6. This suggests that these extensions, although interesting for future applications, are complicated enough and over-parametrized that fail to improve the log-predictive score statistic.

To show the value added from accounting for endogeneity we can consider models I and II under the specifications in section 4.2 where the organizing principle is that of a panel vector autoregression (PVAR) with a joint distribution of  $\varepsilon_{it}$  and  $\xi_{it}$ . We take again the translog as a benchmark with a LPS of zero and the results are presented in Table 8. In case A we have a PVAR model with a general covariance matrix  $\Sigma_{it}$  for  $\varepsilon_{it}$  and  $\xi_{it}$ . In case B we restrict the covariance matrix not to depend on  $\mathbf{x}_{it}$ . In cases C and D we consider the general specification in (40) with a general and fixed covariance matrix respectively. In case D, although the covariance matrix is fixed, it differs among groups ( $g = 1, \dots, G_\zeta$ ).

**Table 8. Additional results for endogeneity**

Notes: FPA and ACD denote “Fractional Polynomial Approximation” and “Additive Cobb-Douglas” respectively. The results presented are for the best orders of FPA and ACD and are not reported to save space but are available on request. In case A we have a PVAR model with a general covariance matrix  $\Sigma_{it}$  for  $\varepsilon_{it}$  and  $\xi_{it}$ . In case B we restrict the covariance matrix not to depend on  $\mathbf{x}_{it}$ . In cases C and D we consider the general specification in (40) with a general and fixed covariance matrix respectively. In case D, although the covariance matrix is fixed, it differs among groups ( $g = 1, \dots, G_\zeta$ ).

	FPA	ACD
translog	0.000	0.00
A. Panel VAR, fixed $\Sigma_{it}$	2.415	3.342
B. Panel VAR, Joint distribution of $\varepsilon_{it}$ and $\xi_{it}$ , fixed $\Sigma_{it}$	3.032	3.430
C. Joint distribution of $\varepsilon_{it}$ and $\xi_{it}$ , fixed $\Sigma_{it}$	1.255	1.717
D. Joint distribution of $\varepsilon_{it}$ and $\xi_{it}$ , general $\Sigma_{it}$	1.121	1.713

From the results in Table 8, there are definitely gains from considering endogeneity relative to a simple translog, so endogeneity matters empirically. The issue is whether modeling endogeneity in this way is better when compared to the alternatives in Tables 7. As the highest LPS is 3.342 it turns out worse than an LPS of 4.919 for the FPA model IV in Table 7 and certainly LPS of 17.091 or 33.817 for models VI and III in the same Table. Therefore, at least in this empirical application, modeling endogeneity through a PVAR as in section 4.2 does not appear viable. *Although there is, clearly, considerable value added from modeling heterogeneity, it turns out PVAR models (section 4.2) turn out to be worst when compared to the alternative, more general and flexible models like II,III or IV,VI.*

The different models can be compared in terms of returns to scale (RTS), efficiency change (EC), productivity growth (PG) which is equal to technical change plus EC, and cost efficiency.

## Figure 1. Sample distribution of RTS

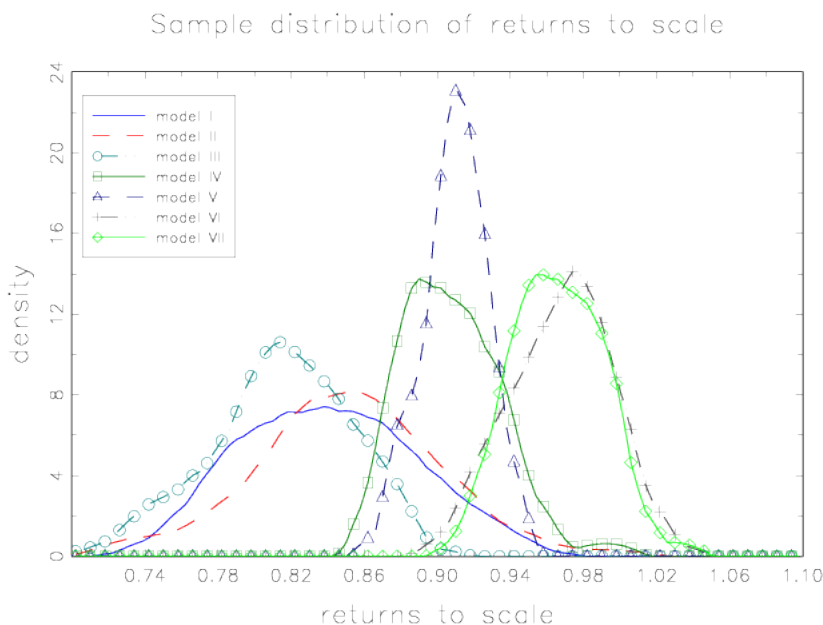
Notes: Returns to scale are computed as  $rts = 1 / \sum_{k=1}^K \frac{\partial \log C(p,y)}{\partial \log y_k}$ . The measure is averaged across all SMC draws to account for parameter uncertainty. Model I: a translog cost function model which depends on input prices, outputs, equity, non-performing loans and a time trend. The model does not allow for inefficiency. This model is used as a benchmark.

Model II: A FPA model with normal and half-normal distributions for the two error components of the model.

Model III: A FPA model with the following modification. The log variance of the two-sided error term is itself an FPA model and the log of the scale parameter of the half-normal distribution for the one-sided error term is also an FPA model.

Models IV, V and VI: As models I, II and III with the modification that we use ACD instead of FPA.

Model VII: A general SMR (or GSMR) model for the cost function. The probabilities and logs of the scale parameters of the normal and half-normal distributions for the two error components of the model are functions of logs of input prices, outputs, equity, non-performing loans and, finally, a time trend.



The good performance of models VI and VII in terms of LPS materializes in the fact that these models have quite different implications for RTS, PG and efficiency. From Figure 1, these models agree that RTS is close to 0.98 whereas the other models provide RTS measures from 0.82 to 0.90 -which is typical in banking studies employing “flexible” functional forms. The translog itself provides average RTS close to 0.82 and very small probability that RTS could be higher than 0.95.

From Figure 2, models VI and VII suggest that PG is very close to zero and very few banks have PG close to 1% or -1%. The other models provide, again, results which indicate that PG averages almost 4% and can be as large as 10% for certain banks. Models VI and VII are inconsistent with this prediction and suggest that PG, if any, is quite small.

## Figure 2. Sample distribution of PG

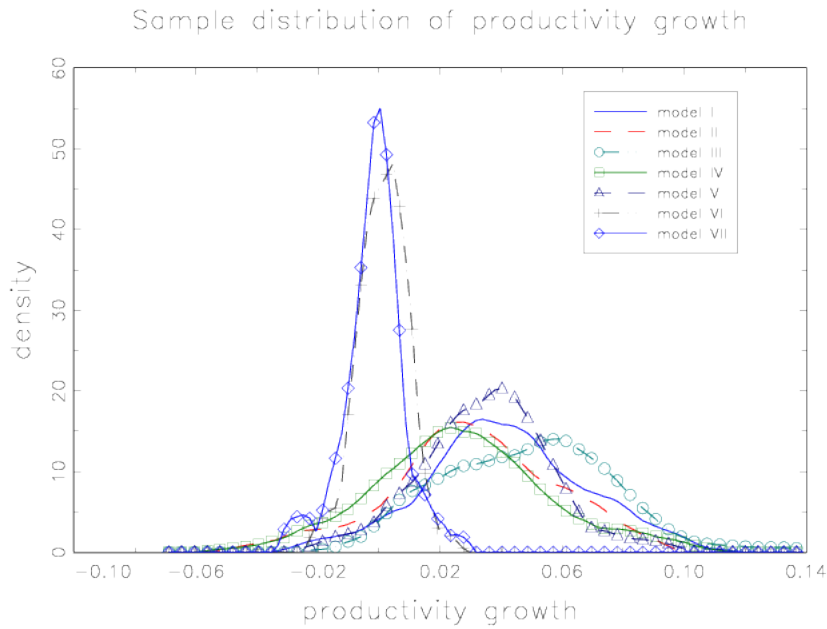
Notes: Model I: a translog cost function model which depends on input prices, outputs, equity, non-performing loans and a time trend. The model does not allow for inefficiency. This model is used as a benchmark.

Model II: A FPA model with normal and half-normal distributions for the two error components of the model.

Model III: A FPA model with the following modification. The log variance of the two-sided error term is itself an FPA model and the log of the scale parameter of the half-normal distribution for the one-sided error term is also an FPA model.

Models IV, V and VI: As models I, II and III with the modification that we use ACD instead of FPA.

Model VII: A general SMR (or GSMR) model for the cost function. The probabilities and logs of the scale parameters of the normal and half-normal distributions for the two error components of the model are functions of logs of input prices, outputs, equity, non-performing loans and, finally, a time trend.



### Figure 3. Sample distribution of technical efficiency

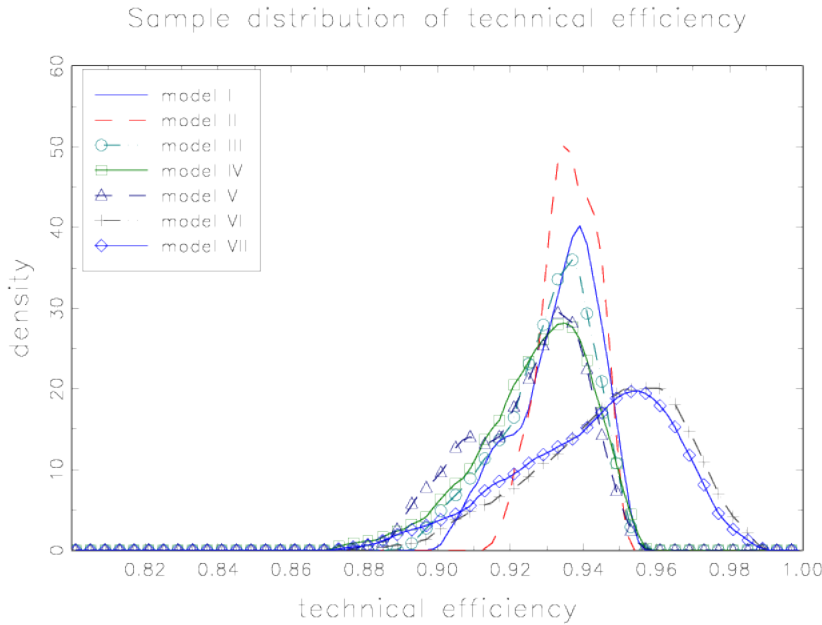
Notes: Model I: a translog cost function model which depends on input prices, outputs, equity, non-performing loans and a time trend. The model does not allow for inefficiency. This model is used as a benchmark.

Model II: An FPA model with normal and half-normal distributions for the two error components of the model.

Model III: An FPA model with the following modification. The log variance of the two-sided error term is itself an FPA model and the log of the scale parameter of the half-normal distribution for the one-sided error term is also an FPA model.

Models IV, V and VI: As models I, II and III with the modification that we use ACD instead of FPA.

Model VII: A general SMR (or GSMR) model for the cost function. The probabilities and logs of the scale parameters of the normal and half-normal distributions for the two error components of the model are functions of logs of input prices, outputs, equity, non-performing loans and, finally, a time trend.



Another difference is in terms of technical efficiency (see Figure 1). Models VI and VII suggest that cost efficiency averages 95%-96% but the sample distribution is highly skewed to the left. The remaining models suggest that cost efficiency averages 92% but the sample distributions are quite different. For example, the remaining models suggest that efficiency in excess of 96% is practically impossible, contrary to the predictions of models VI and VII. It should be mentioned that since one can put confidence intervals on efficiency estimates, 0.96 and 0.98 will probably be statistically the same.

In Figure 4 we examine the temporal pattern of PG. Again, models VI and VII suggest that PG has been fairly close to zero from 2001 to 2010. The other models provide quite different predictions with a generally declining pattern of PG but still positive and well above 3% on the average. Model IV even suggests that PG increased in the aftermath of the sub-prime crisis. In terms of efficiency change (Figure 5) models VI and VII suggest that there has been no serious change in cost efficiency contrary to the remaining models. Some models provide positive

estimates while other provide negative estimates for efficiency change so we have a “mixed bag” in this case, in terms of the temporal behavior of efficiency change. As the models are quite different and have different implications, this is not unexpected. The fact that models VI and VII suggest that PG has been fairly close to zero since 2001 is more in line with reality. In banking most innovations (like ATMs and electronic banking) have exhausted their productivity effects well before the 2000s as they have been introduced extensively by almost all banks. From that point of view it is not surprising to find that productivity growth has been quite low.

### Figure 4. Temporal behavior of PG

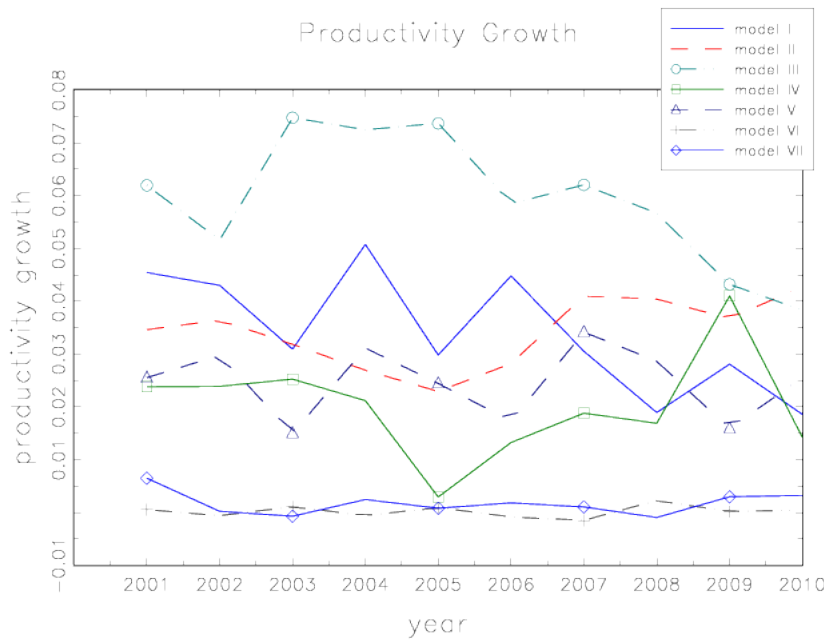
Notes: Model I: a translog cost function model which depends on input prices, outputs, equity, non-performing loans and a time trend. The model does not allow for inefficiency. This model is used as a benchmark.

Model II: A FPA model with normal and half-normal distributions for the two error components of the model.

Model III: A FPA model with the following modification. The log variance of the two-sided error term is itself an FPA model and the log of the scale parameter of the half-normal distribution for the one-sided error term is also an FPA model.

Models IV, V and VI: As models I, II and III with the modification that we use ACD instead of FPA.

Model VII: A general SMR (or GSMR) model for the cost function. The probabilities and logs of the scale parameters of the normal and half-normal distributions for the two error components of the model are functions of logs of input prices, outputs, equity, non-performing loans and, finally, a time trend.



## Figure 5. Cost efficiency change

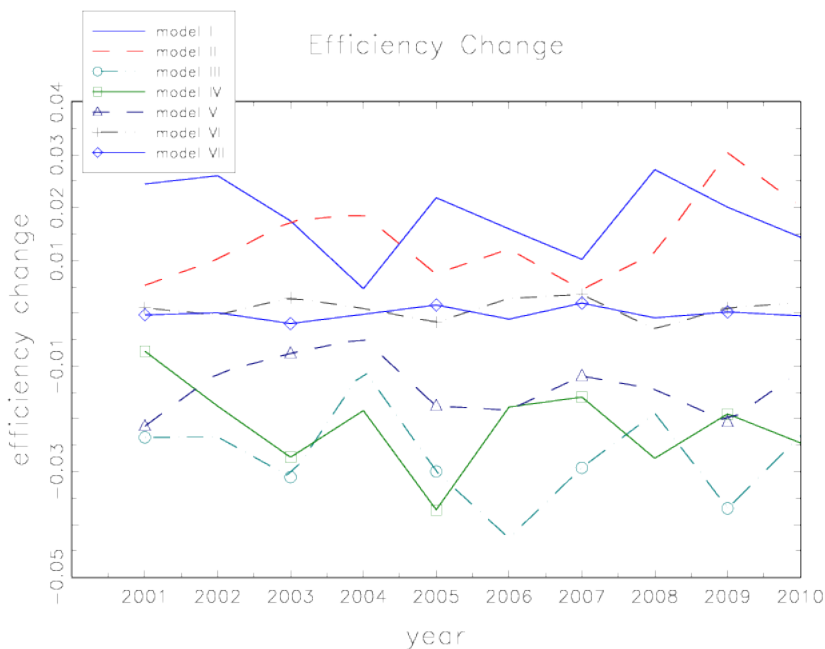
Notes: Model I: a translog cost function model which depends on input prices, outputs, equity, non-performing loans and a time trend. The model does not allow for inefficiency. This model is used as a benchmark.

Model II: A FPA model with normal and half-normal distributions for the two error components of the model.

Model III: A FPA model with the following modification. The log variance of the two-sided error term is itself an FPA model and the log of the scale parameter of the half-normal distribution for the one-sided error term is also an FPA model.

Models IV, V and VI: As models I, II and III with the modification that we use ACD instead of FPA.

Model VII: A general SMR (or GSMR) model for the cost function. The probabilities and logs of the scale parameters of the normal and half-normal distributions for the two error components of the model are functions of logs of input prices, outputs, equity, non-performing loans and, finally, a time trend.



Similar to productivity change is what we observe in terms of cost efficiency change. Most models agree that cost efficiency change has been quite low and in the neighborhood of  $\pm 2\text{-}3\%$  with the preferred models VI and VII showing no evidence of quantitatively important changes in cost efficiency. This challenges the conventional view that efficiency can adjust rapidly to changing economic conditions as, for example, in the aftermath or during the sub-prime crisis. On the contrary, conventional models like model I, deliver the empirical implication that efficiency has constantly improving at an average rate of 1,5-2% per year, which is hard to believe and must be attributed to the inflexibilities of the functional forms in this model.

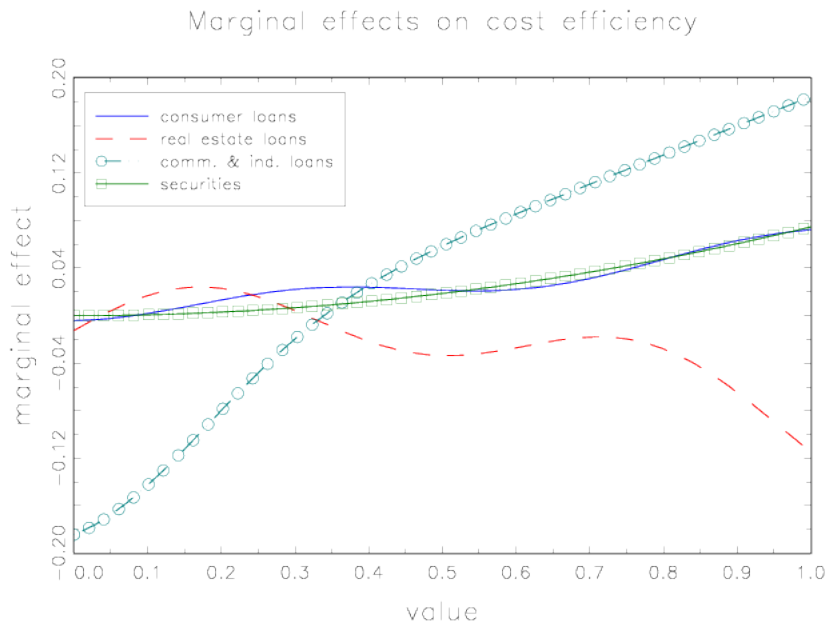
The marginal effects of several variables on cost efficiency are provided in Figures 6a through 6c. These marginal effects may be computed as  $\frac{\partial \log \hat{r}_{it}}{\partial w_{it}}$  where<sup>9</sup>  $\hat{r}_{it} = \exp(-\hat{u}_{it})$  and  $\hat{u}_{it}$  is estimated cost inefficiency. To approximate these effects numerically we use finite differences, viz.  $\frac{\Delta \hat{u}_{it}}{\Delta w_{it}}$ , where  $\Delta w_{it}$  is defined as  $\Delta w_{it} = w_{it} + \mathbf{h}$ , and  $\mathbf{h}$

<sup>9</sup>Here,  $w_{it}$  stands for the vector of variables that affect the ACD specification. We remind that most elements of  $w_{it}$  are already in logs.



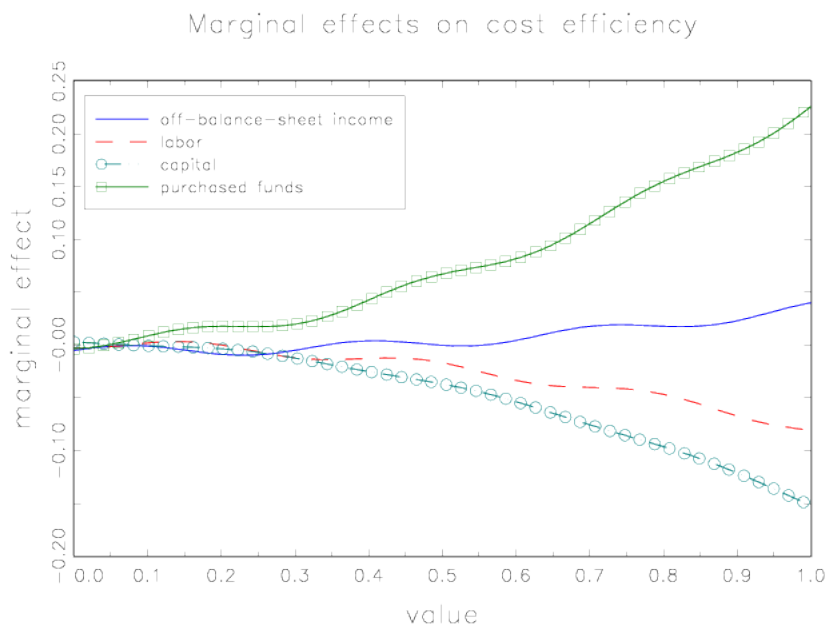
is a vector whose elements are 0.1 times the minimum absolute value of elements of  $w_{it}$ . The marginal effects are computed for each draw and they are averaged across all draws so that they fully account for parameter uncertainty. Finally, results are reported only for our preferred specification, viz. Model VI. For visual clarity all variables are normalized to lie in the interval  $[0, 1]$ .

**Figure 6a. Marginal effects on cost efficiency**



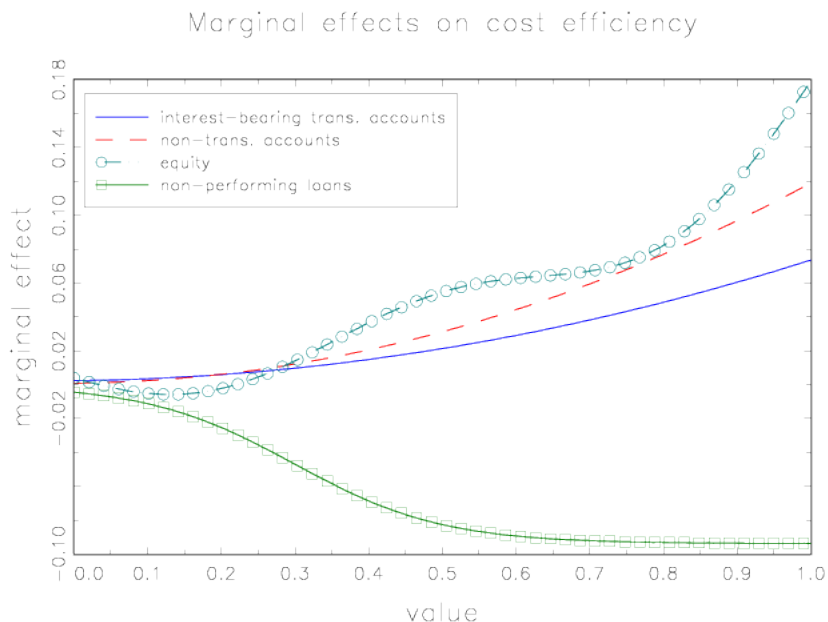
One of the most important technical aspects of our major extensions is the ability to deliver the marginal effects of certain variables on cost efficiency. The marginal effects are plotted in Figures 6a through 6c. The effects are highly nonlinear and do not always have the same sign through the entire (relevant) domain. The effect of consumer loans (Fig. 6a) is positive starting from about zero to finally obtain a value of 0.04. The effect of real estate loans is negative for the most part but positive when these loans are less than their 20% percentile. At the median their effect is -0.04 and increases to -0.12 at higher percentiles -a fact that can be attributed to the volatile character of real estate prices. Commercial and industrial loans have a monotonic impact on efficiency with their effect being negative for values less than the 35% percentile and they increase to reach a maximum of of 0.12. Finally, securities, have a consistently positive and monotonic effect which is close to 0.01 at the median and increases to 0.04 near the maximum.

Figure 6b. Marginal effects on cost efficiency



From Figure 6b, it turns out that labor and capital consistently decreases cost efficiency but purchased funds have a positive effect with a minor positive contribution from off-balance-sheet income. These results indicate that further expansion of the banks in terms of capital or labor cannot possibly increase the efficiency of their operations, although there are clearly ways to do so through re-balancing their activities in terms of scope. Purchased funds, for example, (Fig. 6b) as well as securities and consumer loans (Fig. 6a) can contribute to higher efficiency.

**Figure 6c. Marginal effects on cost efficiency**



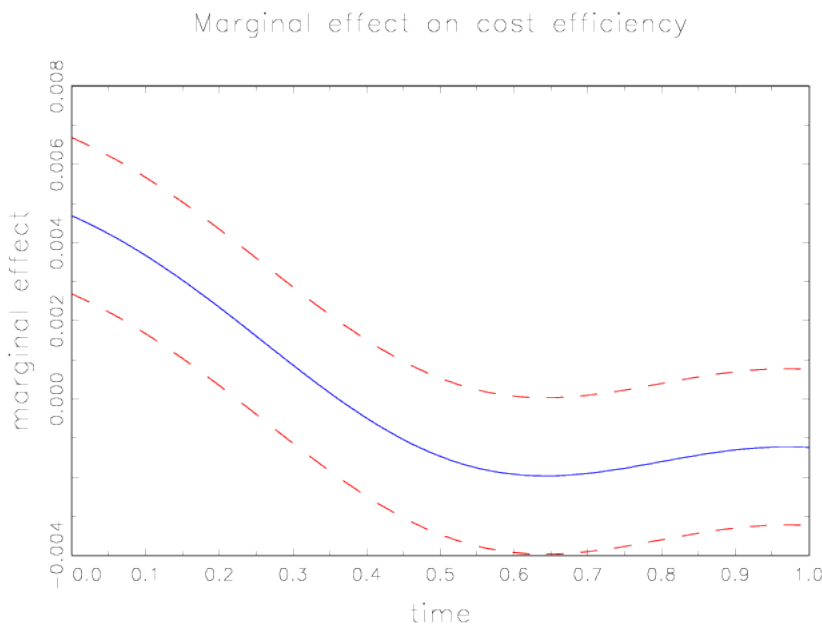
We can see how economies of scope in efficiency work further if we take a look at Figure 6c. Non-performing loans have a clear negative effect on cost efficiency (of nearly -0.10 at the median) while interest-bearing accounts, equity and non-transaction accounts have a positive effect (although non-monotonic) on efficiency. Non-monotonicity is particularly pronounced in the case of equity: Its effect “takes off” just above the 70% percentile although even the effect at the median is quantitatively important (nearly 0.05). In terms of policy implications these results are important. Cost efficiency can be improved through a shift from capital, labor and non-performing loans to equity, interest-bearing transaction and non-transaction accounts, purchased funds and, somewhat less importantly, to consumer loans. The use of commercial and industrial loans and other items is more ambiguous and depends on where a particular bank is, in terms of its current position in the (relevant) domain of this variable.

These results are not just descriptive although we deliberately considered the separate effect of each variable on technical efficiency. As a matter of fact, for each particular bank, we can consider in a formal way, the precise effect (along with 95% Bayes probability intervals) of changing the mix of outputs on its efficiency. In that way, we can design formally a particular scheme to increase cost efficiency by, say, three or four percentage points by keeping constant labor and capital and changing the mix of consumer versus industrial and commercial loans and securities. The flexibility of our new models, therefore, solves another outstanding problem, viz. that *researchers often can estimate efficiency but do not know what to do when it comes to specific proposal about increasing efficiency levels. Our results can be used to design formal schemes to achieve precisely this goal.*

Finally, in Figure 7 we report the marginal effect  $\frac{\partial \log \hat{r}_{it}}{\partial t}$  which we call the “technical change function”. First, the values of the function are quite close to zero, a fact that is consistent with previous results. Second, it seems that in the beginning of 2000s there was some positive, albeit quite small technical change, the general pattern shows a negative trend and a recovery took place in recent years but technical change in the banking sector is still close to zero and negative.

### Figure 7. Technical change function

Note: The dotted lines represent 95% Bayes probability intervals.



In view of the empirical fact that there has been immaterial technical change and productivity growth, it is clear that the banking sector can realize cost savings mainly through increasing its cost efficiency. As we have shown the matter is both quantitative and qualitative, in the sense that a new mix of outputs or re-balancing must be sought with specific changes in the mix that can be computed through an obvious generalization of marginal effects in the direction of changing several outputs at a time. Due to the flexibility of our functional forms and the many stages at which they enter (functional form, mixing weights, variances etc) there are rich patterns that can be modeled in a systematic and attractive manner. Indeed, one of the elusive goals of efficiency estimation so far, has been not only to estimate efficiency (although doing this in a flexible manner has proved to be quite challenging) but also to figure out what has to be done in order to realize cost savings in practice. The models in this paper are likely to contribute to this goal and open new avenues for further research to this important policy issue.

## Concluding remarks

We provide major extensions to smoothly mixing regressions and fractional polynomial approximations for both the functional form of the frontier and the structure of inefficiency. Moreover, we handle endogeneity using copulas thus providing reasonable answers to the problems raised by Stone (2002) and addressed in Badunenko, Henderson and Kumbhakar (2012). Our flexible model is based on nontrivial extensions of Smoothly Mixing Regressions (SMR, Geweke and Keane, 2007) and fractional polynomial approximations (FPA, Petrella and Geweke, 2014). Additionally we provide an alternative model which is based on a sum of Cobb-Douglas functions (ACD). In an empirical application to US banks it turns out that the ACD specification performs very well both in the frontier function and as a model for the scale functions of the two error components of the stochastic frontier model. The new procedures strike a good balance between traditional fully parametric stochastic frontier models and fully nonparametric estimation which has its own shortcomings when the sample sizes are small and / or the number of inputs and outputs is large. Our functional forms are flexible and regularity conditions can be imposed easily while endogeneity is also taken into account via copulas. We believe that the new techniques could become part of the toolbox of economists and engineers interested not only in measuring efficiency but also in examining how a set of environmental variables has an impact on inefficiency. This impact is also modeled in a flexible way within the SMR using FPA or ACD specifications.

## References

- Aigner, D., Lovell, C. K., P. Schmidt (1977). Formulation and estimation of stochastic frontier production function models. *J. Econometr.*, 6, 21–37.
- Badunenko, O., D. J. Henderson, S.C. Kumbhakar (2012). When, where and how to perform efficiency estimation. *J. R. Statist. Soc. A* 175, Part 4, pp. 863–892.
- Berger A.N., L.J. Mester (1997). Inside the black box: what explains differences in the efficiencies of financial institutions? *Journal of Banking and Finance* 21(7): 895–947.
- Berger A.N., L.J. Mester (2003). Explaining the dramatic changes in performance of US banks: technological change, deregulation, and dynamic changes in competition. *Journal of Financial Intermediation* 12(1): 57–95.
- Charnes, A., W.W. Cooper, E. Rhodes (1978) Measuring the efficiency of decision-making units. *Eur. J. Oper. Res.*, 84, 655–676.
- Chopin, N. (2002). A sequential particle filter method for static models, *Biometrika* 89 (3) 539-551.
- Danaher, P.J., M.S. Smith (2011). Modeling multivariate distributions using copulas: Applications in marketing. *Marketing Science*, 30 (1), 4-21.
- Daraio, C., A. Bonaccorsi, L. Simar (2015). Rankings and university performance: A conditional multidimensional approach, *European Journal of Operational Research* 244, 918-930.

Daraio, C., L. Simar (2014). Directional distances and their robust versions: Computational and testing issues, *European Journal of Operational Research* 237, 358-369.

Daraio, C., L. Simar (2007). *Advanced nonparametric methods in efficiency analysis: Methodology and applications*. Springer-Verlag.

Durham, G., J. Geweke (2013). Adaptive sequential posterior simulators for massively parallel computing environments. working paper.

Fan, Y., Q. Li, A. Weersink (1996) Semiparametric estimation of stochastic production frontier models. *J. Bus. Econ. Statist.*, 14, 460–468.

Feng, G., A. Serletis (2009). Efficiency and productivity of the US banking industry, 1998–2005: evidence from the Fourier cost function satisfying global regularity conditions. *Journal of Applied Econometrics* 24(1): 105–138.

Gelfand, A., D. Dey, H. Chang (1992). Model determination using predictive distributions with implementation via sampling-based methods, in *Bayesian Statistics*, ed. by J. Bernardo, J. Berger, A. Dawid, and A. Smith, Oxford University Press, vol. 4, 147-168.

Geweke, J., M. Keane (2007). Smoothly mixing regressions, *Journal of Econometrics*, 138, 252-290.

Hughes J.P., L.J. Mester (1993). A quality and risk-adjusted cost function for banks: evidence on the ‘too-big-to-fail’ doctrine. *Journal of Productivity Analysis* 4(3): 293–315.

Hughes J.P., L.J. Mester (1998). Bank capitalization and cost: evidence of scale economies in risk management and signaling. *Review of Economics and Statistics* 80(2): 314–325.

Jondrow, J., C.A. Knox Lovell, I. S. Materov, P. Schmidt (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics* 19, 233–238.

Kneip, A., L. Simar, L., P.W. Wilson (2008) Asymptotics for DEA estimators in non-parametric frontier models. *Econometric Theory* 24, 1663–1697.

Kumbhakar, S. C., C.A.K. Lovell (2003). *Stochastic Frontier Analysis*. Cambridge: Cambridge University Press.

Kutlu, L. (2010). Battese-coelli estimator with endogenous regressors. *Economics Letters* 109, 79-81.

Malikov, E., S.C. Kumbhakar, M.G. Tsionas (2015). A cost system approach to the stochastic directional technology distance function with undesirable outputs: the case of US banks in 2001–2010. *Journal of Applied Econometrics*, in press.

Meeusen, W., J. van den Broeck (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *Int. Econ. Rev.*, 18, 435–444.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica* 46:69-85

Norets, A., J. Pelenis (2011). Bayesian modeling of joint and conditional distributions, working paper, Princeton University and Institute of Advanced Studies, Vienna.

Perelman, S., D. Santin. On the generation of a regular multi-input multi-output technology using parametric output distance functions, working paper, <http://www2.ulg.ac.be/crepp/papers/crepp-wp200506.pdf>

Petrella, L., J. Geweke (2014). Likelihood-based inference for regular functions with fractional polynomial approximations, *Journal of Econometrics*, 183 (1), 22-30.

Pu, D., C.F. Parmeter, J.S. Racine (2013). Nonparametric Kernel Regression with Multiple Predictors and Multiple Shape Constraints, *Statistica Sinica* 23, 1347-1371

Sealey C.W., J.T. Lindley (1977). Inputs, outputs, and a theory of production and cost at depository financial institutions. *Journal of Finance* 32(4): 1251–1266.

Simar, L. P. W. Wilson (2007). Estimation and inference in two-stage semi-parametric models of production, *Journal of Econometrics* 136, 31-64.

Simar, L., A. Vanhems (2012). Probabilistic characterization of directional distances and their robust versions, *Journal of Econometrics* 166, 342-354.

Simar, L., A. Vanhems, P.W. Wilson (2012). Statistical inference for DEA estimators of directional distances, *European Journal of Operational Research* 220, 853-864.

Sklar, A. (1959), Fonctions de repartition a n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Universite de Paris*, 8:229-231.

Sauerbrei W., P. Royston (1999). Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society A* 162 part 1, 71-94.

Stone, M. (2002). How not to measure the efficiency of public services (and how one might), *Journal of the Royal Statistical Society, Series A* 165(3), 405–434.

Tran, K.C., E.G. Tsionas (2013). GMM estimation of stochastic frontier model with endogenous regressors. *Economics Letters* 118, 233-236.

Villani, M., R. Kohn, P. Giordani (2009). Regression density estimation using smooth adaptive Gaussian mixtures, *Journal of Econometrics* 153, 155-173.

Villani, M., R. Kohn, D.J. Nott (2012). Generalized smooth finite mixtures, *Journal of Econometrics* 171, 121-133.

## APPENDIX

Chopin (2002) recommends the independence Metropolis algorithm to select the kernel, which requires a source distribution. A possible choice, if we sampled from  $p_n$  ( $n < T$ ), with respect to  $p_{n+s}$  is  $\mathcal{N}(\hat{E}_{n+s}, \hat{V}_{n+s})$  where

$$\hat{E}_{n+s} = \frac{\sum_{j=1}^J w_j \theta_j}{\sum_{j=1}^J w_j}, \hat{V}_{n+s} = \frac{\sum_{j=1}^J w_j (\theta_j - E_{n+p}) (\theta_j - E_{n+p})'}{\sum_{j=1}^J w_j}.$$

The strategy can be parallelized easily. If  $K$  processors are available, we can partition the particle system into  $K$  subsets, say  $S_k, k = 1, \dots, K$ , and implement computations for particles of  $S_k$  in processor  $k$ . The algorithm

can deal with new data at a nearly geometric rate and therefore the frequency of exchanging information between processors (after reweighting) decreases at a rate exponential to  $n$ , which is highly efficient.

Resampling according to  $\theta_j^m \sim K_t(\theta_j^r, \cdot)$  reduces particle degeneracy (Gilks and Berzuini, 2001) since identical replicates of a single particle are replaced by new ones without altering the stationary distribution. For the application in the next section, using  $J = 2^{12}$  particles gave a mean squared error in posterior means of  $10^{-5}$  over 100 runs.

Chopin (2004) introduces a variation of MSC in which the observation dates at which each cycle terminates (say  $t_1, \dots, t_L$ ) and the parameters involved in specifying the Metropolis updates (say  $\lambda_1, \dots, \lambda_L$ ) are specified. Therefore,  $0 = t_0 < t_1 < \dots < t_L = T$  and we have the following scheme (we rely heavily on Durham and Geweke, 2013).

**Step 1.** Initialize  $l = 0$  and  $\theta_{jn}^{(l)} \sim p(\theta)$ ,  $j \in \mathcal{J}, n \in \mathcal{N}$ .

**Step 2.** For  $l = 1, \dots, L$ :

(a) **Correction phase:**

(i)  $w_{jn}(t_{l-1}) = 1$ ,  $j \in \mathcal{J}, n \in \mathcal{N}$

(ii) For  $s = t_{l-1} + 1, \dots, t_l$

$$w_{jn}(s) = w_{jn}(s-1)p(y_s|y_{1:s-1}, \theta_{jn}^{(l-1)}), j \in \mathcal{J}, n \in \mathcal{N}.$$

(iii)  $w_{jn}^{(l-1)} := w_{jn}(t_l)$ ,  $j \in \mathcal{J}, n \in \mathcal{N}$ .

(b) **Selection phase**, applied independently to each group  $j \in \mathcal{J}$  : Using multinomial or residual sampling based on  $\{w_{jn}^{(l)}, n \in \mathcal{N}\}$ , select

$$\{\theta_{jn}^{(l,0)}, n \in \mathcal{N}\}$$

from  $\{\theta_{jn}^{(l-1)}, n \in \mathcal{N}\}$ .

(c) **Mutation phase**, applied independently across  $j \in \mathcal{J}, n \in \mathcal{N}$  :

$$\theta_{jn}^{(l)} \sim p(\theta|y_{1:t_l}, \theta_{jn}^{(l,0)}, \lambda_l) \tag{A.1}$$

where the drawings are independent and the pdf above satisfies the invariance condition:

$$\int_{\Theta} p(\theta|y_{1:t_l}, \theta^*, \lambda_l)p(\theta^*|y_{1:t_l})d\nu(\theta^*) = p(\theta|y_{1:t_l}). \tag{A.2}$$

**Step 3.**  $\theta_{jn} := \theta_{jn}^{(l)}$ ,  $j \in \mathcal{J}, n \in \mathcal{N}$ .

At the end of every cycle, the particles  $\theta_{jn}^{(l)}$  have the same distribution  $p(\theta|y_{1:t_l})$ . The amount of dependence within each group depends upon the success of the Mutation phase which avoids degeneracy.