Smooth Approximations to Monotone Concave Functions in Production Analysis: An Alternative to Nonparametric Concave Least Squares

Mike G. Tsionas, Marwan Izzeldin

PII: S0377-2217(18)30472-7
DOI: 10.1016/j.ejor.2018.05.053
Reference: EOR 15166

To appear in: *European Journal of Operational Research*

Received date: 7 November 2017
Revised date: 25 May 2018
Accepted date: 26 May 2018

Highlights

- Smooth alternative to nonparametric segmented concave least squares.
- We use a differentiable approximation using smoothly mixing Cobb-Douglas anchor functions.
- Bayesian techniques organized around Markov Chain Monte Carlo.
- Approximation properties investigated with Monte Carlo experiment.
- Applications to a large US banking data set and global banking data.

Smooth Approximations to Monotone Concave Functions in Production Analysis: An Alternative to Nonparametric Concave Least Squares

Mike G. Tsionas* & Marwan Izzeldin†

June 22, 2018

Abstract

Estimation of banking efficiency and productivity is essential for regulatory purposes and for testing various theories in the context of banking such as the quiet life hypothesis, the bad management hypothesis etc. In such studies it is, therefore, important to place as few restrictions as possible on the functional forms subject to global satisfaction of the theoretical properties relating to monotonicity and concavity. In this paper we propose an alternative to nonparametric segmented concave least squares. We use a differentiable approximation to an arbitrary functional form based on smoothly mixing Cobb-Douglas anchor functions over the data space. Estimation is based on Bayesian techniques organized around Markov Chain Monte Carlo. The approximation properties of the new functional form are investigated in a Monte Carlo experiment where the true functional form is a Symmetric Generalized McFadden. The new techniques are applied to a large U.S banking data set as well as a global banking data set.

Keywords: OR in Banking; Simulation; Nonparametric Concave Least Squares; Segmented Least Squares; Bayesian analysis.

*Corresponding author. Lancaster University Management School, LA1 4YX, U.K., and Athens University of Economics and Business, Greece, m.tsionas@lancaster.ac.uk

†Lancaster University Management School, LA1 4YX, U.K., m.izzeldin@lancaster.ac.uk

1 Introduction

The problem of estimating monotone concave (or convex) functions without imposing parametric assumptions has a long history and is especially important in production economics. Monotonicity and curvature are standard properties of production, cost or profit functions implied by economic theory. In practice, estimating monotone concave functions from noisy data sets has proven to be a difficult task. Even parametric functional forms that are widely used in applied production analysis, like the translog, often suffer from violations of monotonicity and / or curvature and imposing such restrictions is a difficult task. Even when such restrictions are imposed, it is questionable whether the flexibility properties are still in place.

Concave Least Squares (CLS) is a non-parametric technique that has been used extensively in practice to estimate production or utility functions, see for example Allon, Beenstock, Hackman, Passy, & Shapiro (2007), Keshvari & Kuosmanen (2013) and Kuosmanen & Kortelainen (2012). Kuosmanen (2008) provides a quadratic programming formulation for CLS and proofs of its properties. Keshvari (2017) proposes an ingenious formulation for the segmented concave least squares (SCLS) problem using a novel nonparametric piecewise linear regression formulation. The SCLS estimator is defined as the CLS estimator under a cardinality constraint on the number of linear segments and shares common properties with least squares residuals, in particular orthogonality of errors and regressors. In addition, SCLS automatically classifies observations into different groups as shown in a hedonic pricing application by Keshvari (2017).

In this paper we propose an alternative to SCLS. Specifically we use a smooth mixture of anchoring Cobb-Douglas production functions to approximate the true but unknown monotone concave function in a smooth fashion. Smoothness may be desirable in practical applications in addition to monotonicity and curvature restrictions. The new model (Smooth SCLS) can be estimated using maximum likelihood or Bayesian techniques organized around Markov Chain Monte Carlo (MCMC). We use the Bayesian approach as standard algorithms for maximum likelihood can get stuck in local optima of the likelihood. We validate the new approach through a Monte Carlo experiment where the true but unknown functional form is a member of the Symmetric Generalized McFadden form (Diewert and Wales, 1987). The function is globally concave. Additionally, we provide two applications; one to a large U.S banking data set and another on global banking.

Estimation of inefficiency, productivity, and returns to scale in banking is traditionally based on parametric models belonging to flexible functional forms. Parametric models suffer from important drawbacks when it comes to fit in real data sets. Different approaches to flexible functional forms have been proposed by Michaelides et al. (2010, 2015). A common problem in flexible functional forms, is the violation of theoretical of monotonicity and curvature. There are numerical (Gallant and Golub, 1984) and Bayesian techniques (O'Donnell and Coelli, 2005) for imposing curvature conditions that can be used in practice. However, in production it often makes sense to estimate functional forms that satisfy globally the curvature conditions dictated by economic theory. This has been "one of the most vexing problems applied economists have encountered in estimating flexible functional forms" (Diewert and Wales, 1987) and remains "one of the most difficult challenges faced by empirical economists" (Terrell, 1996) as "[u]ltimately, the biggest challenge for researchers remains the issue of the appropriate production function specification to represent the underlying process technology" (Vaneman and Triantis, 2007). In the past many authors argued that the Fourier flexible function is more appropriate in banking (e.g. McAllister and McManus, 1993), since it is more flexible and fits the data better. However, Berger and Mester (1997) do not find significant differences in terms of efficiency.

Although the importance of functional form specification is important in all empirical fields, in banking in particular, its importance is related to significant policy issues. It is widely thought that market power and market concentration can limit bank efficiency and financial deepening. According to the quiet life hypothesis, banks enjoy monopoly profits at the expense of efficiency, so more competitive banking sectors would lead to more inefficiency. However, based on the efficient structure hypothesis, more efficient firms have lower costs, which in turn lead to higher profits. Moreover, according to the competition-fragility hypothesis, more bank competition (less market power) results in decrease of profit margins, and reduced bank risk taking. On the other hand, banks lagging behind in terms of efficiency are likely to have different risk characteristics as a result. In the next section we expand more on these issues.

2 Issues in banking

There are many reasons why technical efficiency estimation is important in banking. With deregulation changes in institutional frameworks, and increased competition adoption of "best practices" becomes all the more important (Goddard et al., 2007 and ECB, 2003). There is a large number of studies on bank competition, bank risk and efficiency (e.g. (e.g. Boyd and De Nicolo, 2003; Cihak, Schaeck and Wolfe, 2009, Casu and Girardone, 2009). It is possible, for example, that more bank risk may signal a deterioration in cost efficiency. Conversely, lower efficiency may force banks to take excessive risks. These arguments are particularly relevant especially after the sub-prime crisis where excessive bank risks and non-performing loans have become a central issue. For example, highly capitalized banks *may* be more efficient and less willing to take excessive risks in which case regulations should aim at better structured capitalization.

Hughes and Mester (1998, 2009) have emphasized that capitalization and bank risk may be related to bank efficiency. For instance, efficient banks may have more capital leverage or a more flexible risk profile while less efficient banks with low capitalization may be prone to taking higher risk due to lost returns and moral hazard problems. For more efficient banks the supervising authorities may also be willing to allow more flexibility and exercise less regulation. The fact that U.S bank efficiency and capital are relevant determinants of bank risk has been shown in Berger and De Young (1997) as well as Kwan and Eisenbeis (1997).

For European banks some authors do not find a positive relationship between inefficiency and bank risk-taking. This means that more inefficient European banks appear more capitalized and take on less risk, see Altunbas et al., (2007) and Williams (2004). Indirectly, this would provide some support for the quiet-life-hypothesis. Of course, there are different rationalizations for conflicting empirical findings as there are several theories.

The "cost skimping" hypothesis assumes that there is a trade-off between cost efficiency and future risk-taking due to moral hazard considerations. According to the "bad luck" hypothesis unexpected shocks can increase non-performing loans irrespective of managerial effiency and /or or their risk-taking profiles. According to the The 'moral hazard' hypothesis there is a negative relationship between capitalization and banking risk pas banks may have incentives to become more risky when capitalization is low or banks are les efficient. Informational frictions or agency problems may aggravate these effects.

According to the "bad management" hypothesis of Berger and DeYoung (1997) and Williams (2004), inefficient banks will have higher costs due to problematic credit monitoring and / or limited control of operating expenses. Inefficiency will precede increases in bank risk due to credit, reputational, operational, and / or other problems.

Additional capital buffers forced by regulatory authorities may force banks to take more risk (Hellman, Murdock and Stiglitz, 2000) but at the same time become less efficient (divert resources from quiet-life "buffer stocks") or more efficient (to counteract the regulatory measures). These actions relating to bank risk and efficiency are not, in general, independent as risk profiles and efficiency are related through diversion of real resources from certain uses to others. Fiordelisi et al. (2010) in a study of European banks summarize their findings as follows: "[L]ower efficiency scores (either cost or revenue) suggest greater future risks and efficiency improvements tend to shore up banks' capital positions. Our findings also emphasize the importance of attaining long-term efficiency gains to support financial stability objectives." (p. 24).

From a regulatory perspective the implications are clear. As efficiency causes risk -an implication of the "bad management" and "moral hazard" hypotheses. Specifically, more efficient banks eventually have better capitalizations and better capitalization leads to improvements in efficiency. As inefficiency also impproves risk according to the "bad management" hypothesis, better managerial practices and a re-organization of the internal structure of the banking production may lead to less risk and, at the same, a reduction in waste that can support better capitalization.

However, measures of efficiency tend to be sensitive to functional forms and inefficiency estimates cannot be interpreted properly when the theoretical restrictions are not satisfied globally. A related issue is that banking productivity seems to be high preceding the sub-prime crisis. Productivity estimates are also sensitive to the choice of functional form and can be meaningless if the theoretical restrictions are not satisfied globally. This major problem can be solved if we can represent banking production in a non-parametric way which enforces the restrictions globally and can provide inefficiency estimates. This difficult problem can be solved using a differentiable approximation to an arbitrary functional form based on smoothly mixing Cobb-Douglas anchor functions over the data space. This can be thought of as an alternative to Segmented Concave Least Squares, a technique that we describe in the next section.

3 Segmented Concave Least Squares

Keshvari (2017) provided an ingenious nonparametric concave least squares estimator. The technique fills the gap between OLS and CLS by estimating a concave linear function in which the number of linear segments can be controlled. Suppose we have a response variable y_i and a set of covariates $\mathbf{x}_i = [x_{i1}, ..., x_{im}]'$. The regression problem is $y_i = f(\mathbf{x}_i) + \varepsilon_i$ where the error has zero mean and is homoskedastic. The CLS problem is the following:

$$\begin{gathered} \min \tfrac{1}{2} \textstyle\sum_{i=1}^{n} \varepsilon_i^2, \\ \text{s.t. } f \in \mathcal{F}, \end{gathered} \tag{1}$$

where $\mathcal{F}$ is the space of continuous and concave functions from $\Re^K$ to $\Re$. It is known that the fitted values of (1) are unique and they are obtained from a piecewise linear function (Kuosmanen, 2008) via the following QP problem:

$$\begin{gathered} \min \textstyle\sum_{i=1}^{n} \varepsilon_i^2, \\ y_i = \zeta_i + \mathbf{x}_i' \psi_i + v_i, \ i = 1, ..., n, \\ \zeta_i + \mathbf{x}_i' \psi_i \leq \zeta_j + \mathbf{x}_j' \psi_j, j = 1, ..., n. \end{gathered} \tag{2}$$

As Keshvari (2017) mentioned there are n regression hyperplanes but some of them may be equivalent to each other. Hence the number of separate hyperplanes is $n_o \leq n$. Available QP solvers handle this problem, however, since the number of constraints is of order $O(n^2)$ and grows rapidly the solution time may be excessive or the solution process may stop prematurely before an optimal solution is obtained. In turn Keshvari (2017) proposed SCLS where the number of hyperplanes n_o is at most M:

$$\begin{gathered} \min \textstyle\sum_{i=1}^{n} \varepsilon_i^2, \\ y_i = \zeta_i + \mathbf{x}_i' \psi_i + v_i, \ i = 1, ..., n, \\ \zeta_i + \mathbf{x}_i' \psi_i \leq \zeta_j + \mathbf{x}_j' \psi_j, j = 1, ..., n, \\ \mathrm{card}(Q) \leq M, \end{gathered} \tag{3}$$

where $Q = \{q_1, ..., q_n\}$ is the set of unique hyperplanes and the last constraint implies that they are at most M. There are other published articles that propose methods to reduce the solution time of CLS (as it is cited in Keshvari 2017), such as Hannah and Dunson (2013), Keshvari (2017), Lee, Johnson, Moreno-Centeno, and Kuosmanen (2013).

Finally, the problem can be solved using mixed integer programming:

$$\begin{gathered} \min \tfrac{1}{2} \textstyle\sum_{i=1}^{n} \varepsilon_i^2 \\ \text{s.t.} \\ y_i = \alpha_h + x_i' \beta_h + \delta_{ih}, \ i \in J, \ h \in \mathcal{K}, \\ \delta_{ih} - \varepsilon_i \leq 0, \ i \in J, \ h \in \mathcal{K}, \\ \varepsilon_i - \delta_{ih} \leq (1 - \eta_{ih}) B, \ i \in J, \ h \in \mathcal{K}, \\ \textstyle\sum_{h=1}^{k} \eta_{ih} = 1, \ i \in J, \\ \eta_{ih} \in \{0, 1\}, i \in J, \ h \in \mathcal{K}, \end{gathered} \tag{4}$$

where B is a big positive number, η_{ih} ($i \in J = \{1, ..., n\}$, $h \in \mathcal{K} = \{1, ..., M\}$) are binary variables. If $\mathcal{H}_M = \{p_1, ..., p_M\}$ is the set of hyperplanes in the optimal solution of (3) then the binary variables reveal which hyperplanes in $\mathcal{H}_M$ are the same with which hyperplanes in Q. A hyperplane p_m is associated with a pair (α_m, β_m) and δ_{ih} is the distance of an observation to the hyperplane. The first constraint of (4) defines M regression hyperplanes, the next two impose concavity and the last provides that each fitted value is located on one hyperplane.

4 Smooth SCLS

In a general setting we can use a Cobb-Douglas anchor to come up with flexible functional forms for production functions. For $\mathbf{x} \in \Re^m$ we can use:

$$f(\mathbf{x};\alpha) = \alpha_0 + \sum_{k=1}^{m} \alpha_k x_k, \tag{5}$$

provided x_ks are in logs for simplicity in notation and $f(\mathbf{x};\alpha)$ denotes the log of the production function. The function $\exp\{f(\mathbf{x};\alpha)\}$ is automatically monotonic if $\alpha_j > 0,\ j = 0, 1, ..., K$ and concave provided $\alpha_j < 1,\ j = 1, ..., K$. Therefore. monotonicity and concavity can be enforced easily. Of course, the problem is that the Cobb-Douglas functional form is not flexible. However, it can be flexible enough locally, that is at particular regions of the data space.

Our approach to obtain a nonparametric version of (4) and adapt to different subsets of $\mathfrak{X}$ is described next. Suppose we have a sample $\{y_i, \mathbf{x}_i,\ i = 1, ..., n\}$ and $\mathfrak{X} = \{\mathbf{x}_i, i = 1, ..., n\}$. We write the model in the form of a usual regression:

$$y_i = \mathbf{x}_i'\beta + \varepsilon_i,\ i = 1, ..., n. \tag{6}$$

To turn the model into semi-parametric we use:

$$y_i = \sum_{g=1}^{G} (\mathbf{x}_i'\beta_g)\, p_g(\mathbf{x}_i;\gamma_g) + \varepsilon_i,\ i = 1, ..., n, \tag{7}$$

where G is the number of groups in the data (analogous to M in Keshvari, 2017), $\beta_g \in \Re^K$ ($K = m+1$) denotes group-specific parameter vectors and $p_g(\mathbf{x}_i;\gamma_g)$ are weights applied to each observation $\mathbf{x}_i \in \mathfrak{X}$. These weights are parametrized as follows:

$$p_g(\mathbf{x}_i;\gamma_g) = \frac{\exp\{\mathbf{x}_i'\gamma_g\}}{\sum_{g'=1}^{G} \exp\{\mathbf{x}_i'\gamma_{g'}\}},\ i = 1, ..., n,\ g = 1, ..., G. \tag{8}$$

For the error term we assume $\varepsilon_i \sim i.i.d.\mathcal{N}(0, \sigma^2)$ although many other alternatives are possible. In particular, we assume heteroskedasticity of the following form:

$$\sigma_i^2 = \sum_{g=1}^{G'} \exp\{\mathbf{x}_i'\zeta_g\}\, p_g(\mathbf{x}_i;\psi_g), \tag{9}$$

where $\zeta_g, \psi_g \in \Re^K$ are parameters. Therefore, we assume that the variance also has the same functional form as the regression function in (7) but the number of groups G' can be potentially different than G. The purpose of introducing flexible heteroskedasticity as in (9) is to reduce the approximation error which can thought of as the error in (7) and distinguish between a systematic component that depends on the regressors and a "pure error" of approximation and / or noise.

The model is not unlike an artificial neural network (ANN):

$$y_i = \sum_{g=1}^{G} \delta_g \varphi(\mathbf{x}_i'\lambda_g) + \varepsilon_i,\ i = 1, ..., n, \tag{10}$$

where δ_gs are parameters and $\lambda_g \in \Re^K$ (for each $g = 1, ..., G$). Typically, $\varphi(z) = \frac{1}{1+\exp(-z)}$, $z \in \Re$. However, Smooth SCLS uses explicitly a Cobb-Douglas anchor to impose monotonicity and concavity whereas this is burdensome in a standard ANN. Global approximation properties of ANNs have been established in Hornik (1991), Hornik et al. (1989) and White (1989).

Before proceeding it may be instructive to examine how the present model improves on existing alternatives. First, in nonparametric regression the researcher adopts a model of the form: $y_i = \varphi(\mathbf{x}_i) + v_i$, $i = 1, ..., n$, where $x_i \in \Re^M$ and, typically, $v_i \sim i.i.d.\mathcal{N}(0, \sigma^2)$. The functional form $\varphi(\cdot)$ is unknown but can be estimated using kernel techniques. Since $\varphi(\mathbf{x}) = E(y|\mathbf{x} = \mathbf{x}_i)$ we have $\varphi(\mathbf{x}) = \frac{\int yp(\mathbf{x},y)dy}{f(\mathbf{x})}$, where $f(\mathbf{x})$ is the marginal density of $\mathbf{x}$. Kernel regression can be implemented using the Nadaraya (1964) and Watson (1964) approach which delivers: $\varphi(\mathbf{x}) = \frac{(nh)^{-1}\sum_{j=1}^{n} K_h(x-x_i)y_j}{(nh)^{-1}\sum_{j=1}^{n} K_h(x-x_i)}$, where $K_h(\cdot)$ is a kernel function and h is a bandwidth parameter. The imposition of monotonicity and concavity is very hard in nonparametric estimation.

Local likelihood estimation (Tibshirani and Hastie, 1987) avoids some of these problems. In the model: $y_i = \mathbf{x}_i'\beta + \varepsilon_i$ where $\varepsilon_i \sim i.i.dN(0, \sigma^2)$, $i = 1, ..., n$, the parameters are approximated using local polynomials. Typically, local linear estimation is used which means that the researcher uses: $\beta(\mathbf{x}) = \gamma_o + \Gamma(\mathbf{x} - \mathbf{x}_i)$, where$\gamma_o$ is $M \times 1$ and Γ is $M \times M$. These are unknown coefficients to be estimated.[1] In local likelihood, the researcher uses the following log - likelihood function to maximize:

$$\log L(\gamma_o, \Gamma) = -\tfrac{n}{2}\log(2\pi\sigma^2) - \tfrac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - \gamma_o - \Gamma(\mathbf{x} - \mathbf{x}_i)\mathbf{x}_i]^2 K_H(\mathbf{x} - \mathbf{x}_i), \tag{11}$$

where $K_H(u) = |H|^{-1}K(H^{-1}u)$ is a kernel and H is a matrix bandwidth parameter which can be simplified to $H = hI_M$, for some positive constant h and $K(u) = [K_1(u), ..., K_M(u)]'$. The local likelihood must be maximized for each observation of the sample (conditional on the bandwidth parameter). The kernel should satisfy the standard conditions: $\int K(u)du = 1$ and $\int uu'K(u)du = cI_M$ for some c >0. If the parameters in Γ guarantee the imposition of monotonicity and concavity, local likelihood methods have excellent properties as shown in Kumbhakar et al. (2007) but they are computationally expensive if the sample size is (n) is large. In addition, cross-validation must be used to determine h which increases further the computational burden.

To avoid such expensive optimizations the kernel function is made part of the functional form as in (8) when used in (7). Of course in (8) a parametric form is used instead of a kernel. However, the model is effectively semi-parametric as the number of terms G or G' increases and arbitrary functional forms can be approximated arbitrarily well. Whether the approximation is useful, is a question that we take up in section 5 through a Monte Carlo experiment. By "useful" we mean whether the number of terms G or G' is not excessively large in order for the method to be feasible in the case where $\mathbf{x}_i$ is high-dimensional. Additionally, the model does not need cross-validation to determine bandwidth parameters and, in fact, a single optimization is needed to determine all its parameters (for given values of G or G').

Another approach is the Smoothly Mixing Regression (SMR) concept which has been introduced in Geweke and Keane (2007). The method has been explored further, in the context of production analysis, in Tsionas (2017). The SMR approach assumes a normal mixture of the form: $y_i = \mathbf{x}_i'\beta_g + \varepsilon_{i,g}$, $i = 1, ..., n$ where $g \in \{1, ..., \bar{G}\}$ denotes the group,

[1] For simplicity of presentation we abstract from the case where σ^2 can be also localized.

$\varepsilon_{i,g} \sim \mathcal{N}(0, \sigma_g^2(\mathbf{x}_i;\gamma_1))$, $i = 1, ..., n$ and the probability of each group is parametrized as $p_{i,g} = p_{i,g}(x_i;\gamma_2)$ where γ_1 and γ_2 denotes parameters. In fact $\sigma_g^2(\mathbf{x}_i;\gamma_1) = \exp\left(\mathbf{x}_i'\gamma_{1,g}\right)$ and $p_{i,g}(\mathbf{x}_i;\gamma_2) = \frac{\exp\left(\mathbf{x}_i'\gamma_{2,g}\right)}{\sum_{g'=1}^{G}\exp\left(\mathbf{x}_i'\gamma_{2,g'}\right)}$. The model proposed here is clearly simpler as (i) the number of parameters is, in principle, lower, and (ii) we do not adopt a mixture formulation. In fact, imposing monotonicity and concavity on the regression function $E(y_i|\mathbf{x}_i) = \sum_{g=1}^{G} p_{i,g}(\mathbf{x}_i;\gamma_2)\mathbf{x}_i'\beta_g$ is not easy as the $\mathbf{x}_i$s are involved in the mixing probabilities. From this point of view, the model proposed here is much easier to handle in terms of estimation and, of course, in imposing monotonicity and concavity.

5 Estimation

To estimate the model we consider the likelihood function which is given by:

$$L_{G,G'}(\theta;\mathcal{Y}) = (2\pi)^{-n/2}\left\{\prod_{i=1}^{n}\sum_{g=1}^{G'}\exp\left\{\mathbf{x}_i'\zeta_g\right\}p_g(\mathbf{x}_i;\psi_g)\right\}^{-1/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{\left[y_i-\sum_{g=1}^{G}(\mathbf{x}_i'\beta_g)\,p_g(\mathbf{x}_i;\gamma_g)\right]^2}{\sum_{g=1}^{G'}\exp\left\{\mathbf{x}_i'\zeta_g\right\}p_g(\mathbf{x}_i;\psi_g)}\right\}, \tag{12}$$

where $\theta \in \Theta \subseteq \Re^d$ is the entire parameter vector and $\mathcal{Y}$ denotes the data. Although it is possible to maximize the likelihood (for given values of G and G') using standard numerical techniques, there is always the problem to get stuck at a local maximum. A useful alternative is to follow a Bayesian approach which can explore thoroughly the likelihood of the model and avoid such problems. We use a flat prior for all parameters:

$$p(\theta) \propto \text{const.} \tag{13}$$

The Cobb-Douglas parameters are reparametrized to $\alpha_j = \exp\left(-p_j^2\right)$ where p_js are free parameters. To implement the Bayesian approach we make use of Markov Chain Monte Carlo (MCMC) which delivers samples $\left\{\theta^{(s)}, s = 1, ., .., S\right\}$ that converge in distribution to the posterior:

$$p_{G,G'}(\theta|\mathcal{Y}) \propto L_{G,G'}(\theta;\mathcal{Y}) \cdot p(\theta). \tag{14}$$

As the model and its posterior depend on specific values for G and G' model selection is implemented using the marginal or integrated likelihood:

$$\mathcal{M}_{G,G'}(Y) = \int_{\Theta} L_{G,G'}(\theta;\mathcal{Y}) \cdot p(\theta)d\theta. \tag{15}$$

Computation of the marginal likelihood is complicated by the fact that it relies on computation of a multivariate integral. Our MCMC techniques and computation of the marginal likelihood are discussed in the Technical Appendix.

6 Monte Carlo

To decide whether the model is useful in approximating functional forms we use a Monte Carlo experiment. There are $m = 5$ inputs which are generated from a uniform distribution in the interval $(1, 10)$. There are n observations (n =100, 500, 1,000 and

5,000) and the error $\varepsilon_i \sim i.i.d.\mathcal{N}(0,\sigma^2)$. To determine reasonable values for σ we first set $\sigma = 0$ and compute the y_i values as detailed below. Suppose the sample variance of these values is $s_y^2 = n^{-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$. In turn, we set $\sigma = hs_y$ for h=0.1, 0.5 and 1.

The underlying functional form is a Symmetric Generalized McFadden (SGM) due to Diewert and Wales (1987)[2]:

$$Y_i = a'\mathbf{X}_i + \tfrac{1}{2}\frac{\mathbf{X}_i' B \mathbf{X}_i}{\mathbf{X}_i' c} + \varepsilon_i, i = 1, ..., n, \tag{16}$$

where $a \in \Re^m$, B is a symmetric negative definite matrix and $c \in \Re^m$ is a vector whose elements are equal to one. We set $\alpha_j = m^{-1},\ j = 1, ..., m$ and we generated the elements of B from $B = -L'L$ where L is a lower triangular matrix whose elements are set to standard normal random numbers. The SGM is automatically linearly homogeneous in $\mathbf{x}_i$, and globally concave. In order to be increasing we must have:

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = a + \frac{B\mathbf{X}}{c'\mathbf{X}} - \tfrac{1}{2}\frac{\mathbf{X}'B\mathbf{X}}{(c'\mathbf{X})^2}c > \mathbf{0}_m. \tag{17}$$

We enforce these restrictions by rejection at each data point $\mathbf{X} = \mathbf{X}_i$. If the SGM in (16) is, in fact, a profit function then the derivatives in (17) provide the input and output demand functions provided $\mathbf{X}$ contains input and output prices. We are interested in the approximation of the functional form (16) and its derivatives (17). To make the results interpretable we use the fact that we are often interested in the log of the function, $\ln f(\mathbf{X})$ and its elasticities, $\frac{\partial \ln f(\mathbf{X})}{\partial \ln X_j}$. Although the imposition of (17) seems costly, in fact we can use the following approach. Suppose $\mathbf{X} = [X_1, X_2]'$ for simplicity. Let $\underline{X}_j = \min_{i=1,...,n} X_{j,i}$ and $\bar{X}_j = \max_{i=1,...,n} X_{j,i}$ $(j = 1, 2)$. Then the restrictions in (17) can be enforced only at the four points: $A_1(\underline{X}_1, \underline{X}_2)$, $A_2(\underline{X}_1, \bar{X}_2)$, $A_3(\bar{X}_1, \underline{X}_2)$ and $A_4(\bar{X}_1, \bar{X}_2)$. With five inputs this is still somewhat expensive so we follow another strategy: We impose the restrictions at $\underline{\mathbf{X}}$, $\bar{\mathbf{X}}$ and the sampling average, that is at only three points. Then we impose the restrictions in (17) using rejection. This strategy resulted in a vast increase of the acceptance rate requiring, on the average, 10 rejections per acceptance.

We implement MCMC using S =15,000 iterations the first 5,000 of which are discarded to mitigate potential start up effects. We use 50 different MCMC chains of the same length (15,000) starting from widely differing initial conditions generated from a normal distribution with zero mean and covariance matrix $V = h^2 I_d$ where d is the dimensionality of the parameter vector θ and $h = 100$. We report our results in Table 1. The table reports root mean absolute errors (RMSE) between the actual and estimated functional forms in percentage terms (multiplied by 100). Specifically, RMSE= $\sqrt{n^{-1}\sum_{i=1}^{n}\left(f_i - \hat{f}_i\right)^2}$, where $f_i = f(\mathbf{x}_i)$ is the actual function of the function or its derivatives and $\hat{f}_i = \hat{f}(\mathbf{x}_i)$ is the estimated function of the function or its derivatives. In the column headed "average (G, G')" we report the average (across Monte Carlo simulations) numbers of groups for the function (G), see (7)-(8) and the variance of the error term, see (9).

The results from the Monte Carlo approach are quite favorable to the new approach. In fact, with a small number of terms G and G' the approximation properties are excellent and improve as the sample size increases. In fact, RMSEs scale approximately as the square of sample size. As noise (σ) increases, of course, RMSEs increase as well being lowest when σ =0.1. Other values of σ were examined and RMSEs behave as expected but the results are omitted in the interest of space.

[2]We use capital letters to denote the arguments in levels as opposed to their logs.

Table 1: Monte Carlo results

	approximation error (%)						average (G, G')
	$\ln f(\mathbf{x})$	$\frac{\partial \ln f(\mathbf{x})}{\partial \ln x_1}$	$\frac{\partial \ln f(\mathbf{x})}{\partial \ln x_2}$	$\frac{\partial \ln f(\mathbf{x})}{\partial \ln x_3}$	$\frac{\partial \ln f(\mathbf{x})}{\partial \ln x_4}$	$\frac{\partial \ln f(\mathbf{x})}{\partial \ln x_5}$	
σ =1							
n =100	2.31	1.44	1.80	2.55	2.80	1.77	(2, 1)
n =500	0.91	0.58	0.71	1.01	1.11	0.70	(2, 1)
n =1,000	0.63	0.39	0.51	0.72	0.75	0.49	(3, 2)
n =5,000	0.29	0.16	0.24	0.32	0.35	0.22	(3, 2)
σ =0.5							
n =100	1.45	0.97	0.82	0.90	1.04	1.03	(2, 1)
n =500	0.56	0.38	0.32	0.37	0.41	0.40	(2, 1)
n =1,000	0.39	0.26	0.27	0.25	0.28	0.31	(2, 1)
n =5,000	0.18	0.12	0.09	0.11	0.14	0.12	(2, 1)
σ =0.1							
n =100	0.43	0.71	0.54	0.43	0.48	0.52	(3, 2)
n =500	0.17	0.29	0.21	0.16	0.17	0.20	(3, 1)
n =1,000	0.13	0.19	0.16	0.12	0.13	0.13	(3, 1)
n =5,000	0.05	0.09	0.07	0.05	0.07	0.08	(3, 1)

7 Empirical application to U.S banking

As an empirical application we use the same data as in Malikov, Kumbhakar and Tsionas (2016). The data are from large American commercial banks through the period 2001–2010. There is an unbalanced panel with 2,397 bank-year observations for 285 large commercial banks whose total assets exceed one billion dollars (in 2005 dollars). The data come from Call Reports of the Federal Reserve Bank of Chicago. For detailed description of the data construction, see Section 5 of Malikov, Kumbhakar and Tsionas (2016) and the Appendix of Malikov, Restrepo-Tobon and Kumbhakar (2015).

We have five outputs (Y_1= Consumer Loans, Y_2= Real Estate Loans,Y_3= Commercial & Industrial Loans,Y_4= Securities, Y_5= Off-Balance Sheet Activities Income). These are all in thousands of 2005 real U.S dollars. five inputs (X_1= Labor, number of full-time Employees, X_2= Physical Capital (Fixed Assets) , X_3= Purchased Funds, X_4= Interest-Bearing Transaction Accounts , X_5= Non-transaction Accounts) and their prices. Variables X_2, X_3, X_4, and X_5 are expressed in thousands of 2005 U.S dollars.Capital letters denote levels and lower case denotes logs.

Since we have multiple outputs we cannot use a single production function. Instead we use an input distance function representation of the production technology[3]. The input distance function is defined as:

$$\mathcal{D}_I(X, Y) = \max\left\{\lambda : \ \lambda^{-1}X \in \mathcal{L}(Y)\right\}, \tag{18}$$

where $\mathcal{L}(Y) = \{(X, Y) : \ X \text{ is feasible given } Y\}$ in the input set. The input distance function must be increasing and linearly homogeneous in X, decreasing in Y, and concave in X (Kumbhakar and Lovell, 2000, p. 30). Feasible input - output

[3]Both input-oriented and output-oriented distance functions are valid representations of the technology. To choose among them we have to think about the behavior of banks. If they are cost-minimizing then inputs are endogenous while outputs as taken as predetermined. An output-oriented distance function would be consistent with revenue-maximizing behavior where inputs are taken as given and outputs are choice variable and, therefore, endogenous. In this application we stick with the cost-minimizing approach as it has been employed widely in banking studies.

vectors are those for which $\mathcal{D}_I(X,Y) \geq 1$. Technical efficiency is defined as: $\mathcal{E}(X,Y) = \min\{\vartheta : \mathcal{D}_I(\vartheta X, Y) \geq 1\}$. Therefore, $\mathcal{E}(X,Y) = \mathcal{D}_I(X,Y)^{-1}$. Econometric implementation of the input distance function relies on the following representation:

$$\log \mathcal{D}_I(x_{1,it}, ..., x_{5,it}, y_{1,it}, ..., y_{5,it}, T_{it}) = v_{it} + u_{it},\ i = 1, ..., n,\ t = 1, ..., T, \tag{19}$$

where $u_{it} \geq 0$ is a non-negative error term representing technical inefficiency in production, $T_{it} = t$ is a time trend and x_{it}, y_{it} are in log terms to simplify notation. To maintain linear homogeneity with respect to $\mathbf{x}_{it}$ the Cobb-Douglas exponents corresponding to the inputs are restricted to be positive and have unit sum. The Cobb-Douglas exponents corresponding to the outputs are are constrained to be negative to preserve monotonicity of the input distance function . For estimation of technical inefficiency we follow the Cornwell, Schmidt and Sickles (1990, CSS) approach. The CSS approach parametrizes technical inefficiency as follows:

$$u_{it} = a_i + b_i t + c_i t^2,\ \ i = 1, ..., n,\ t = 1, ..., T, \tag{20}$$

that is, we have a quadratic function of trend with bank - specific coefficients. Once the parameters have been estimated, technical inefficiency is estimated as:

$$\tilde{u}_{it} = \hat{u}_{it} - \min_{i,t} \hat{u}_{it}, \tag{21}$$

where

$$\hat{u}_{it} = \hat{a}_i + \hat{b}_i t + \hat{c}_i t^2,\ \ i = 1, ..., n,\ t = 1, ..., T. \tag{22}$$

For the translog as well as the Smooth SCLS, the input distance function is expressed directly in log terms. Therefore, we can write the input distance function in (19) as follows:

$$\log \mathcal{D}_I(\mathbf{x}_{it}, \mathbf{y}_{it}, T_{it}) = v_{it} + u_{it} \Longleftrightarrow x_{1,it} = F(\mathbf{x}^*_{(-1),it}, \mathbf{y}_{it}, T_{it}) + v_{it} + u_{it}, \tag{23}$$

where $\mathbf{x}^*{}_{(-1),it} = [x_{2,it} - x_{1,it}, ..., x_{5,it} - x_{1,it}]'$, $\mathbf{y}_{it} = [y_{1,it}, ..., y_{5,it}]'$ and $F(\cdot)$ is a functional form. Solving with respect to $x_{1,it}$ is possible because of linear homogeneity of the input distance function with respect to inputs. In turn, we define input elasticities $\frac{\partial \log \mathcal{D}_I(\mathbf{x}_{it}, \mathbf{y}_{it}, T_{it})}{\partial \log x_{j,it}} \geq 0, j = 1, ..., 5$, output elasticities $\frac{\partial \log \mathcal{D}_I(\mathbf{x}_{it}, \mathbf{y}_{it}, T_{it})}{\partial \log y_{m,it}} \leq 0, m = 1, ..., 5$, returns to scale $RTS_{it} = -\sum_{m=1}^{5} \frac{\partial \log \mathcal{D}_I(\mathbf{x}_{it}, \mathbf{y}_{it}, T_{it})}{\partial \log y_{m,it}}$, and technical change $TC_{it} = -\frac{\partial \log \mathcal{D}_I(\mathbf{x}_{it}, \mathbf{y}_{it}, T_{it})}{\partial \log T_{it}}$. Finally, we define productivity growth as $PG_{it} = TC_{it} + EC_{it}$, where efficiency change is computed as $EC_{it} = \tilde{u}_{it} - \tilde{u}_{i,t-1}$.

With the SGM functional form we have the problem that it is defined in levels rather than in logs. For econometric implementation let $\mathbf{Z}_{it} = [\mathbf{X}'_{it}, \mathbf{Y}'_{it}]'$ so that (16) can be written as follows:

$$\mathcal{D}_I(\mathbf{Z}_{it}) = a'\mathbf{Z}_{it} + \frac{1}{2} \frac{\mathbf{Z}'_{it} B \mathbf{Z}_{it}}{\mathbf{Z}'_{it} c} + a_\tau T_{it} + \frac{1}{2} a_{\tau\tau} T_{it}^2 + b' \mathbf{Z}_{it} T_{it} + v_{it} + u_{it},\ i = 1, ..., n,\ t = 1, ..., T. \tag{24}$$

We set $\mathcal{D}_I(\mathbf{Z}_{it}) = 1$ as our dependent variable and we proceed with formulating the likelihood function as in (12).

We summarize our results in graphical form. In Figure 1 we report sample densities of input elasticities (upper panel) and output elasticities (bottom panel). Input elasticities (and also returns to scale, technical scale, inefficiency etc) are functions of both the data and the parameters. First, elasticities are computed for all observations in the sample and all parameter draws. In turn, they are averaged across parameter draws to account for parameter uncertainty, resulting in posterior mean estimates. Finally, the kernel density across all observations is presented. As expected, input elasticities are positive and output elasticities are negative. In Figure 2 we report returns to scale from the Smooth SCLS, as well as corresponding results from a translog and an SGM input distance function. Returns to scale estimates from the translog and the SGM are much lower compared to those delivered by the Smooth SCLS -which is evidence that the approximation properties of the former are inferior.

In Figure 3 we report sample densities for technical change (TC_{it}) defined as the derivative of the input distance function with respect to trend. Estimates from the Smooth SCLS and the translog suggest that technical change averages almost 1% per year and ranges from -3% to 4% in the Smooth SCLS compared to -1% to 3% for the translog. For the SGM technical change averages -1% and ranges from -3% to 2.5%. The sample densities of TC estimates are, of course, quite different.

In Figure 4 we report sample densities for technical inefficiency ($\tilde{u}_{it}$) computed using the CSS approach in (20) - (22). Inefficienmcy estimates are close to 15% for the Smooth SCLS but close to 20-25% on the average from translog and the SGM; a fact that has to do with the approximation properties of the latter.

Finally, in Figure 5 reported are sample densities of productivity growth (PG). Estimates from the Smooth SCLS suggest that PG averages close to zero and is likely to be close to zero for most banks in the sample. On the contrary, estimates from SGM and the translog suggest that there is positive productivity growth of, approximately, 2-2.5% on the average and ranging from 1% to 3.7% although the SGM delivers productivity estimates that have larger spread compared to the translog and the Smooth SCLS. These results imply that for important functions of interest, like productivity growth, Smooth SCLS delivers completely different estimates compared to the translog and the SGM. As in Smooth SCLS, productivity growth is close to zero but technical change is not, it must be the case that efficiency change has been, in fact, an important determinant. According to the translog and the SGM, productivity growth is estimated to be at an unrealistic 2.5%. This is hard to believe as there are no channels to feed the productivity growth in banking given that ATM and electronic technologies have exhausted long ago their positive effects. See for example Brissimis et al. (2009), Casu et al. (2016), Kenjegalieva et al. (2009), Kevork et al. (2017)), Tzeremes (2017), and Tsionas and Mamatzakis (2017) .

It is perhaps, surprising, that productivity growth is so different in Smooth SCLS relative to the parametric alternatives (translog and Symmetric Generalized McFadden). To examine this issue further we use another data set that has been used before in Tsionas and Mamatzakis (2017). It is an unbalanced panel that includes 17,399 observations for 31 advanced countries, 7,130 observations for 35 emerging countries, and 2,471 observations for 40 developing countries for the period 2000 - 2013. Bank-specific financial variables are obtained from Bankscope and country-level variables are collected from the World Bank Indicators database. We have three bank outputs: net loans (y1), other earning assets (y2), and off balance sheet items (y3). There are three inputs: funds (x1) which is total customer deposits; physical capital (x2) defined as fixed assets; and labor (x3). Equity (E) is included as a netput (Berger & Mester, 1997). Tsionas and Mamatzakis (2017) used an alternative profit function

Figure 1: Sample densities of posterior mean input elasticities

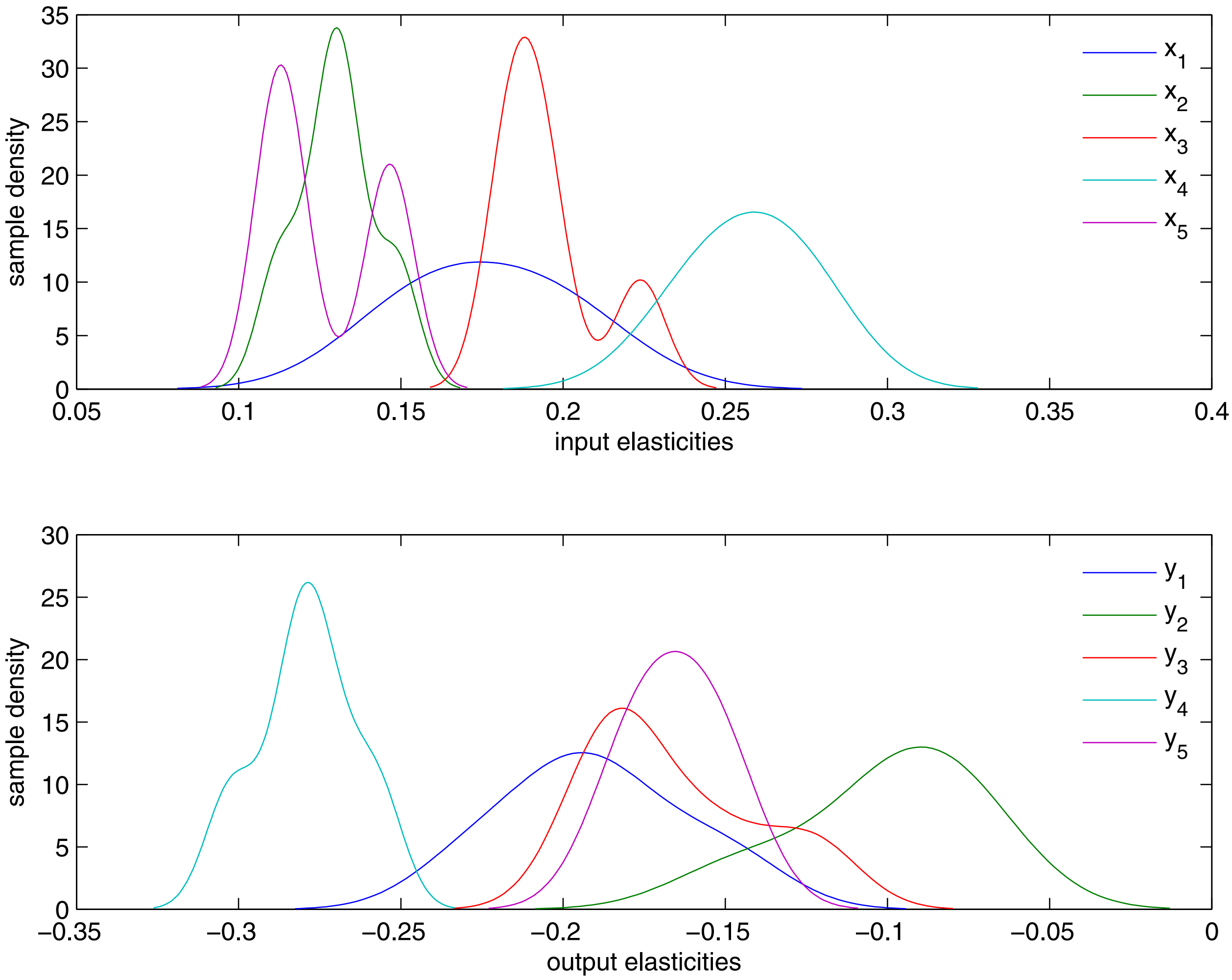


Figure 2: Sample densities of returns to scale

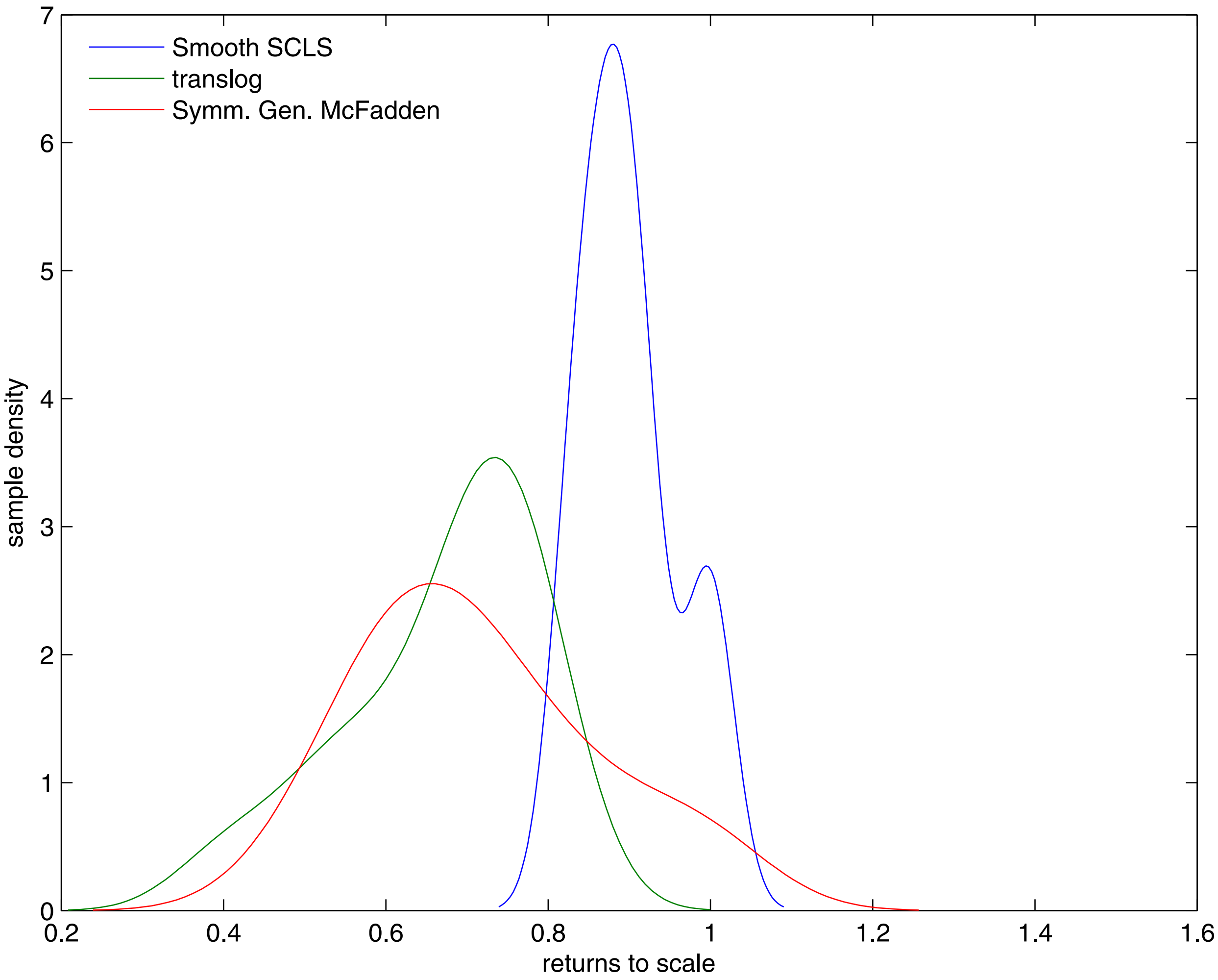


Figure 3: Sample densities of posterior mean technical change

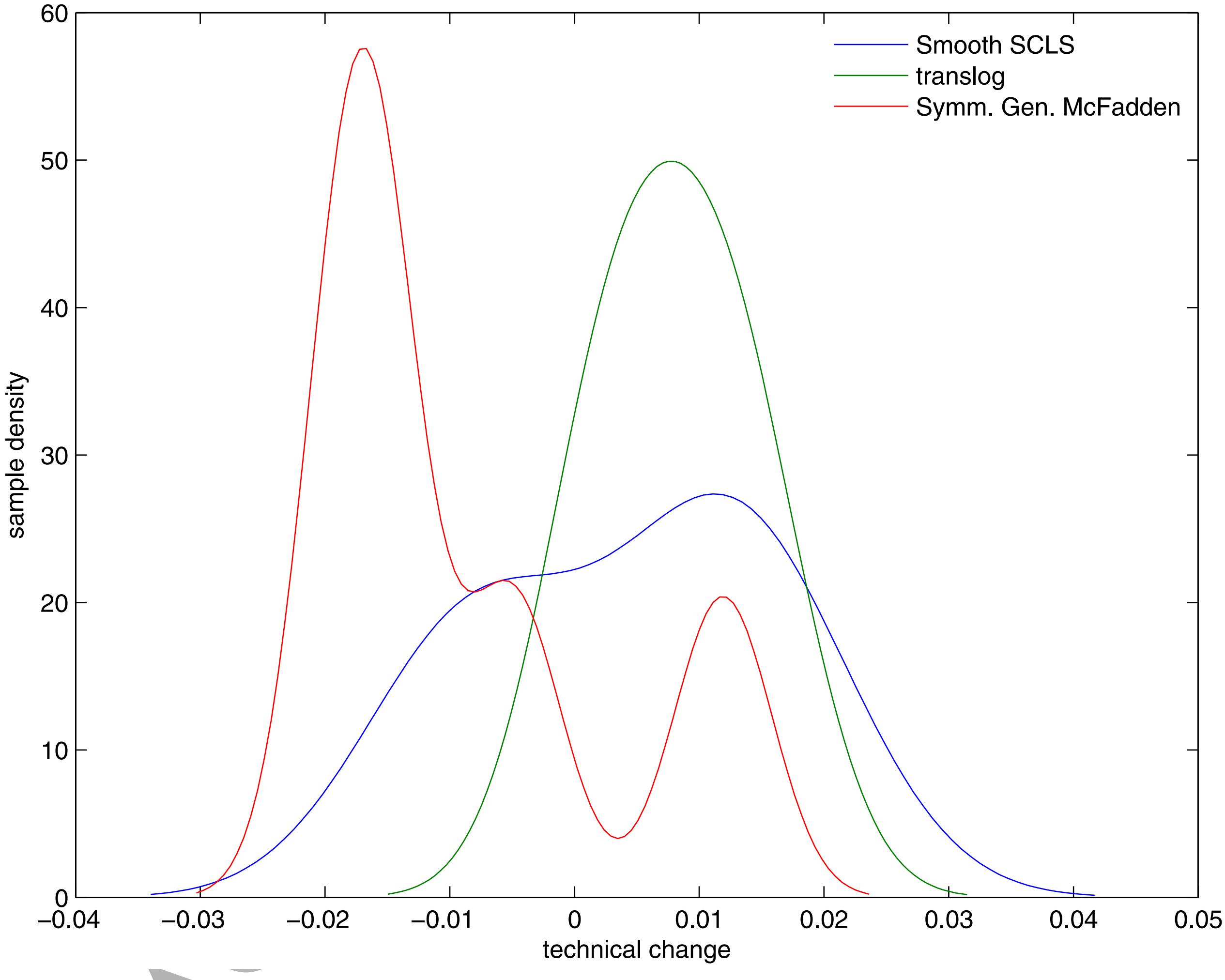


Figure 4: Sample densities of posterior mean technical inefficiency

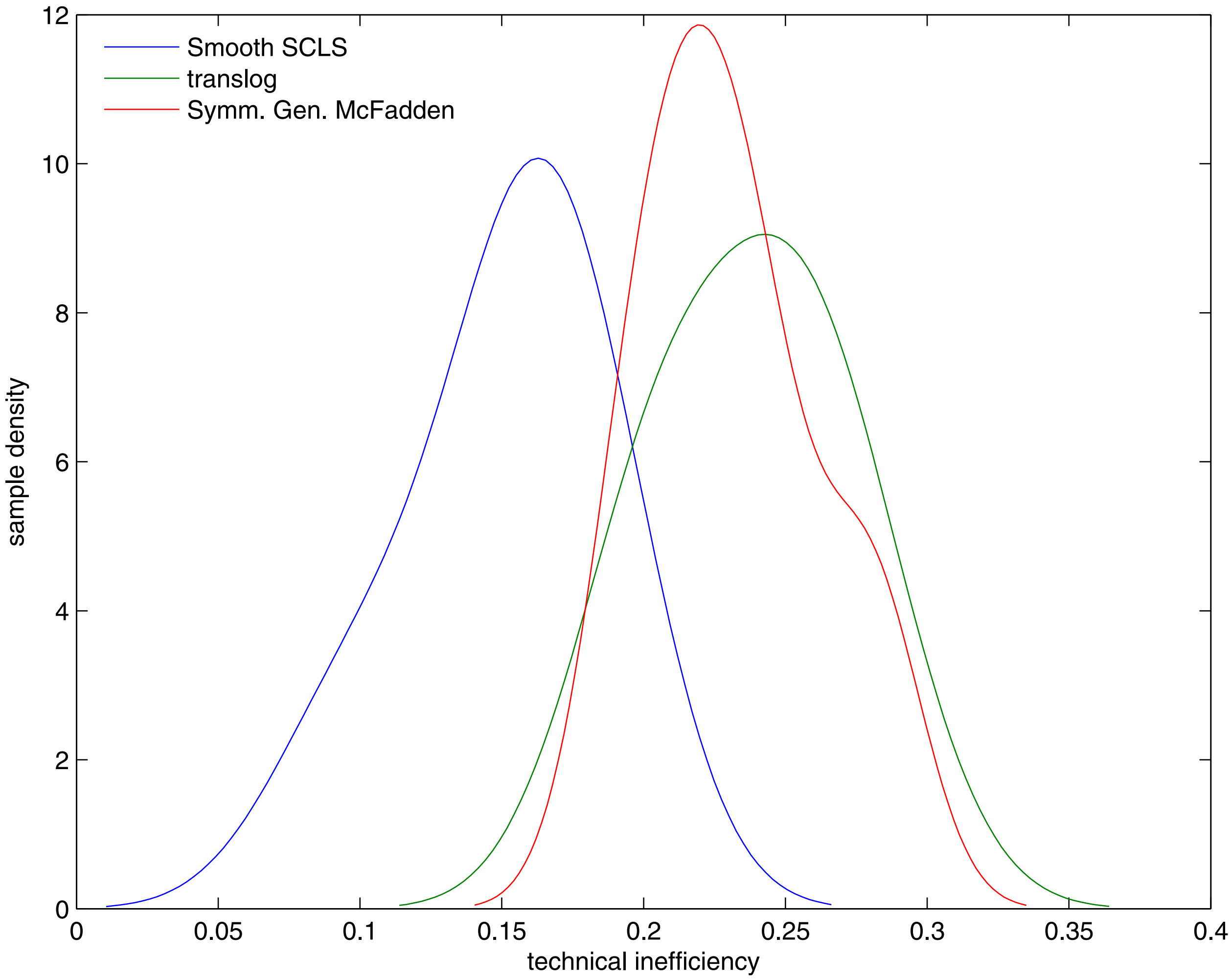


Figure 5: Sample densities of posterior mean productivity growth

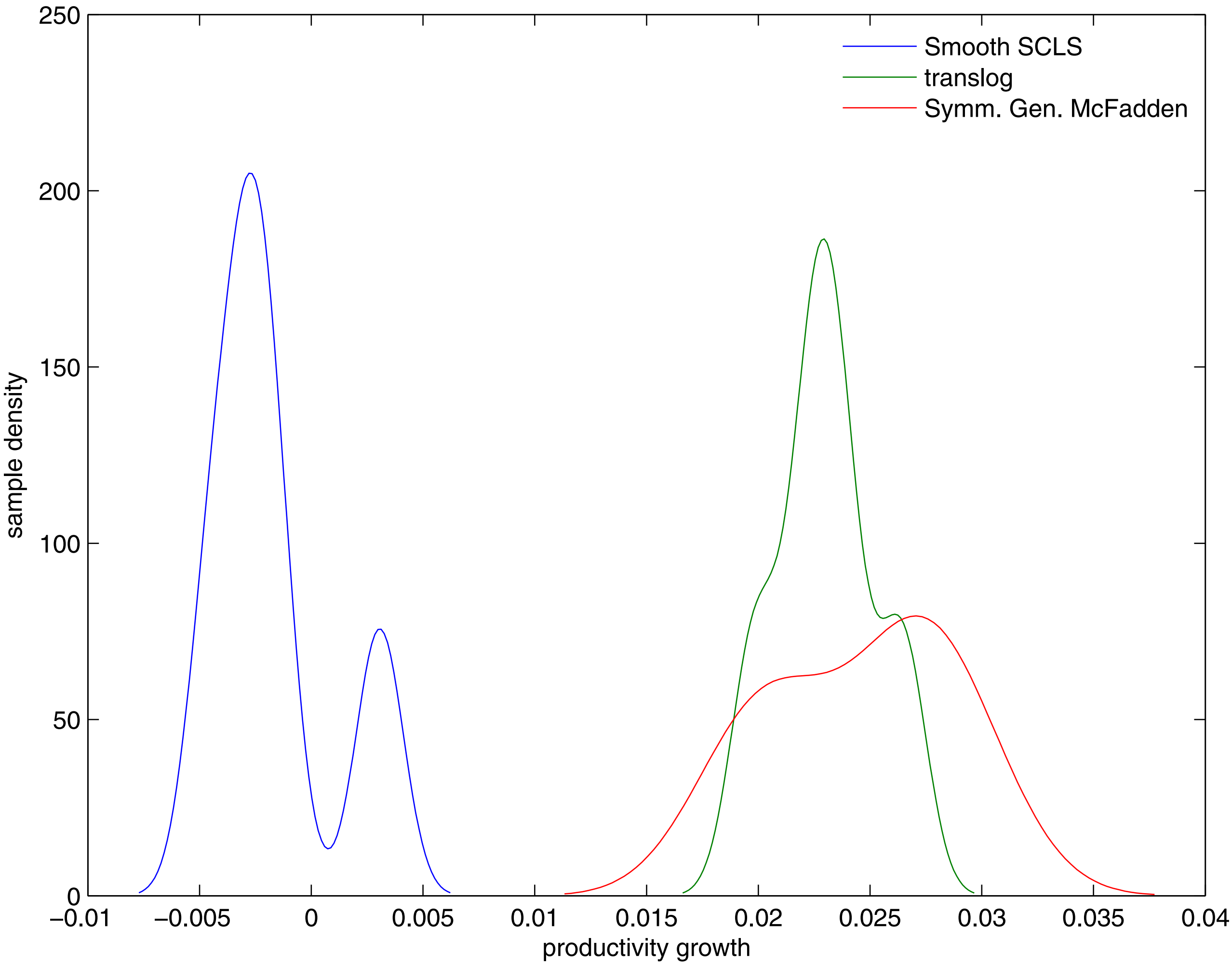
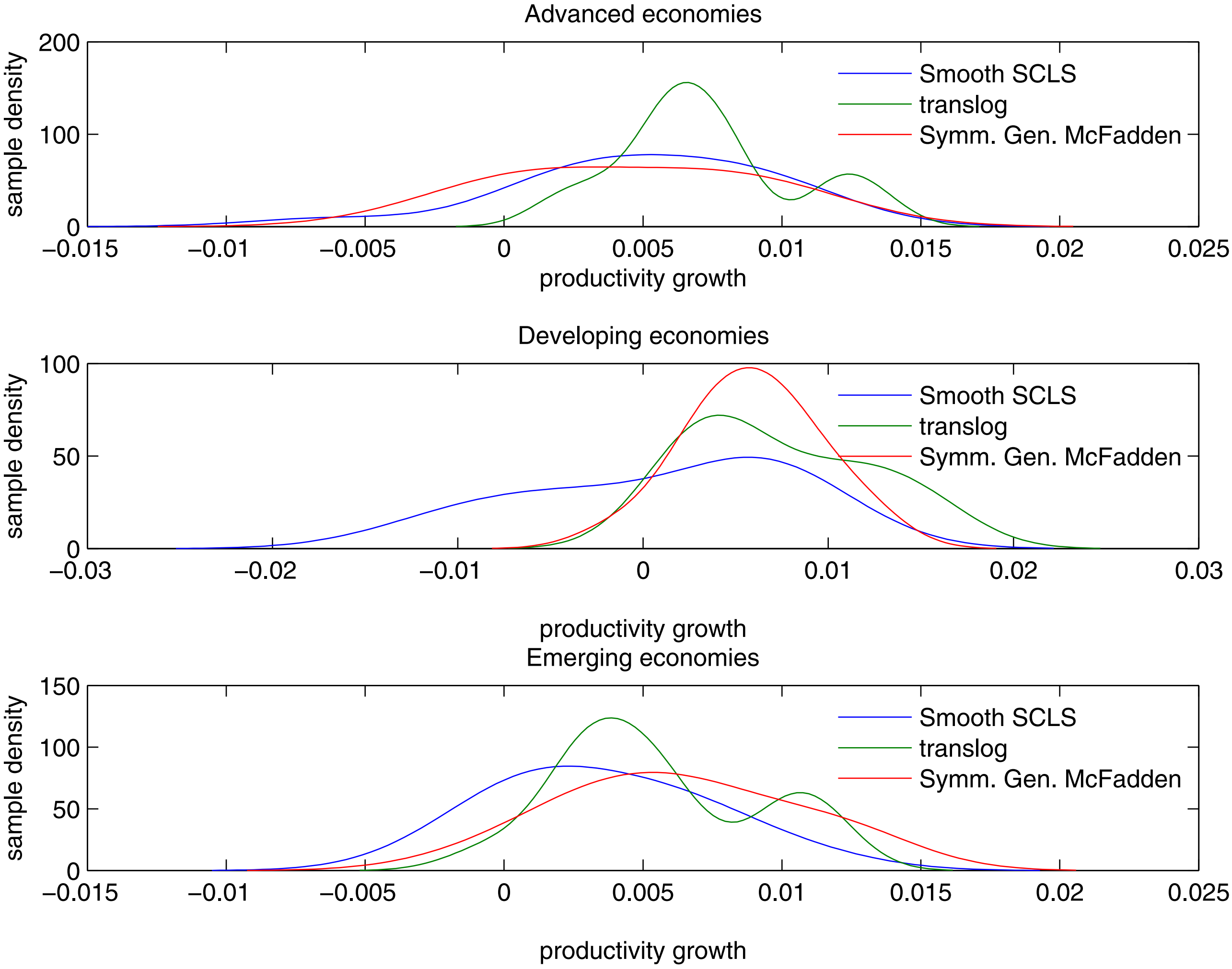


Figure 6: Sample distributions of productivity growth (global sample)



and local maximum likelihood techniques in their empirical application. Here, we use an input distance function so the results are not directly comparable although we can compare Smooth SCLS with the parametric alternatives.

As our interest focuses on productivity growth we omit estimation details for input elasticities, returns to scale and inefficiency (which are available on request). Sample distributions of productivity growth are reported in Figure 6.

The distributions are, again, different but apparently less so compared to the U.S sample. The question that we want to address is whether differences between Smooth SCLS and the parametric models can be explained systematically by country - specific variables. If this is so then we can argue that Smooth SCLS captures aspects of the data that cannot be captured by the parametric models (notice that Smooth SCLS is non - parametric). We use the same country - specific variables as in Tsionas and Mamatzakis (2017, Table 6, p. 648). The dependent variables are productivity growth from SCLS minus productivity growth from translog (DTL) and productivity growth from SCLS minus productivity growth from the Symmetric Generalized McFadden (DSGM). We include country effects, time effects and bank fixed effects in our regressions which are estimated using standard panel data methods (a Hausman test rejected the random effects specification at conventional levels of statistical significance).

Table 2: Panel regressions to explain productivity growth differences from Smooth SCLS relative to translog and Symmetric Generalized McFadden

	DTL	DSGM
Z-score	0.017 (1.044)	0.0032 (0.414)
Capital ratio	-0.015 (-0.930)	-0.027 (-1.20)
Non-interest income	0.043 (4.717)	0.015 (0.717)
Securities	0.022 (7.313)	0.014 (0.912)
per capita GDP	-0.0035 (-6.212)	0.0017 (1.731)
inflation	0.0017 (9.310)	0.0014 (0.810)
population density	0.0021 (1.313)	-0.0013 (0.415)
market size	0.0025 (4.112)	-0.0011 (1.310)
adjusted R^2	0.112	0.053

The empirical results are reported in Table 2 (t-statistics are reported in parentheses).

Differences of productivity growth between the Smooth SCLS and the translog seem to be explained by a number of variables including non-interest income, securities, per capita GDP, inflation, market size and a quadratic function of time trend. Therefore, these differences are systematic. For the Symmetric Generalized McFadden we fail to find really strong evidence to explain the differences, with the exception of per capita GDP which is marginally significant (t-statistic is 1.731). Ojn the other hand country effects, bank effects and time effects are statistically significant (p-values < 0.001 in three all cases).

From these results it turns out that the differences relative to translog can be explained by a number of variables but this is not so for the differences from Symmetric Generalized McFadden. This is reasonable in view of the fact that the former is a *local* second - order approximation to the unknown functional form at the means of the data, whereas the latter is a *global* approximation. A global approximation is expected to perform better when compared to a non-parametric specification like the Smooth SCLS. *A better approximation does not imply, however, that the results will be the same as we find important differences of productivity growth that can be attributed to country effects, bank effects and time effects.* It is difficult to say where these differences arise from in the case of the Symmetric Generalized McFadden although in the case of the translog it is clear that we have misspecification as both country-specific and bank-specific variables are responsible for the differences at least to some extent. Based on the empirical results in Table 2, higher non-interest income, securities, inflation, population density and market size contribute to *higher* productivity growth in the Smooth SCLS relative to the translog. On the other hand higher per capita GDP contributes to *lower* productivity growth implying that it systematically *overstates* productivity growth in economies with lower per capital GDP. It also understates productivity growth in economies with higher non-interest income, securities, inflation, population density and market size. For the Symmetric Generalized McFadden we cannot have such statements as none of the variables seem to explain differences in productivity growth. All differences are attributed to fixed bank, country and time effects which implies that (i) either the differences are due to wrong specification of the functional form in the parametric alternatives, or (ii) there are structural differences beyond those predicted by econometric specifications but are related to the allocation of credit in different economies. One cannot exclude the possibility that both factors are at work. Relative to the second explanation there is some evidence that credit distortions explain productivity differences between the core and periphery in Europe (Hassan et al., 2017, see also Gopinath et al., 2015). Hartmann et al. (2007) also provide an interesting explanation by developing indicators of different aspects of the financial system related to size, innovation, complete markets, transparency, corporate governance, regulation, and competition. Hassan et al. (2017) focus on complete versus incomplete markets to make the point that credit constraints in the European periphery are responsible for incomplete markets and therefore lower productivity.

Concluding remarks

In this paper we provide a semi-parametric alternative to the segmented concave least squares procedure (SCLS) of Keshvari (2017). The method avoids expensive optimizations and relies instead on straightforward estimation by maximum likelihood or Bayesian techniques using Markov Chain Monte Carlo (MCMC). We implement an efficient MCMC scheme that uses second-derivative information from the log posterior. The model is a smooth mixture of anchor Cobb-Douglas production, cost or distance functions. Smoothness is often a desirable property besides monotonicity and curvature but cannot be delivered by SCLS. We show that the new techniques perform well in Monte Carlo studies where the data generating process is the Symmetric Generalized McFadden functional form. Given the good performance of the new method in the Monte Carlo study we take up an empirical application to US banking where the results are compared to those from translog as well as a Symmetric Generalized McFadden input distance function. Given the productivity growth differences between our Smooth SCLS and the parametric alternatives, we examine the three models using a global banking data set. It turns out that the differences are still systematic in the case of the translog as they depend on country and bank specific variables. In the case of the Symmetric Generalized McFadden the differences cannot be attributed to these variables although bank, country and time effects are still significant in explaining these differences.

As the Smooth SCLS model proposed here is non-parametric one can attribute these banking productivity growth differences to the specification of the functional form. Although this is possible it would be interesting to examine the relative performance of the three methods in other banking contexts as it seems that productivity differences can be attributed to size, innovation, complete markets, transparency, corporate governance, regulation, and competition (Hartmann et al., 2007). As we do not have data on such indicators the issue is left necessarily for future research. Another possibility that could be examined in future research is the choice of anchoring functions in SCLS or Smooth SCLS. Here we have used loglinear anchoring functions as it is easy to impose curvature. Another alternative is the use of Symmetric Generalized McFadden anchors which also preserve curvature under parameter restrictions independently of the data. It remains to be seen how technical difficulties associated with the Symmetric Generalized McFadden can be resolved and whether it is worth trying more complicated anchoring functions.

Allon, G., Beenstock, M., Hackman, S., Passy, U., & Shapiro, A. (2007). Nonparametric estimation of concave production technologies by entropic methods. Journal of Applied Econometrics, 22(4), 795–816.

Altunbas, Y. Carbo, S., Gardener, E.P.M., & Molyneux, P. (2007). Examining the relationships between capital, risk and efficiency in European banking. European Financial Management 13, 49-70.

Berger, A.N., & De Young R. (1997). Problem loans and cost efficiency in commercial banking. Journal of Banking and Finance 21, 849-870.

Berger, A., & Mester, L. (1997). Inside the black box: what explains differences in the efficiencies of financial institutions? Journal of Banking and Finance, 21, 895–947.

Boyd, J.H., & Nicolo, G.D. (2003). The theory of bank risk taking and competition revisited. Journal of Finance 60, 1329-1343.

Brissimis, S.N., Delis, M.D. & Tsionas, E.G. (2009). Technical and allocative efficiency in European banking. European Journal of Operational Research, 204 (1), 153-163.

Casu, B., & Girardone, C. (2009). Testing the relationship between competition and efficiency in banking: A panel data analysis. Economics Letters 105, 134-137.

Casu, B., Ferrari, A., Girardone, C. & Wilson, J. O. S. (2016). Integration, productivity and technological spillovers: Evidence for eurozone banking industries. European Journal of Operational Research, 255 (3), 971-983.

Cihak, M., Schaeck, K., & Wolfe, S. (2009). Are competitive banking systems more stable? Journal of Money, Credit, and Banking 41, 711-734

Cornwell, C., Schmidt, P. & Sickles, R. C. (1990), Production frontiers with cross-sectional and time-series variation in efficiency level. Journal of Econometrics 46, 185-200.

DiCiccio, T. J., Kass, R. E., Raftery, A. E. & Wasserman, L. (1997) Computing Bayes factors by combining simulation and asymptotic approximations. Journal of the American Statistical Association 92, 903-15.

Diewert, W.E. & Wales, T. J. (1987). Flexible Functional Forms and Global Curvature Conditions. Econometrica 55 (1), 43-68.

ECB, European Central Bank (2010). Financial Integration in Europe, April.

Gallant, A.R. & G.H. Golub (1984). Imposing curvature restrictions on flexible functional forms. Journal of Econometrics 26, 295-321.

Fiordelisi, F., D. Marques-Ibanez & P. Molyneux (2010). Efficiency and risk in European banking. European Central Bank, Working Paper 1211/June.

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

Girolami, M., & Calderhead, B. (2011) Riemann manifold Langevin and Hamiltonian Monte Carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73(2), 123–214.

Goddard, J., Molyneux, P., Wilson, J.O.S., & Tavakoli, M. (2007). European Banking: An Overview. Journal of Banking and Finance 31, 1911–35.

Gopinath, G., Kalemli-Ozcan S., L. Karabarbounis L., & C. Villegas-Sanchez (2015). Capital Allocation and Productivity in South Europe. NBER Working Paper No. 21453.

Hannah, L. A., & Dunson, D. B. (2013). Multivariate convex regression with adaptive partitioning. The Journal of Machine Learning Research, 14, 3261-3294.

Hartmann, P., Heider, F., Papaioannou, E., & Lo Duca, M. (2007). The Role of Financial Markets and Innovation in Productivity and Growth in Europe. European Central Bank Occasional Paper no. 72

Hassan, F., F. di Mauro & G. J. P. Ottaviano (2017). Banks credit and productivity growth. European Central Bank, working paper no. 2008.

Hellmann, T. F., Murdock, K. C., & Stiglitz, J.E. (2000). Liberalization, moral hazard in banking, and prudential regulation: Are capital requirements enough? American Economic Review 90, 147-65.

Hornik, K. (1989). Approximation Capabilities of Multilayer Feedforward Networks. Neural Networks, 4, 251-257 .

Hornik, K. Stinchcombe, M. & White, H. (1989). Multilayer Feedforward Networks are Universal Approximators, Neural Networks 2, 359-366.

Hughes, J.P. & Mester, L.J. (1998). Bank capitalization and cost: Evidence of scale economies in risk management and signalling. The Review of Economics and Statistics 80, 314-325.

Hughes, J.P. & Mester, L.J. (2009). Measuring the Performance of Banks: Theory, Practice, Evidence, and Some Policy Implications, in Berger, A.N., Molyneux, P. & Wilson, J.O.S (Eds.). Oxford Handbook of Banking. Oxford University Press, 2013, Chapter 10m 247-270.

Kenjegalieva, K., Simper, R. Weyman-Jones, T. & Zelenyuk, V. (2009). European Journal of Operational Research, 198 (1), 326-340.

Keshvari, A. (2017). A penalized method for multivariate concave least squares with application to productivity analysis. European Journal of Operational Research, 257 (3), 1016-1029.

Keshvari, A., & Kuosmanen, T. (2013). Stochastic non-convex envelopment of data: Applying isotonic regression to frontier estimation. European Journal of Operational Research, 231(2), 481–491.

Kevork, I.S., Pange, J., Tzeremes, P. & Tzeremes, N.G. (2017). Estimating Malmquist productivity indexes using probabilistic directional distances: An application to the European banking sector. European Journal of Operational Research, 261 (3), 1125-1140.

Kumbhakar, S.C., Park, B.U., Simar, L., & Tsionas, E.G. (2007). Nonparametric stochastic frontiers: A local maximum likelihood approach. Journal of Econometrics 137 (1), 1-27.

Kuosmanen, T. (2008). Representation theorem for convex nonparametric least squares. Econometrics Journal, 11(2), 308–325.

Kuosmanen, T. & Kortelainen, M. (2012). Stochastic non-smooth envelopment of data: Semiparametric frontier estimation subject to shape constraints. Journal of Productivity Analysis, 38 (1), 11–28.

Kwan, S. & Eisenbeis, R., 1997. Bank risk, capitalization and operating efficiency. Journal of Financial Services Research 12, 117-131.

Lee, C. Y., Johnson, A. L., Moreno-Centeno, E. & Kuosmanen, T. (2013). A more efficient algorithm for Convex Nonparametric Least Squares. European Journal of Operational Research 227 (2), 391-400.

Malikov, E., Kumbhakar, S.C. & Tsionas, E.G. (2016). A Cost System Approach to the Stochastic Directional Technology Distance Function with Undesirable Outputs: The Case of U.S. Banks in 2001-2010, Journal of Applied Econometrics, 31(7), 1407-1429.

Malikov, E., Restrepo-Tobon, D.A. & Kumbhakar, S.C. (2015) Estimation of Banking Technology under Credit Uncertainty. Empirical Economics, 49 (1):185-211.

McAllister, P. H., McManus, D. (1993). Resolving the Scale Efficiency Puzzle in Banking. Journal of Banking & Finance 17 (2-3), pp. 389–405.

Michaelides, P. G., A. T., Vouldis, & Tsionas, E. G. (2015). Globally flexible functional forms: The neural distance function. European Journal of Operational Research 206 (2), 456-469.

Michaelides, P. G., Tsionas, E. G., Vouldis, A. T. & Konstantakis, K. N. (2015). Global approximation to arbitrary cost functions: A Bayesian approach with application to US banking. European Journal of Operational Research 241 (1), 148-160.

Nadaraya, E. A. (1964). On estimating regression. Theoretical Probability and Applications 9 141-142.

O'Donnell, J.C. & J.T. Coelli (2005). A Bayesian approach to imposing curvature on distance functions Journal of Econometrics 126 (2), 493-523.

Terrell, D. (1996). Incorporating monotonicity and concavity conditions in flexible functional forms. Journal of Applied Econometrics 11, 179-194.

Tibshirani, R. J. & Hastie, T. J. (1987). Local likelihood estimation. Journal of the American Statistical Association 82, 559–567.

Tzeremes, N.G. (2017). Efficiency dynamics in Indian banking: A conditional directional distance approach. European Journal of Operational Research 240 (3), 807-818.

Tsionas, M.G. (2017). "When, Where, and How" of Efficiency Estimation: Improved Procedures for Stochastic Frontier Modeling. Journal of the American Statistical Association 112 (519), 948-965.

Tsionas, M.G. & Mamatzakis, E. (2017). Adjustment costs in the technical efficiency: An application to global banking. European Journal of Operational Research, 256 (2), 640-649.

Vaneman, W. K. & K. Triantis (2007). Evaluating the productive efficiency of dynamical systems. IEEE Transactions on Engineering Management 54 (3), 600-612.

Watson, G. S. (1964). Smooth regression analysis. Sankhya A 26, 359-372.

White, H. (1989). Some asymptotic results for learning in single hidden layer feedforward network models. Journal of the American Statistical Association, 84, 1003–1013.

Williams, J. (2004). Determining management behaviour in European banking. Journal of Banking and Finance 28, 2427-2460.

TECHNICAL APPENDIX. Inference and Markov Chain Monte Carlo (MCMC)

To explore the posterior we use Markov Chain Monte Carlo (MCMC) techniques. Specifically we use the Girolami and Calderhead (2011, GC) Riemannian manifold Hamiltonian MCMC. The algorithm uses local information about both the gradient and the Hessian of the log-posterior conditional of parameter vector $\theta \in \Theta \subseteq \Re^d$ at the existing draw. Suppose $\mathscr{L}(\theta) = \log p(\theta|\mathcal{Y})$ is used to denote for simplicity the log posterior of θ. Moreover, define:

$$\boldsymbol{G}(\theta) = \text{est.cov}\tfrac{\partial}{\partial\theta}\log p(\mathcal{Y}|\theta), \tag{A.1}$$

the empirical counterpart of

$$\boldsymbol{G}_o(\theta) = -E_{\mathcal{Y}|\theta}\tfrac{\partial^2}{\partial\theta\partial\theta'}\log p(\mathcal{Y}|\theta). \tag{A.2}$$

The Langevin diffusion is given by the following stochastic differential equation:

$$d\theta(t) = \tfrac{1}{2}\tilde{\nabla}_\theta\mathscr{L}\{\theta(t)\}\,dt + d\mathbf{B}(t), \tag{A.3}$$

where

$$\tilde{\nabla}_\theta\mathscr{L}\{\theta(t)\} = -\boldsymbol{G}^{-1}\{\theta(t)\}\cdot\nabla_\theta\mathscr{L}\{\theta(t)\}, \tag{A.4}$$

is the so called "natural gradient" of the Riemann manifold generated by the log posterior. The elements of the Brownian motion are

$$\begin{aligned}\boldsymbol{G}^{-1}\{\theta(t)\}d\mathbf{B}_i(t) =& |\boldsymbol{G}\{\theta(t)\}|^{-1/2}\sum_{j=1}^{d}\tfrac{\partial}{\partial\theta}\left[\boldsymbol{G}^{-1}\{\theta(t)\}_{ij}|\boldsymbol{G}\{\theta(t)\}|^{1/2}\right]dt \\ &+\left[\sqrt{\boldsymbol{G}\{\theta(t)\}}d\boldsymbol{B}(t)\right]_i\end{aligned} \tag{A.5}$$

The discrete form of the stochastic differential equation provides a proposal as follows:

$$\begin{aligned}\tilde{\theta}_i =& \theta_i^o + \tfrac{\varepsilon^2}{2}\left\{\boldsymbol{G}^{-1}(\theta^o)\nabla_\theta\mathscr{L}(\theta^o)\right\}_i - \varepsilon^2\sum_{j=1}^{d}\left\{\boldsymbol{G}^{-1}(\theta^o)\frac{\partial\boldsymbol{G}(\theta^o)}{\partial\theta_j}\boldsymbol{G}^{-1}(\theta^o)\right\}_{ij} \\ &+\tfrac{\varepsilon^2}{2}\sum_{j=1}^{K_\theta}\left\{\boldsymbol{G}^{-1}(\theta^o)\right\}_{ij}\text{tr}\left\{\boldsymbol{G}^{-1}(\theta^o)\frac{\partial\boldsymbol{G}(\theta^o)}{\partial\theta_j}\right\} + \left\{\varepsilon\sqrt{\boldsymbol{G}^{-1}(\theta^o)}\boldsymbol{\xi}^o\right\}_i \\ =& \boldsymbol{\mu}(\theta^o,\varepsilon)_i + \left\{\varepsilon\sqrt{\boldsymbol{G}^{-1}(\theta^o)}\boldsymbol{\xi}^o\right\}_i,\end{aligned} \tag{A.6}$$

where θ^o is the current draw. The proposal density is:

$$q\left(\tilde{\theta}|\theta^o\right) = \mathcal{N}_d\left(\tilde{\theta}, \varepsilon^2 \boldsymbol{G}^{-1}\left(\theta^o\right)\right), \tag{A.7}$$

and convergence to the invariant distribution is ensured by using the standard form Metropolis-Hastings probability:

$$\min\left\{1, \frac{p\left(\tilde{\theta}|\cdot,\mathcal{Y}\right) q\left(\theta^o|\tilde{\theta}\right)}{p\left(\theta^o|\cdot,\mathcal{Y}\right) q\left(\tilde{\theta}|\theta^o\right)}\right\}. \tag{A.8}$$

To compute the marginal likelihood, $\mathcal{M}(\mathcal{Y}) = \int_\Theta L(\theta;\mathcal{Y})p(\theta)d\theta$, we use the well known identity:

$$\mathcal{M}(\mathcal{Y}) = \frac{L(\theta;\mathcal{Y})p(\theta)}{p(\theta|\mathcal{Y})},\ \forall\theta\in\Theta, \tag{A.9}$$

which holds for all parameter vectors in Θ. Therefore, if $\bar{\theta}\in\Theta$ and is, say, the posterior mean we have:

$$\mathcal{M}(\mathcal{Y}) = \frac{L(\bar{\theta};\mathcal{Y})p(\bar{\theta})}{p(\bar{\theta}|\mathcal{Y})}. \tag{A.10}$$

The numerator can be computed easily. The denominator is not known but can be approximated using a multivariate normal distribution with mean $\bar{\theta}$ and covariance matrix $\bar{V}$ which we set equal to the posterior covariance matrix, viz. $\bar{V} = S^{-1}\sum_{s=1}^{S}\left(\theta^{(s)}-\bar{\theta}\right)\left(\theta^{(s)}-\bar{\theta}\right)'$. Then it is easy to see that:

$$\mathcal{M}(\mathcal{Y}) = (2\pi)^{d/2}|\bar{V}|^{1/2}L(\bar{\theta};\mathcal{Y})p(\bar{\theta}). \tag{A.11}$$

This technique is known as the Laplace approximation, see DiCiccio et al. (1997). Two additional things are of importance. First, how the MCMC scheme works in terms of autocorrelation and, second, how sensitive are MCMC results to initial conditions. High autocorrelation prevents MCMC from thorough exploration of the posterior whereas sensitivity to initial conditions would indicate that the posterior is multimodal and autocorrelation has a destructive effect or exploring the posterior. As we have a large number of parameters[4] we report in Figure A.1 the autocorrelation function (acf) for $||\theta|| = \left(\sum_{j=1}^{d}\theta_j^2\right)^{1/2}$. Individual acf plots are available on request along with other convergence diagnostics which show that MCMC chains converge well in a relatively low number of iterations which is expect as the GC algorithm uses information from the Hessian of the log posterior. To examine sensitivity to different initial conditions we generate random initial conditions from a multivariate normal centered at zero with covariance $h^2 I_d$ where h =100. Posterior means are computed discarding the first 5,000 iterations and the percentage differences for all parameters jointly are presented in Figure A.2.

[4]The number of parameters includes the bank-specific parameters in (20) - (22).

Figure A.1: Sensitivity of posterior means to 1,000 randomly generated initial conditions

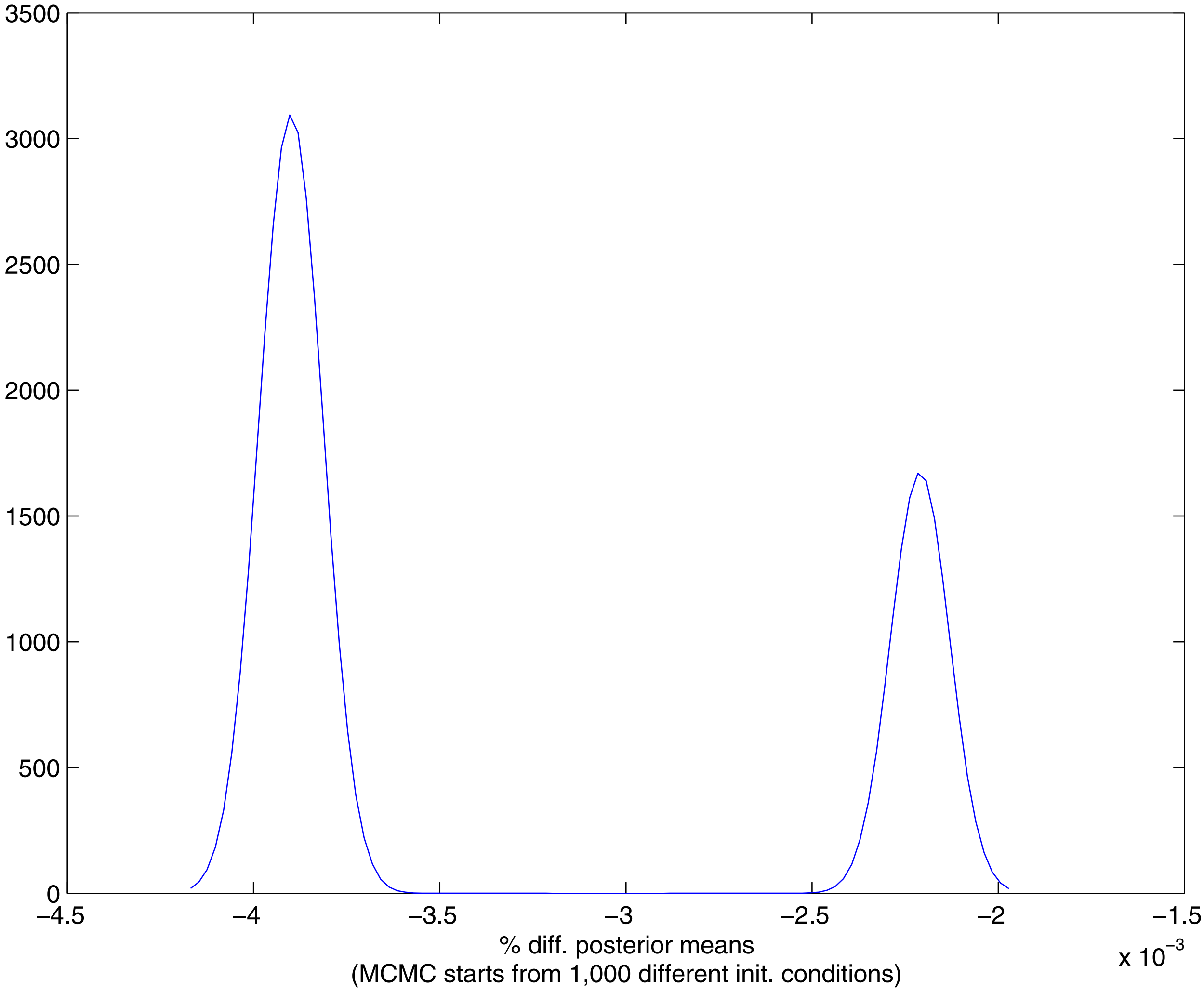


Figure A.2: Autocorrelation function for $||\theta|| = \left(\sum_{j=1}^{d} \theta_j^2\right)^{1/2}$