On a high-dimensional model representation method based on Copulas

Mike G. Tsionas*, Athanasios Andrikopoulos†

Abstract

This article provides an alternative to High-Dimensional Model Representation (HDMR) using a Copula approximation of an unknown functional form. We apply our methodology in the context of an extensive Monte Carlo study and to a sample of large US commercial banks. In the Monte Carlo experiment, the approximations errors of the Copula approach are small and behave randomly. In our empirical application, we find that the Copula Approximation performs much better, in terms of Bayes factors for model comparison, compared to HDMR, which, in turn, provides better results when compared with standard flexible functional forms, like the translog, the minflex Laurent, and the Generalized Leontief (GL), or a Multilayer Perceptron (MLP). Moreover, the choice of approximation has significant implications for productivity and its components (returns to scale, technical inefficiency, technical change, and efficiency change).

Keywords: Productivity and Competitiveness; Copula; High dimensional model Representation; Multilayer Perceptron; Bayesian Analysis.

* Mike G. Tsionas. Lancaster University Management School, LA1 4YX, UK. Tel. +44 (0) 1524 592668, m.tsionas@lancaster.ac.uk.
† Corresponding Author: Athanasios Andrikopoulos. Hull University Business School, University of Hull, Hull HU6 7RX, UK. Tel. +44 (0)1482 46-3310, A.Andrikopoulos@hull.ac.uk.
Acknowledgments: The authors are grateful to two anonymous reviewers for their several ingenious remarks and comments on an earlier version.

1. Introduction

Although many functional forms have been proposed in relation to estimating cost and production functions, there is no consensus as to what one should do in practice (Zhang and Berardi, 2001; Michaelides et al., 2010; 2015). To deal with this problem, many previous studies have used local approximations such as the translog (Christensen et al., 1971; Kumbhakar, 2011; and Kumbhakar, 2013), the Generalized Leontief (Diewert, 1971; Barnett and Lee, 1987; and Genius et al., 2012) or the symmetric McFadden form (Diewert and Wales, 1987; Michaelides et al., 2010; 2015; and Tsionas and Izzeldin, 2018) and global approximations such as the neural cost function (NCF) (Michaelides et al., 2010; 2015).

Estimation of cost and production functions derived using stochastic frontier or envelopment methods is another common practice in the literature (see, for example Sun, Kumbhakar, and Tveterås, 2015; Olesen and Petersen, 2016; among others). Stochastic frontier methods rely on the specification of the functional form of a cost function as well as on specification on distributional assumptions about the two-sided and one-sided error terms (noise and cost inefficiency, respectively). Data envelopment analysis (DEA) models do not involve statistical errors and use a non-parametric approximation to the frontier. As statistical errors are ignored, DEA may be more sensitive to outliers and yields biased estimates of inefficiency.

Tsionas (2003) combines stochastic frontier models with linear programming methods by using DEA measures as priors of efficiency in the stochastic frontier model to measure efficiency in US airlines. Tsionas and Mallick (2019) use latent dynamic stochastic productivity and perform Bayesian analysis using a Sequential Monte Carlo Particle-Filtering approach. They apply their proposed techniques in Indian non-financial firms.

Moreover, following the High Dimensional Model Representation (HMDR) theory (also known as functional Analysis of variance ANOVA), Beccacece et al., (2015) evaluate the strength of interactions in a value function, without making any *a priori* assumptions on its functional form. Their result is that in general their experiment confirmed that the value functions have the desired properties of additivity and monotonicity.

Feil et al. (2009) develop a method based on the HDMR theory for approximation of local volatility functions. They show that the HDMR model can produce more precise results than an alternative model using cubic splines. Tsionas and Izzeldin (2018) suggest an alternative to nonparametric segmented concave least squares. They use a differentiable approximation to an arbitrary functional form based on smoothly mixing Cobb-Douglas anchor functions over the data space. They use Bayesian techniques and Markov Chain Monte Carlo. The approximation properties of the new functional form are examined in a Monte Carlo study where the real functional form is a Symmetric Generalized McFadden.

In this paper, we tackle the approximation problem in a different way, using copulas as approximations of cost functions. Copulas are prominent in modelling multivariate distributions, bivariate for the most part (see, for example, Genest and Mackay1986; Poon et al., 2004, and Hong et al., 2007, among others). However, their use as approximations of functions is less common and appears to be novel at least in the case of modelling cost functions if not in the case of approximating functional forms in general. For the Copula Approximation we provide a novel multivariate mixture-of-normals formulation; a mixture-of-normals has never been used in copulas contrary to other copulas (see Joe, 1997; Cherubini et al., 2004; Nelsen; 1999; Karlis, and Meligkotsidou, 2006; Patton, 2012; Fan and Patton, 2014; Kakouris and Rustem, 2014; Dellaportas and Tsionas, 2019).[1]

From the point of view of global approximation, Gallant (1981) shows the correspondence (in Sobolev metric) between an efficient quadratic (projection) estimation method, SURE (Seemingly Unrelated Regressions Estimation) in its specific context, and the mathematical algorithm for calculating the multivariate Fourier approximation coefficients. This finding is not much diffused in the literature for computational economics and possibly the major difficulty of the mathematical method is yet less diffused in the banking field (for an extensive use see Maggi and Guida, 2011). Nonetheless, Gallant's result clearly disenables the translog and similar approximations, and casts doubts on the Diewert and Wales (1987) results in order to find the true function (their paper is on the empirical application of the theoretical properties of the cost function without the warranty of finding the true function).[2]

In an extensive Monte Carlo experiment, we find that the Copula Approximation performs better compared to a High-Dimensional Model Representation (HDMR) and Multi-Layer Perceptrons (MLP). It also performs best in our empirical application to large US banks using as metric of comparison the Bayes factor.

As already discussed, there is a massive empirical literature on estimating cost and production functions for many industries with quite interesting findings. However, the literature with applications in the banking sector is even more extensive (see for example, Lovell, 1995; Lozano Vivas, 1997; Sathye, 2003; Bos, et al., 2009; Galán et al., 2015; Tzeremes, 2015; and Dong, et al, 2016).[3]

We compare the performance of the new Copula Approximation approach relative to alternative methods in the literature such as the HDMR, the translog, the minflex Laurent, the Generalized Leontief

---

[1] Common parametric copulas are the Gaussian or Normal copula, the Student's t-copula, the Frank copula, the Gumbel copula, and the Clayton copula (for properties of other parametric copulas see Joe, 1997; and Nelsen; 2006).

[2] We wish to thank an anonymous reviewer for making these points and in helping us to motivate the paper in the proper way.

[3] For other studies in banking, efficiency see Koutsomanoli-Filippaki and Mamatzakis, 2009; Ray and Das, 2010; Tecles, and Tabak, 2010; Wanke, et.al., 2015; Badunenko, and Kumbhakar, 2017 and Torri et.al., 2018).

(GL), and a Multilayer Perceptron (see section 4). The Copula Approximation performs much better in Bayes factor terms when employed in a substantive application to large US commercial banks.

We find that the HDMR and commonly used functional forms deliver different results compared to our method. Moreover, based on predictive Bayes factors that are computed over left-out sub-samples of the data, we find that our data strongly support the Copula Approximation providing more corroboration for our Monte Carlo experiments.

2. The HDMR approximation

Suppose $\boldsymbol{x} \in \mathcal{X} \subseteq \Re^n$ represents a vector of the logs of input prices and outputs. An arbitrary cost function can be approximated using an HDMR (see Alis and Rabitz, 1999, Li et al., 2002):

$$f(\boldsymbol{x}) \simeq f_0 + \sum_{i=1}^{n} f_i\,(x_i) + \sum_{1\le i<j\le n} f_{ij}\left(x_i, x_j\right) + \cdots$$
$$+ \sum_{1\le i_1<\cdots<i_{k'}\le n} f_{i_1,i_2\ldots i_{k'}}\left(x_{i_1}, x_{i_2}, \ldots, x_{i_{k'}}\right) + \cdots \tag{1}$$
$$\ldots + f_{1,2,\ldots n}(x_1, x_2, \ldots, x_n),$$

where the first order component function $f_i(x_i)$ gives the independent contribution to $f(\boldsymbol{x})$ by the *ith* input variable acting, and $f_{ij}(x_i, x_j)$ gives the pair correlated contribution to $f(\boldsymbol{x})$ by the input variables x_i and x_j. Here, we use orthonormal polynomials $\{\varphi_k(x), k = 1,2, \ldots\}$ in the domain [0, 1], in order to reduce the sampling effort, such that:[4]

$$\int_0^1 \varphi_k\,(x)dx = 0,$$
$$\int_0^1 \varphi_k{}^2(x)\,dx = 1, \tag{2}$$
$$\int_0^1 \varphi_k\,(x)\varphi_{k'}(x)dx = 0, k \neq k' = 1,2, \ldots,$$

i.e., they have a zero mean, unit norm and are mutually orthogonal. The polynomials can be constructed directly from their definition and the first three are given by:

[4] Other suitable functions are spline functions, or simple polynomial functions (see for example, Li, et al., 2002).

$$
\begin{aligned}
\varphi_1(x) &= \sqrt{3}(2x-1),\\
\varphi_2(x) &= 6\sqrt{5}\left(x^2 - x + \frac{1}{6}\right),\\
\varphi_3(x) &= 20\sqrt{7}(x^3 - \frac{3}{2}x^2 + \frac{3}{5}x - \frac{1}{20}),\\
&\dots
\end{aligned} \tag{3}
$$

Given the set of orthonormal polynomials, we have the following approximation:

$$
\begin{aligned}
f_i(x_i) &= \sum_{k=1}^{\infty} \alpha_k^i \, \varphi_k(x_i),\\
f_{ij}(x_i, x_j) &= \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \beta_{kl}^{ij} \, \varphi_k(x_i)\varphi_l(x_j), \dots
\end{aligned} \tag{4}
$$

In most cases, to achieve a desired accuracy using $\varphi_1(x), \varphi_2(x)$ and $\varphi_3(x)$ (that is, $k, l \leq s = 3$) is sufficient. In this case, (1) becomes:

$$
f(\boldsymbol{x}) \simeq f_0 + \sum_{i=1}^{n} \sum_{k=1}^{s_i} \alpha_k^i \, \varphi_k(x_i) + \sum_{1 \leq i < j \leq n} \sum_{k=1}^{s_i'} \sum_{l=1}^{s_j} \beta_{kl}^{ij} \, \varphi_k(x_i)\varphi_l(x_j), \tag{5}
$$

where $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_k^i\}$ and $\boldsymbol{\beta} = \{\boldsymbol{\beta}_{kl}^{ij}\}$ are sets of coefficients. Notice that as the sum $\sum_{1 \leq i < j \leq n}(.)$ applies only to i<j, implicitly there are symmetry restrictions imposed among the $\boldsymbol{\beta}_{kl}^{ij}$.

$$
\min_{\{\alpha_k^i\}}: \int_0^1 \left[f_i(x_i) - \sum_{k=1}^{s_i} \alpha_k^i \, \varphi_k(x_i) + \right]^2 dx_i + \lambda_i \int_0^1 [\partial^2 \left(\sum_{k=1}^{s_i} \alpha_k^i \, \varphi_k(x_i) \right) / \partial x_i^2]^2 \, dx_i \tag{6}
$$

$$
\begin{aligned}
\min_{\{\beta_{kl}^{ij}\}}: \int_0^1 \int_0^1 \left[f_{ij}(x_{ij}) - \sum_{k=1}^{s} \sum_{l=1}^{s} \beta_{kl}^{ij} \, \varphi_k(x_i)\varphi_l(x_j) \right]^2 dx_i dx_j \, + \\
+ \lambda_{ij} \sum_{s,t \in \{i,j\}} \int_0^1 \int_0^1 \left[\partial^2 \left(\sum_{k=1}^{s} \sum_{l=1}^{s} \beta_{kl}^{ij} \, \varphi_k(x_i)\varphi_l(x_j) \right) \Big/ \partial x_s \partial x_t \right]^2 dx_i dx_j,
\end{aligned} \tag{7}
$$

where λ_i and λ_{ij} are smoothing coefficients. The optimization problems can be solved explicitly. For (6) and the case $s = 3$ we have a linear system of the form:

$$
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 + 720\lambda_i & 0 \\ 0 & 0 & 1 + 8400\lambda_i \end{bmatrix} \cdot \begin{bmatrix} \alpha_1^i \\ \alpha_2^i \\ \alpha_3^i \end{bmatrix} = \boldsymbol{b}, \tag{8}
$$

where $\boldsymbol{b} = \begin{bmatrix} \int_0^1 f_i\,(x_i)\varphi_1(x_i)dx_i \\ \int_0^1 f_i\,(x_i)\varphi_2(x_i)dx_i \\ \int_0^1 f_i\,(x_i)\varphi_3(x_i)dx_i \end{bmatrix} = \begin{bmatrix} \int_X f\,(\boldsymbol{x})\varphi_1(x_i)d\boldsymbol{x} \\ \int_X f\,(\boldsymbol{x})\varphi_2(x_i)d\boldsymbol{x} \\ \int_X f\,(\boldsymbol{x})\varphi_3(x_i)d\boldsymbol{x} \end{bmatrix}$.

Similarly, for (7) we have a linear system of the form $\boldsymbol{A}\beta = \boldsymbol{b}$, where $\boldsymbol{A}$ and $\boldsymbol{b}$ are 6×6 and 6×1 respectively and are given in equations (31)-(34) of Li, et al., (2002). The elements of $\boldsymbol{A}$ are simple to evaluate and the elements of b require the evaluation of $\int_0^1 \int_0^1 f_{ij}(x_i, x_j)\varphi_p(x_i)\varphi_q(x_j)dx_i dx_j, p, q \in \{1, \dots, s\}$ which is equal to $\int_{\mathcal{X}} f(\boldsymbol{x})\varphi_p(x_i)\varphi_q(x_j)d\boldsymbol{x}\; p, q \in \{1, \dots, s\}$.

For our purposes, we use the following extended approximation, which consists of a standard quadratic form plus an HDMR representation:

$$f(\boldsymbol{x}) = a_0 + \boldsymbol{\gamma}^\top \boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^\top \Gamma \boldsymbol{x} + \sum_{i=1}^{n} f_i(x_i) + \sum_{1 \le i < j \le n} f_{ij}(x_i, x_j), \tag{9}$$

where $\boldsymbol{\gamma} \in \Re^n$ and $\Gamma \in \Re^{n \times n}$. We include the quadratic form in the hope that the number of orthogonal polynomials will be reduced (a similar idea is put forward in Gallant, 1981 but in a different context). After approximating using orthonormal polynomials we have:

$$f(\boldsymbol{x}) = a_0 + \boldsymbol{\gamma}^\top \boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^\top \Gamma \boldsymbol{x} + \sum_{i=1}^{n}\sum_{k=1}^{s} \alpha_k^i \varphi_k(x_i) + \sum_{1 \le i < j \le n}\sum_{k=1}^{s}\sum_{l=1}^{s} \beta_{kl}^{ij} \varphi_k(x_i)\varphi_l(x_j). \tag{10}$$

3. A new approximation

Suppose we transform all variables $\boldsymbol{x} \in \mathcal{X} \subseteq \Re^n$ so that $\mathcal{X} = [0,1]^n$. Suppose also that for each variable it is reasonable to assume, as a first approximation, that $f_i(x_i) = (\alpha_i + 1)x_i^{\alpha_i}, x_i \in [0,1], i = 1, \dots, n$. Notice that $\int_0^1 f_i(x_i)dx_i = 1$ so that we may consider $f_i(x_i)$ as a density function. The corresponding distribution function is $F_i(x_i) = x_i^{\alpha_i+1}$ whose inverse is $x_i = F_i^{1/(1+\alpha_i)}$ assuming $\alpha_i > -1$ By Sklar's theorem (see Sklar, 1959; Faugeras, 2013; and Durante et al., 2015) we can then express the density function as:

$$f(\boldsymbol{x}) = \prod_{i=1}^{n} \left\{(1 + \alpha_i)x_i^{\alpha_i}\right\} \cdot c(\boldsymbol{U}; \beta), \tag{11}$$

where $\beta \in \mathcal{B} \subseteq \Re^d$ is a parameter vector, $c(\boldsymbol{U}; \beta)$ is a Copula function, and $\boldsymbol{U} = [U_1, \dots, U_n]^\top$, $U_i = x_i^{\alpha_i+1}$. Therefore:

$$\log f(\boldsymbol{x}) = \sum_{i=1}^{n} \log(1 + \alpha_i) + \sum_{i=1}^{n} \alpha_i \log x_i + \log c(\boldsymbol{U}; \beta). \tag{12}$$

To specify the Copula function, we use a multivariate mixture of normal distributions:

$c(\boldsymbol{U};\beta) =$

$$= \sum_{g=1}^{G} p_g \,(2\pi)^{-d/2} |\Omega_g|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{U}-\boldsymbol{\mu}_g)' \Omega_g^{-1} (\boldsymbol{U}-\boldsymbol{\mu}_g)\right\}, \tag{13}$$

where $p_g \geq 0$, $\sum_{g=1}^{G} p_g = 1$ are mixing probabilities and $\mu_{\boldsymbol{g}}, \Omega_g$ are a $d \times 1$ vector and a $d \times d$ covariance matrix, respectively. Here, $g = 2, \dots, G$ is the number of mixing components and it is to be selected along with the other Copula parameters which are denoted by β. We parametrize Ω_1 in terms of each Cholesky factor. The remaining matrices are $\Omega_g = h_{g1}\Omega_1 + h_{g2}\boldsymbol{I}$, where h_{g1}, h_{g2} are unknown positive constants (to be estimated). This parametrization reduces considerably the number of parameters as each Ω_g has $\frac{d(d+1)}{2}$ unknown elements.

4. Functional forms

In this section, we describe the most common flexible functional forms used in the literature to model cost functions (see Fuss and McFadden, 1978; Barnett and Lee, 1985; and Diewert and Wales, 1987). We start with the translog cost function which takes the form:

$$\log\, C_{TL}\,(W,Y) = \beta_o + \sum_{j=1}^{J} \beta_{wj} \log\, W_j + \sum_{k=1}^{N} \beta_{yk} \log\, Y_j + \frac{1}{2}\sum_i \sum_j \beta_{ww,ij} \log\, W_i \log\, W_j + \frac{1}{2}\sum_i \sum_j \beta_{yy,ij} \log\, Y_i \log\, Y_j + \sum_j \sum_k \beta_{wy,jk} \log\, W_j \log\, Y_k\,. \tag{14}$$

where W is a vector of input prices, Y is a vector of outputs, and $\beta_{ij}s$ are unknown parameters. Linear homogeneity in prices requires:

$$\sum_{j=1}^{J} \beta_{wj} = 1,\ \sum_j \beta_{ww,ij} = 0\ \forall i,\ \sum_k \beta_{wy,jk} = 0\ \forall j \tag{15}$$

The Generalized Leontief due to Diewert (1971) for two outputs is given by:

$$C_{GL}(W,Y) = Y_1 \sum_{j=1}^{K} \beta_{1,jj} W_j + Y_1 \sum_{k\neq j} \sum_{j\neq k} \beta_{1,kj} W_k^{1/2} W_j^{1/2} +$$
$$Y_2 \sum_{j=1}^{K} \beta_{2,jj} W_j + Y_2 \sum_{k\neq j} \sum_{j\neq k} \beta_{2,kj} W_k^{1/2} W_j^{1/2} +$$
$$Y_1^2 \sum_{j=1}^{K} \beta_{1,j} W_j + Y_2^2 \sum_{j=1}^{K} \beta_{2,j} W_j + \tag{16}$$
$$Y_1 Y_2 \sum_{j=1}^{K} \beta_{12,j} P_j + \sum_{j=1}^{K} \alpha_j P_j.$$

See also Hall (1973). This specification permits flexibility in economies of scale while it also imposes no a priori restrictions on the elasticities among factor inputs; see also Li and Rosenman (2001). It is also homogeneous of degree one in prices. The minflex (translog) Laurent expansion is:

$$\log\ C_{MTL} = \log\ C_{TL}\,(W,Y) - \sum_{i\neq j,k>j} \beta_{ijk} W_i^2 \left(W_j W_k\right)^{-1}, \tag{17}$$

where $\beta_{ij} \geq 0$, (see Barnett and Lee, 1985; Barnett, et al., 1985; and Diewert and Wales, 1987). Similarly, one can define the minflex (Generalized Leontief) Laurent expansion as (see Barnett and Lee, 1985; Barnett, et al., 1985):

$$\log\ C_{MGL} = \log\ C_{GL}\,(W,Y) - \sum_{i\neq j,k>j} \beta_{ijk} W_i^2 \left(W_j W_k\right)^{-1}. \tag{18}$$

Finally, we use the most common class of artificial neural networks known in the literature as Multilayer Perceptron (see Rosenblatt, 1962; Rumelhart et al., 1986; Desai et al., 1996; Decoste and Scholkopf, 2002; Hinton and Salakhutdinov, 2006; Ciresan, 2010; Kristjanpoller, and Minutolo, 2018) which is as follows. Suppose we have L layers, N units per layer, and K variables. Then the MLP is characterized as:

$$S_j^{(1)} = \sum_{k=1}^{K} w_{jk}^{(1)} x_k, \qquad j = 1,\ldots,N,$$

(19)

$$A_j^{(1)} = \vartheta_j^{(1)}\left(S_j^{(1)}\right), \quad j = 1,..,N,$$

$$S_j^{(l)} = \sum_{i=1}^{N} w_{ji}^{(l)} A_i^{(l-1)}, \;\; j = 1,\ldots,N,$$

(20)

$$A_j^{(l)} = \vartheta_j^{(l)}(S_j^{(l)} + b_l), l = 2, \ldots, L-1,$$

$$S_1^{(L)} = \sum_{i=1}^{N} w_{1i}^{(L)} (S_1^{(L)}),$$

$$A_1^{(L)} = \psi_1^{(L)} \left(S_1^{(L)}\right), \tag{21}$$

where $\vartheta_j^{(l)}$ is a sigmoid activation function for layer l and unit j, $\psi_1^{(l)}$ is a linear activation function, and $w_{ji}^{(l)}$ is the weight from unit i of the previous layer to unit j in layer l. (19) refers to the first layer, (20) refers to the intermediate layers and (21) to the final layer.

Finally, we should note that, as with other functional forms in production analysis, it is difficult to impose curvature restrictions globally. For example, curvature does not hold globally for (14), (16), (17), and (18). For the curvature constraints, we know that one should avoid enforcing it at all points as flexibility is then compromised. For example, the translog reduces to the Cobb-Douglas which is, of course, not flexible. Imposing curvature at a few points it is difficult for all functional forms like for (14), (16), (17), and (18), and could be conducted as follows: First, select a number of points, say the means of the data. Second, compute the Hessian of the cost function with respect to prices. Third, check if the curvature conditions hold. If not, reject the particular draw in Bayesian Markov Chain Monte Carlo (MCMC) and take another draw until the constraint is satisfied. To the best of our knowledge, there is no other procedure that curvature can be imposed in flexible functional forms unless other simplifying assumptions are to be made. In our work, we enforce curvature at the means only. For the MLP, imposition of such constraints is very difficult and we leave it for future research.

5. Data and empirical results

5.1 Data

Our sample includes quarterly data on all commercial banks insured by the Federal Deposit Insurance Corporation (FDIC) insured commercial banks from Call Reports available from the Federal Reserve Bank of Chicago, from 2001: Q1 to 2010: Q4. The data has been also used by Malikov et al. (2016).

Commercial banks can be starkly different from one another as regards to the size, capitalization, regulatory environment, etc., indicating potential heterogeneity in production technologies across banks (see Mester, 1997; Bos, et al., 2009; Dong et al., 2016). To deal with heterogeneity, we focus on a sample of banks with total assets in excess of one billion US. dollars (in 2005 prices) in the first three years of observation, which is supposedly a homogeneous sample. We exclude internet banks, commercial banks dealing primarily with credit card activities and banks operating outside the continental US. We also exclude observations with negative values for assets, equity, outputs and input prices, since these are likely to have resulted because of erroneous data reporting. The remaining unbalanced panel contains 2,397 observations for 285 banks. All nominal stock variables are deflated to 2005 US. dollars using the Consumer Price Index (for all urban consumers).

We use the "intermediation approach" of Sealey and Lindley (1977), to model the production technology of a bank. According to this method, a bank's balance sheet expresses the main structure of its core business.[5] Liabilities (core deposits and purchased funds) together with physical capital and labor are the inputs to the bank's production process, whereas assets (loans and trading securities) are the outputs. We extend the basic framework of modeling banking technology by acknowledging that the bank's production of desirable outputs, such as earning loans, is usually followed by the simultaneous by-production of non-performing loans, which are treated as an undesirable output (formally, this is treated as an input following Malikov et al., 2016). We use five outputs, viz. y_1 which is consumer loans, y_2 is real estate loans, y_3 is commercial and industrial loans, y_4 is securities, and y_5 is off-balance-sheet income. These are the desirable outputs following Berger and Mester (1997, 2003) and Hughes and Mester (1998, 2013). Total non-performing loans (b) is the bank's undesirable output. Our variables inputs are labor, x_1, (the number of full-time equivalent employees), x_2 (physical capital), x_3 (purchased funds), x_4 (interest-bearing transaction accounts), and x_5 (non-transaction accounts). We further include financial (equity) capital (e) as an additional input. However, we follow Berger and Mester (1997, 2003) and Feng and Serletis (2009) in modeling e as a quasi-fixed input, due to unavailability of each price. Treating equity capital as an input to banking production technology is in line with the argument that banks may use it as a cushion against losses (see Hughes and Mester, 1993, 1998). We divide total expenses on each input by the corresponding input quantity to determine prices of variable inputs (w_1 through w_5).

We compare results obtained through the HDMR and MLP approximations with our novel Copula Approximation along with three well-known functional forms: translog, GL, and minflex Laurent. In all cases, we use Bayesian MCMC techniques to perform the computations, namely the Girolami and Calderhead (2011) MCMC Riemannian Manifold Hamiltonian technique, which uses first - and second-order derivative information from the log posterior. The method is briefly described in the Technical Appendix. We may, of course, use sampling-theory based approaches to estimate the model. The advantage of the Bayesian method is that it provides access to exact, finite—sample densities and, additionally, it is not subject to the risk of being stuck in local optima as it operates, more or less, like a simulated annealing method. In all cases, we use 150,000 MCMC passes the first 50,000 of which are discarded to mitigate possible start up effects. We calibrate the MCMC algorithm so that its acceptance

[5] See also Berger and Mester (1997) and Wanke et al., (2016).

rate is 20-30% in the pre- and post-burn phases (through the constant ε described in the Appendix). We examined convergence thoroughly (starting from different initial conditions). The results are not reported here in the interest of brevity but they are available on request.

In HDMR, MLP and Copula Approximation we determine the various constants involved (number of terms for HDMR, L and N for MLP, and G for Copula Approximation) using Bayes factors resulting from computation of marginal likelihood of each model. We find L=4, N=5, and G=3. To get a sense of the number of parameters involved, the translog and similar functional forms, with five outputs, four relative input prices, one bad output and a time trend, has 75 parameters. The copula approach has 244 parameters if all elements of $\Omega_g, g = 1, \ldots, G$ are unrestricted and 116 when the covariance matrices are restricted as in the discussion following (13).

Below we define the measures of interest. In what follows, *t* represents a time trend. The cost function $C(W, Y, t)$ is defined as follows:

$$C(W, Y, t) = \min_{X \in R_{++}^K} : W'X, s.t.\ F(X, Y, t) \leq 1,$$

for some transformation function, $F(X, Y, t)$, which represents the feasibility of plans (X, Y), i.e. inputs and outputs. In turn, we define returns to scale (*RTS*) as: $RTS^{-1} \equiv \sum_{m=1}^{M} \frac{\partial lnC(W,Y)}{\partial lnY_m}$; viz. the reciprocal of the output cost elasticity, technical change (*TC*) as: $TC \equiv \frac{\partial \ln(C,W,t)}{\partial t}$; cost efficiency (r_{it}) as: $r_{it} \equiv e^{-u_{it}}$ and efficiency change (*EC*) as: $EC = \frac{r_{it} - r_{i,t-1}}{r_{i,t-1}}$; u_{it} is cost inefficiency which we estimate from cost function residuals (say V_{it}) as follows: $u_{it} = V_{it} - \min_{i,t} V_{it}$.

5.2 Discussion of Results

This section presents the empirical results. As we have discussed, different approaches have been suggested in the literature in order to estimate productivity growth and its components (returns to scale, cost inefficiency, technical change, and efficiency change), however, there is no consensus as to what one should do in practice. We use an HDMR, a translog, a minflex Laurent, a GL cost function and MLP as benchmarks to compare our results with the Copula approximation. Comparisons with these models will help us understand the unknown functional form.

5.2.1. Components of productivity growth

We distinguish between different components, namely input elasticities for each of the input prices (w_1 through w_5), returns to scale (Figure 1a), cost inefficiency (Figure 1b), technical change (Figure 1c) and efficiency change (Figure 1d). Figures 1a–1d report the sampling distributions of these four bank-specific aspects of our model and in Table 1 appear their summary statistics.

In Figure 1a we present the sampling distributions of posterior mean returns to scale from the Copula model and the four benchmarks. In the Copula (HDMR) model, returns to scale average near 0.95 (unity) and extend from about 0.9 (0.95) to 1(1.05), as we see in Table 1, which is a reasonable estimate. The Copula and the HDMR models show higher returns to scale estimates from the other three traditional production functions. The other four models produce very different estimates from the Copula and the HDMR models but very similar among them. All of them average around 0.82 and extend from 0.65 to 0.90 for (minflex). The Copula method delivers higher returns to scale, as shown in Figure 1a. The benchmark methods underestimate returns to scale compared to Copula, with only exception the HDMR approach. From Figure 1a and Table1, it is clear that the sampling distribution of returns to scale estimated using the Copula model has the lowest standard deviation (s.d.) and using the HDMR comes second, while the minflex has the highest s.d.

Table 1.

Summary statistics of Bank-specific characteristics.

	RTS	u_{it}	TC	EC	PG=TC+ EC
TL	0.8180	0.2418	0.0030	-0.0111	-0.0082
	(0.0186)	(0.0133)	(0.0182)	(0.0016)	(0.0183)
Minflex	0.8126	0.2853	-0.0243	-0.0105	-0.0348
	(0.0236)	(0.0627)	(0.0192)	(0.0026)	(0.0201)
GL	0.8174	0.2204	-0.0159	0.2204	-0.0257
	(0.0165)	(0.0435)	(0.0277)	(0.0019)	(0.0282)
HDMR	1.0072	0.1263	-0.0073	-0.0121	-0.0195
	(0.0128)	(0.0191)	(0.0136)	(0.0101)	(0.0129)
Copula	0.9439	0.1298	0.0114	0.0123	0.0237
	(0.0163)	(0.0551)	(0.0104)	(0.0051)	(0.0102)

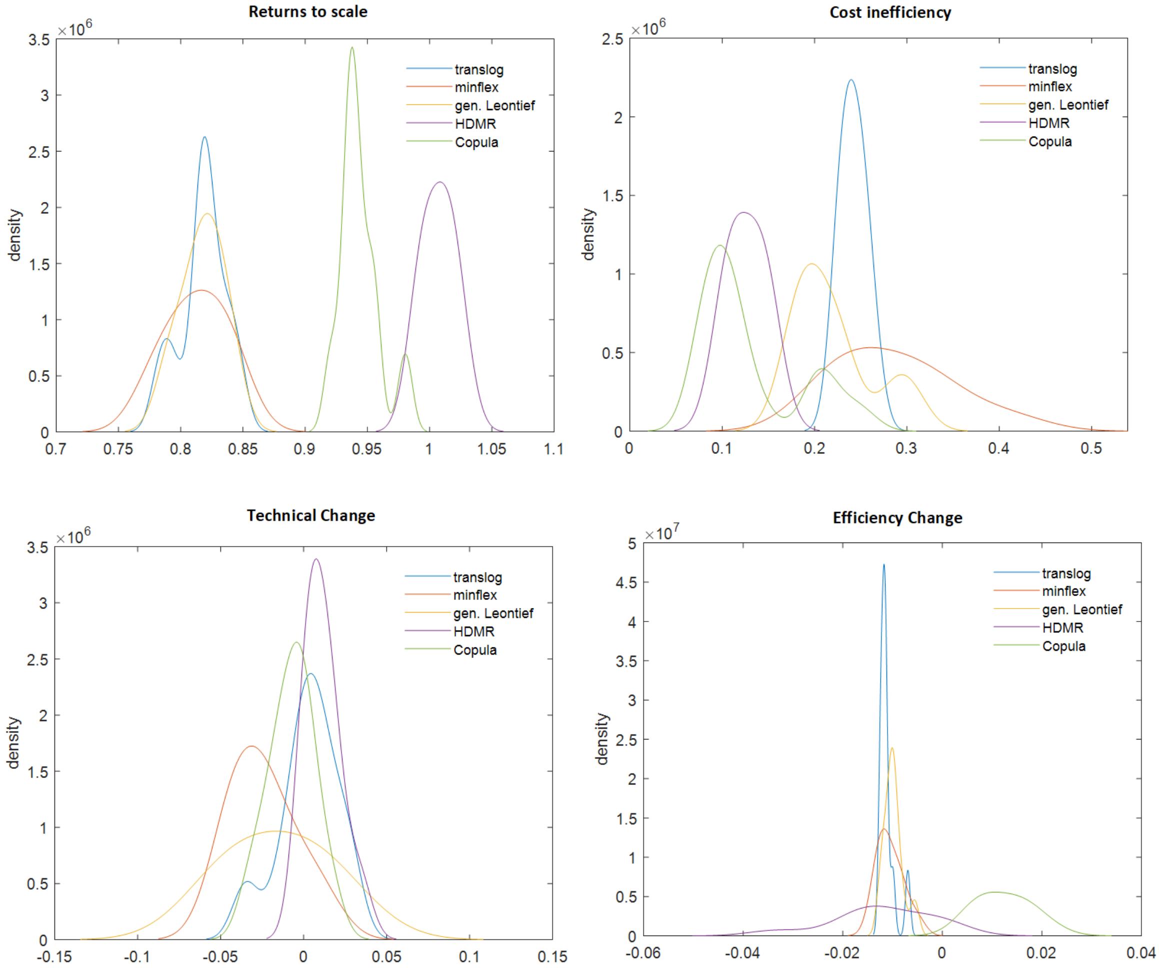
Notes: The Table provides the summary statistics (posterior mean and posterior standard deviation) of bank-specific characteristics, namely, returns to scale (*RTS*), cost inefficiency (u_{it}), Technical change (*TC*), Efficiency change (*EC*) and productivity growth (PG), which is the sum of cost and efficiency change. The summary statistics have been calculated for five different models, Translog (TL), Minflex Laurent, Generalized Leontief (GL), High-Dimensional Model Representation (HDMR) and Copula Approximation. Posterior standard errors of estimates are reported in parentheses.

Sampling distributions of posterior mean cost inefficiencies are reported in Figure 1b. The sampling distribution from the Copula and GL are clearly bimodal. According to the Copula model (HDMR), cost inefficiency estimates average around 0.13 (0.13) (see Table 1) and extending from near zero to 0.30 (0.20), see Figure 1. Sampling distributions of posterior mean cost inefficiencies for translog average around 0.25 and extend from 0.20 to 0.30. The sampling distribution of cost inefficiency estimated using the translog model has the lowest s.d. and using the HDMR comes second, while the minflex Laurent has the highest s.d. From Table 1, according to the Copula and HDMR models, cost inefficiency averages 13% while according to the traditional flexible functional forms cost inefficiency averages much higher, from 22% (for GL) to 29% (for minflex) which is quite common in many previous banking studies. This puts into question the highly popular view that banks are quite cost-inefficient (25-30%).

In Figure 1c, we present the sampling distributions of posterior mean technical change. The Copula and the HDMR methods give more or less similar results with the translog function for technical change, with estimates being concentrated around zero and extending from -5% to 5%. The other flexible functional forms are concentrated around negative values and extend from -10% to 10% with the GL having the highest s.d. The sampling distribution of technical change estimated using the Copula model has the lowest s.d. and the HDMR comes second (see Table 1).

The results regarding efficiency change are quite interesting (Figure 1d), as the HDMR and the Copula have the highest s.d., while all the flexible functional forms are concentrated around zero with a very low s.d., of which the lowest is for the translog. However, the translog and the GL are bimodal revealing the possible presence of multiple groups in the data. All the benchmark models deliver an average efficiency change very close to zero, while the Copula method suggests that efficiency change was positive averaging 1.23% and ranging roughly from slightly below zero to almost 3%.

Figure 1.



5.2.2. Productivity growth.

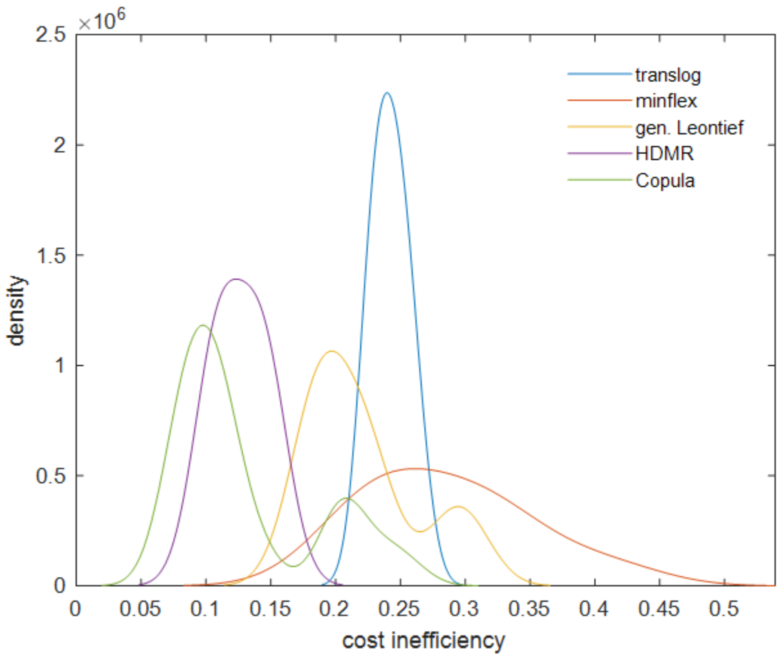
Having discussed the components of productivity growth, now we discuss productivity growth per se. Figure 2, reports the sampling distribution of productivity growth which is the sum of technical and efficiency change, and Table 1 provides its posterior mean and posterior s.d. The estimated productivity growth from the Copula method proposed in this study is compared with the other four methods as benchmarks.

According to the Copula model, productivity growth estimates are concentrated around positive values with an average of 2% and extend from near zero to 5–6%. According, however, to the HDMR the minflex Laurent, and a generalized Leontief the estimates are concentrated around negative values with an average of around -3% (for minflex) to -2% (for HDMR), and extending from -12% to 15% for GL and from -6% to 4% for HDMR. Some of these findings for HDMR and traditional flexible functional forms are, clearly, unreasonable. While for the translog function, the productivity growth estimates are concentrated around zero and extend from -6% – 4%. From Figure 2 and Table 1, it is clear that the sampling distribution of productivity growth estimated using the Copula model has the lowest spread. Using the HDMR comes second, while the generalized Leontief has the highest s.d.

So, in terms of estimating productivity growth, all flexible functional forms give more or less, similar results with the translog function perhaps with the exception of the GL. However, the Copula Approximation receives much greater support in the light of the data, even compared to HDMR.

The Copula method delivers higher productivity growth as shown in the distribution plot, and the benchmark models underestimate productivity. According to HDMR, we obtain an average productivity growth of around -2%, whereas the Copula method suggests that productivity growth was positive and around 2%.

Figure 2.



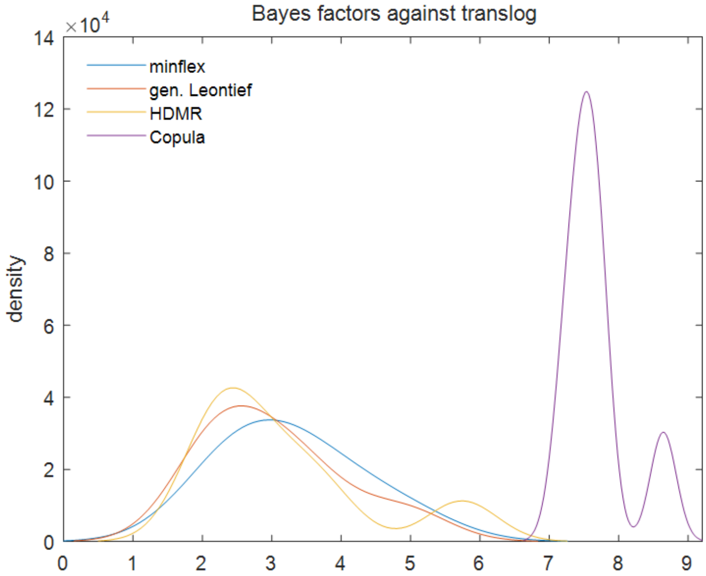
5.2.3. Bayes model comparisons.

To address the question of model comparison, we omit a random number B of banks (viz. all their temporal observations, B ranging from 1 to 20 with equal probability) and we re-estimate the model without these observations. We perform this exercise 10,000 times.

We use the concept of Bayes factor which takes into account both model fit and model complexity.[6] We compute the predictive Bayes factor in favor of each of the four models (Copula, HDMR, GL and minflex Laurent) against the translog model, and plot the densities of Bayes factors in Figure 3.

[6] The Bayes factor is defined as the ratio of the posterior probabilities of the null and alternative hypothesis, ranging from zero to infinity (see Assaf and Tsionas, 2018).

Figure 3.



Summary statistics for the predictive Bayes factors appear in Table 2. Clearly, all flexible functional forms are equivalent, more or less, to the translog, but the Copula Approximation receives much greater support in the light of the data. This result is quite important as it shows that the Copula Approximation performs better even compared to HDMR (the second best performing approximation) as its Bayes factor (relative to translog) has an average of 7.66 and ranges from 7.19 to 8.64 and indicates substantial evidence in favor of the Copula Approach. The Bayes factor for the parametric forms and HDMR ranges from 2.01, for HDMR, to 5.74, again for HDMR.[7]

Table 2.

Summary statistics of Predictive Bayes factors.

Method	Mean	Median	Minimum	Maximum	Std. Dev.
Minflex	3.2836	3.2579	2.2449	4.9304	0.9466
GL	3.0361	2.8258	2.0783	4.9037	1.0055
HDMR	3.1663	2.7642	2.0149	5.7430	1.2890
Copula	7.6661	7.5317	7.1964	8.6435	0.4685

Notes: This Table provides the summary statistics of the Predictive Bayes Factors in favor of each of the four models for Minflex, Generalized Leontief (GL), High-Dimensional Model Representation (HDMR) and Copula Approximation, all of them against the translog model.

Regarding timing, we provide evidence in Table 3. All computations were performed using Fortran 77 (GNU compiler) with access to IMSL and NAG software libraries on an Intel i9-7900 CPU @ 3.30 GHz with RAM 32 GB.

[7] A Bayes factor between 3 and 10 is interpreted as substantial evidence, while a A Bayes factor of 1 is interpreted as no evidence (see Assaf and Tsionas, 2018).

Table 3.

Timing statistics for various models.

Model	Timing (CPU in seconds)
TL	1.220
Minflex	1.351
GL	1.366
HDMR	2.215
Copula	16.332

Notes: This Table provides timing statistics for translog (TL), Minflex, Generalized Leontief (GL), High-Dimensional Model Representation (HDMR) and Copula Approximation.

Of course, copula approximations are more computationally demanding but they seem to be quite feasible in terms of implementation on modern desktop computers. The MLP approximation with its optimal settings for L and N was computationally far more demanding, taking over 2,000 seconds of central processing unit (CPU) time.

To investigate the issue of correlations of cost inefficiency among alternative models, we save for each MCMC draw the rank correlation coefficient for any two pairs of simulated inefficiencies for the Copula, Minflex, GL, HDMR, and the translog. In turn, we average across all MCMC draws (after the burn-in phase) to obtain the posterior mean rank correlation coefficients, reported in Table 4. All posterior s.d.s of these measures were of the order 10^{-3}.

Table 4.

Rank correlation coefficients of u_{it}s.

Method	TL	Minflex	GL	HDMR	Copula
TL	1.000	0.971	0.986	0.312	0.313
Minflex		1.000	0.944	0.413	0.317
GL			1.000	0.382	0.212
HDMR				1.000	0.773
Copula					1.000

Notes: This Table provides rank correlations for the translog (TL), Minflex, Generalized Leontief (GL), High-Dimensional Model Representation (HDMR), and the Copula Approximation.

Evidently, rank correlations are quite high between translog, Minflex and GL but not so for these parametric models versus HDMR and copula approximation. HDMR and copula approximation have higher rank correlation which is 0.773.

6. Monte Carlo experiment

In this section, we set up a Monte Carlo experiment where the true functional form is either a translog, a GL, a minflex Laurent or a minflex translog.

(a) In the first step, we randomly select parameters for the functional forms.

(b)Then, we fit a translog, HDMR, MLP and Copula Approximation. All the parameters like L, N, G, are selected based on marginal likelihood delivered by the Gorilami and Calderhead (2011) MCMC technique to perform the computations.

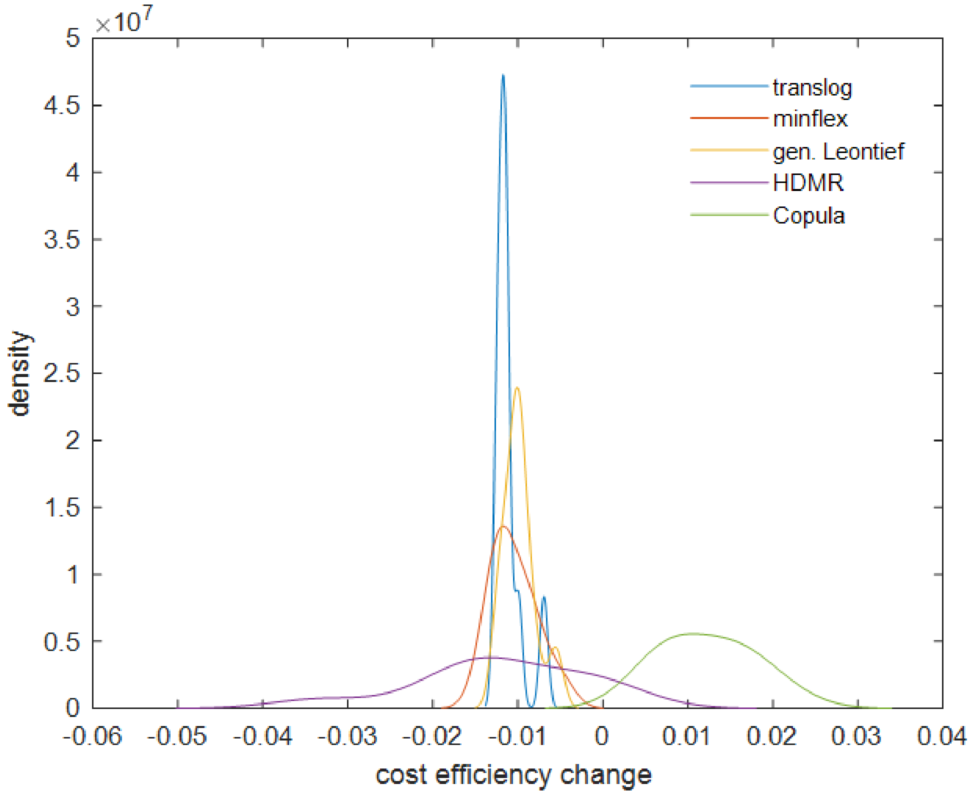
(c) In the third step, we compute maximum signed absolute errors for each approximation and we select the specification with the least minimum absolute errors.

(d) In the fourth step, we select another parameter set and go to step (b). The process is repeated 10,000 times.

We assume that we have two outputs and three input prices. Artificial data for these are produced from lognormal distributions with location parameter zero and scale parameter one, and they are equi-correlated with common correlation coefficient 0.90. We keep generating data sets until they satisfy the monotonicity constraints and curvature restrictions imposed at the mean of data. We add normally distributed noise with zero mean and s.d. equal to 10% of the generated cost series. We abstract from cost inefficiency as our focus is on approximation of functional forms.

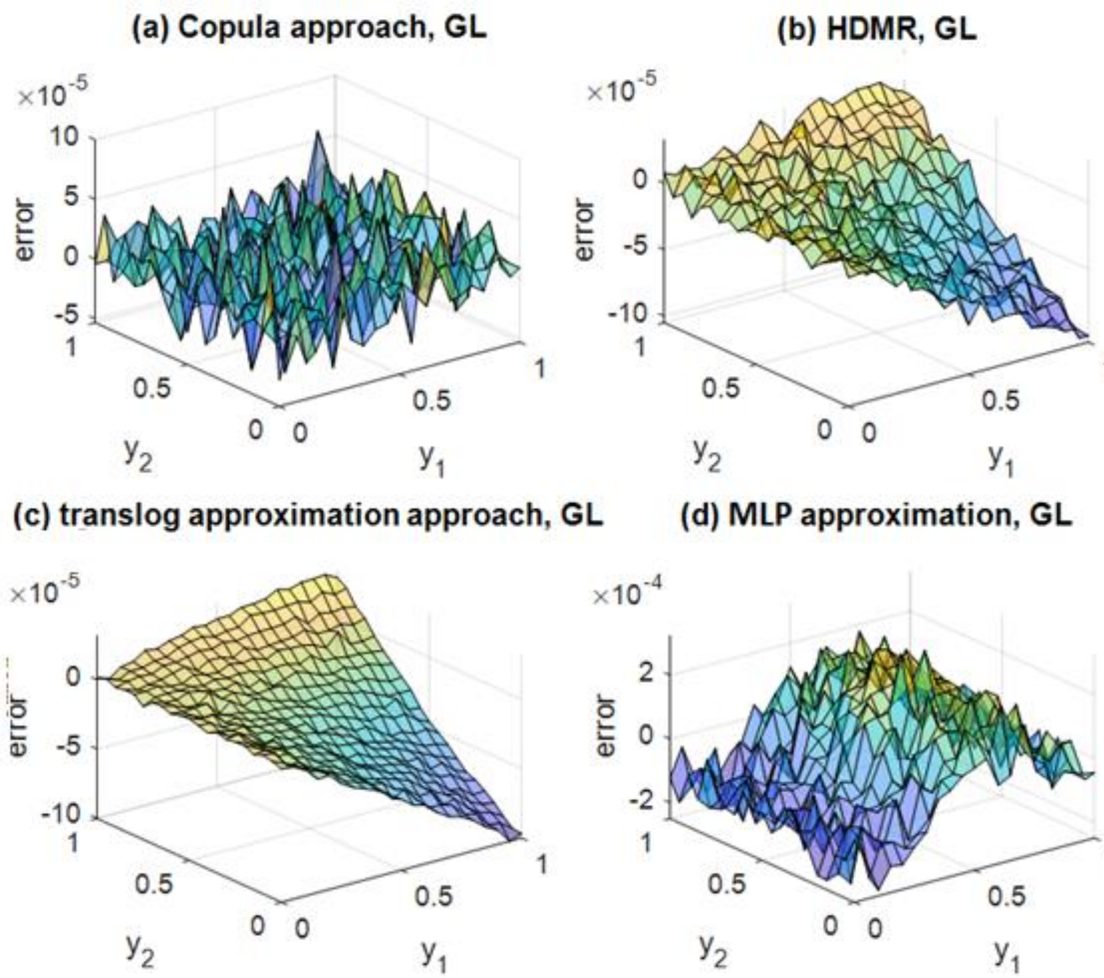
Next, we present errors from the various approximations when we use a Copula, HDMR, translog or an MLP approximation. In all cases, we present the maximum signed absolute error as a function of y_1 and y_2 when the other variables are fixed at their median values. Errors produced by Copula approach, in Figure 4, are quite small (although much smaller for the HDMR and roughly the same compared to a translog form used to estimate from data generated by the translog). No specific pattern is visually evident in Figure 4, so all approximations perform well but the best approximation are the Copula.

Figure 4.



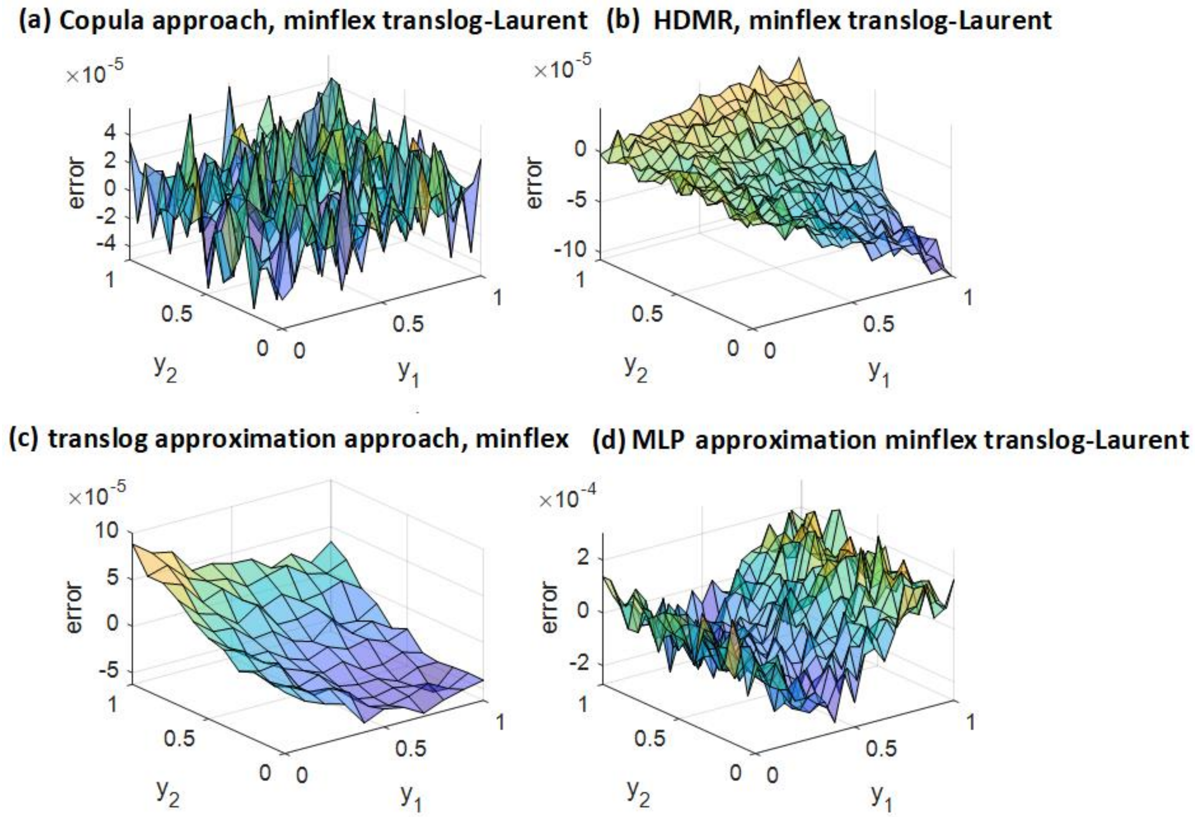
×10^7
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
density
-0.06
-0.05
-0.04
-0.03
-0.02
-0.01
0
0.01
0.02
0.03
0.04
cost efficiency change
translog
minflex
gen. Leontief
HDMR
Copula

Figure 5.



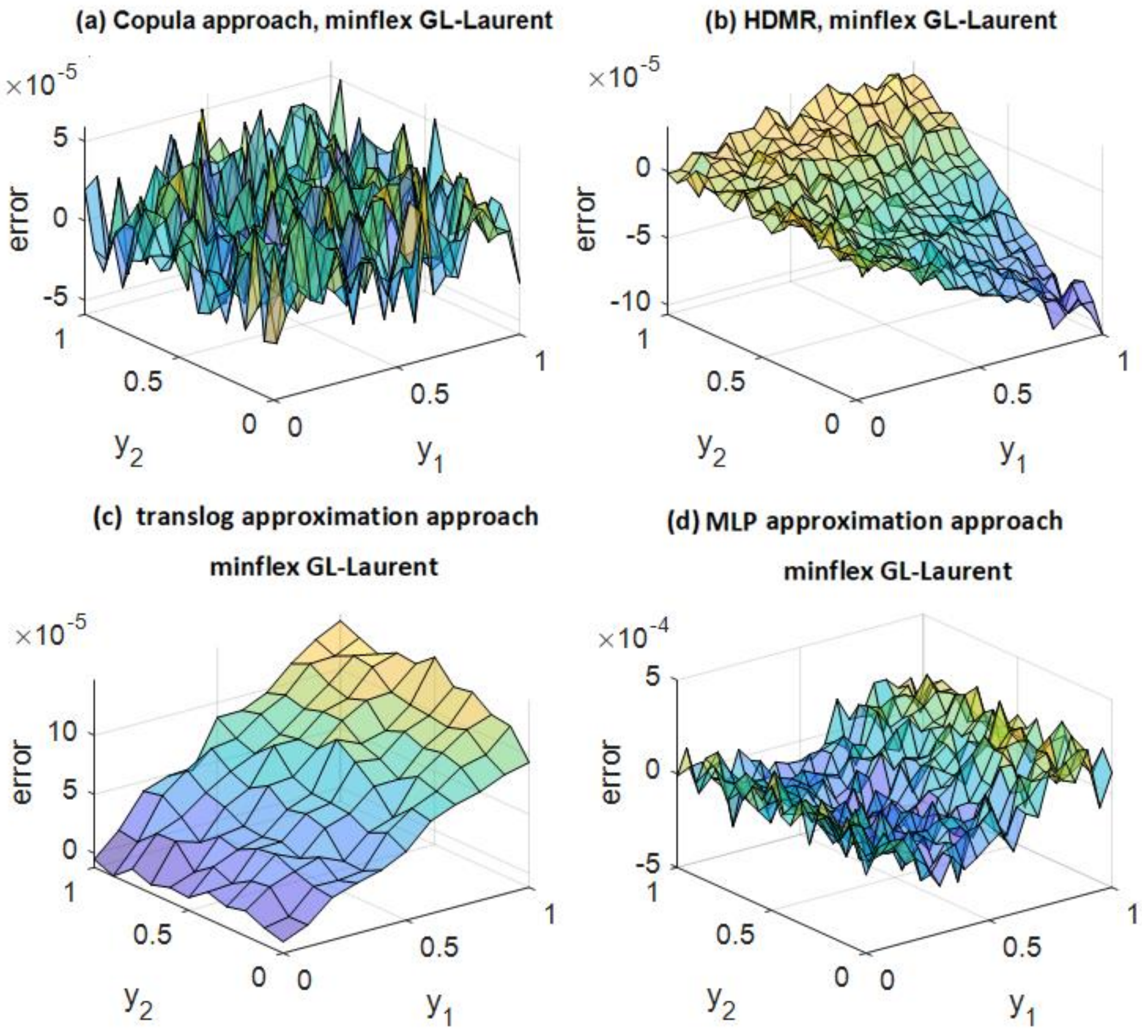
When we use the Generalised Leontief as the data generating process (see Figure 5), the Copula approach delivers small errors without a specific pattern although all the other approximations deliver higher errors that follow a pattern. The HDMR and translog approximations are good at the lowest values of outputs but become worse as their values increase. The MLP approach produces also errors that follow a quadratic pattern and they are larger than the Copula approach and these errors are higher in the center of the data.

Figure 6.



When we use the minflex Laurent as the data generating process (see Figure 6), again, the Copula approach delivers the smallest errors without a pattern but this is not the case for the HDMR, translog or MLP approximations. It appears that, in this instance, the HDMR and the translog approximations perform better near the center of the data, and the MLP approximation performs well in the same part but the approximation at the extremes is not accurate.

Figure 7.



Finally, we use the minflex Laurent as the data generating process (see Figure 7), the good performance of the Copula approach is, again evident, although the HDMR, translog, and MLP have higher signed absolute errors with a non-random pattern (see Figure 7). In our view, this provides compelling evidence in favor of a Copula approximation and against HDMR or, perhaps surprisingly, even for the MLP approximation.

In Figure 8, we show the distribution of the number of layers and the number of units/layer when we use the Copula Approximation in the case of GL. This is in the context of our Monte Carlo experiment. The modal value of the number of layers is 4 although in many cases we need as many as 10. The number of units per layer is 5 to 8, on the average, although 8-14 units per layer are needed. Similar patterns are observed in all other flexible functional forms. They are not discussed here as they are available on request.

In Table 5 we report root-mean-squared errors (RMSE) for the different approximations used in Figures 4 – 7.

Table 5.

Root-mean-squared errors of approximations in Figures 4 - 7.

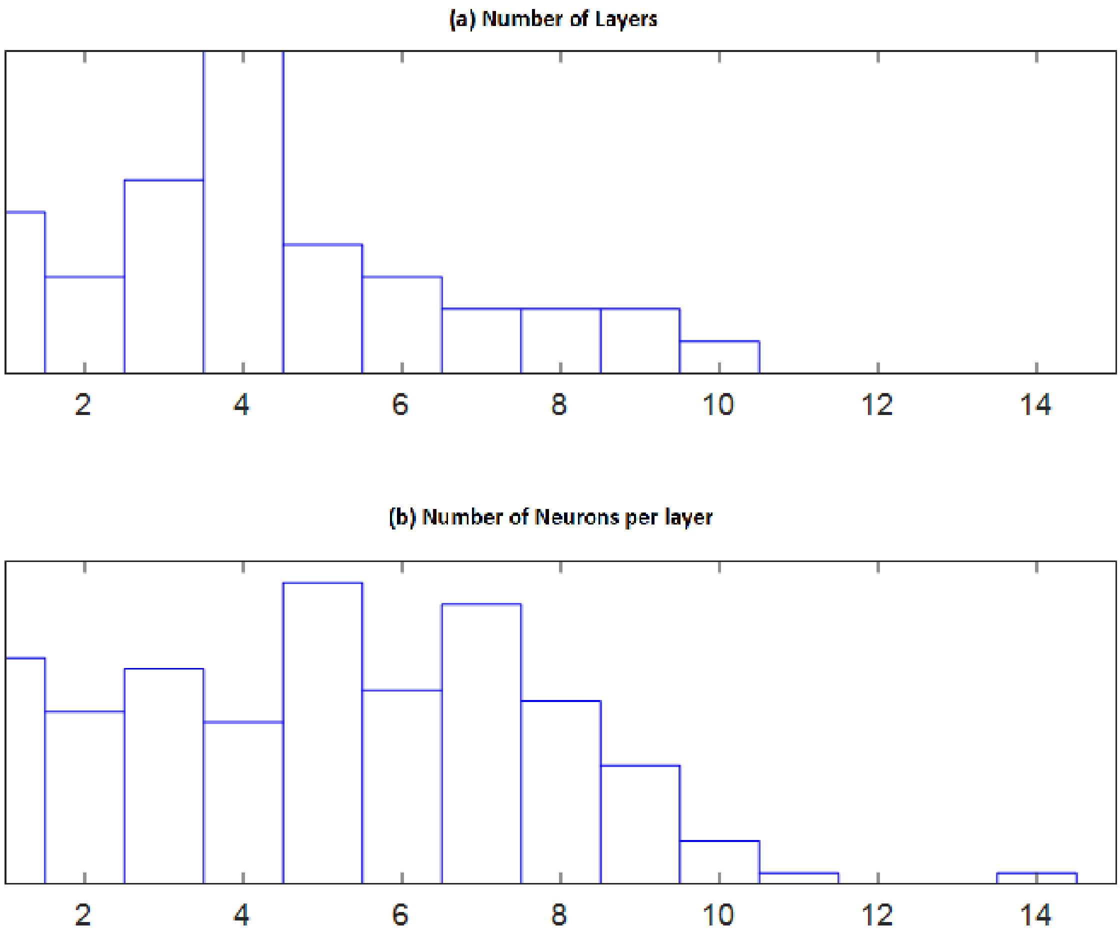
Method	Copula	HDMR	TL	MLP
TL	0.12	0.31	0.42	3.1
GL	0.10	5.42	5.50	0.67
Minflex TL	1.21	5.53	5.30	0.12
Minflex L	0.10	5.41	5.22	0.33

Notes: All numbers should be multiplied by 10^{-5}. This Table provides mean absolute errors for each of the four models: Minflex, Generalized Leontief (GL), High-Dimensional Model Representation (HDMR) and Copula Approximation. The rows correspond to specifications in Figures 4 – 7. TL stands for translog and L for Laurent.

It is necessary to stress the similarities with the Gallant results and provide a hint (besides the Monte Carlo experiment) towards the theoretical capacity of HDMR approaching the true function, which reasonably might be due to the orthonormality of the polynomials used.[8] As our main model relies on the Copula approach, there are still similarities with the Gallant results in that as G increases in (12) and (13) any functional form can be approximated to an arbitrary degree of accuracy. This follows from the standard results of Hornik, Stinchcombe, and White (1989) who show that similar models are "capable of approximating any Borel measurable function from one finite-dimensional space to another to any desired degree of accuracy" (op. cit., p. 359) with a sufficiently large G. A theorem by Kolmogorov states that for any N-variate continuous function the unit cube can be written as the sum of $2N+1$ functions, where each function depends on a single variable. Although there does not exist, as of yet, a constructive approximation based on this idea, MLP, HDMR and the Copula approach rely on this concept. Gallant (1981) actually proves that a consumer's indirect utility function must of the Fourier form. The same result applies to cost or other functions see his first equation in p. 219. We believe that this idea has been underexplored in the literature and definitely deserves further investigation in terms of general functional approximations.

[8] We are indebted to an anonymous reviewer for making this important comment.

Figure 8. Number of layers and number of units per layer in Monte Carlo experiment



Another important issue is to report parameter issues for the different true data generating processes (DGP) and their various approximations. As parameters themselves do not have a structural interpretation we report partial derivatives of the cost function with respect to input prices and outputs. As we mentioned before we have 3 input prices and two outputs. Of course, linear homoigeneity with respect to prices has to be imposed. The results are presented in Table 6 Panels A and B, for input prices and outputs respectively.

Table 6.

Panel A: **Features of the different approximations, input price elasticities**

Approximations →	TL		HDMR		MLP		Copula	
DGP↓	$\frac{\partial \ln C}{\partial \ln w_1}$	$\frac{\partial \ln C}{\partial \ln w_2}$	$\frac{\partial \ln C}{\partial \ln w_1}$	$\frac{\partial \ln C}{\partial \ln w_2}$	$\frac{\partial \ln C}{\partial \ln w_1}$	$\frac{\partial \ln C}{\partial \ln w_2}$	$\frac{\partial \ln C}{\partial \ln w_1}$	$\frac{\partial \ln C}{\partial \ln w_2}$
	0.51	0.25	0.51	0.25	0.51	0.25	0.51	0.25
TL	(0.49)	(0.24)	(0.63)	(0.34)	(0.43)	(0.35)	(0.48)	(0.21)
	0.47	0.32	0.47	0.32	0.47	0.32	0.47	0.32
GL	(0.35)	(0.31)	(0.38)	(0.48)	(0.36)	(0.44)	(0.46)	(0.29)
	0.51	0.38	0.51	0.38	0.51	0.38	0.51	0.38
minflex L	(0.33)	(0.55)	(0.30)	(0.14)	(0.44)	(0.25)	(0.48)	(0.35)
	0.38	0.45	0.38	0.45	0.38	0.45	0.38	0.45
minflex TL	(0.27)	(0.31)	(0.22)	(0.39)	(0.24)	(0.32)	(0.37)	(0.44)

Panel B: Features of the different approximations, output elasticities

Approximations →	TL		HDMR		MLP		Copula	
DGP↓	$\frac{\partial \ln C}{\partial \ln y_1}$	$\frac{\partial \ln C}{\partial \ln y_2}$	$\frac{\partial \ln C}{\partial \ln y_1}$	$\frac{\partial \ln C}{\partial \ln y_2}$	$\frac{\partial \ln C}{\partial \ln y_1}$	$\frac{\partial \ln C}{\partial \ln y_2}$	$\frac{\partial \ln C}{\partial \ln y_1}$	$\frac{\partial \ln C}{\partial \ln y_2}$
	0.44	0.19	0.44	0.19	0.44	0.19	0.44	0.19
TL	(0.43)	(0.20)	(0.33)	(0.28)	(0.31)	(0.20)	(0.43)	(0.22)
	0.26	0.72	0.26	0.72	0.26	0.72	0.26	0.72
GL	(0.19)	(0.81)	(0.32)	(0.59)	(0.33)	(0.61)	(0.24)	(0.70)
	0.61	0.35	0.61	0.35	0.61	0.35	0.61	0.35
minflex L	(0.55)	(0.41)	(0.32)	(0.44)	(0.57)	(0.22)	(0.59)	(0.32)
	0.25	0.87	0.25	0.87	0.25	0.87	0.25	0.87
minflex TL	(0.14)	(0.79)	(0.19)	(0.74)	(0.44)	(0.52)	(0.22)	(0.81)

Notes: True values are stated in the Table. Values from the approximations are reported in parentheses. TL stands for translog, GL for Generalized Leontief, and L for Laurent.

Although MLP, Copula, and HDMR perform well in terms of function approximation (see Figures 4 – 7) this is not the case for the derivatives or elasticities reported in Tables 6 and 6b. Only the Copula approximation seems capable of getting these elasticities right on the average. We believe that this matter should be investigated further, although it seems clear that the Copula approximation performs best due to its flexibility despite the fact that it relies on multivariate mixtures of normal. Along with the results in Gallant (1981), we believe that more work is needed in these directions.

7. Conclusions

In this article, we considered High-Dimensional Model Representation (HDMR) as a way to analyze cost functions and we proposed a novel approximation of a cost function using Copulas. Using Copulas as approximations of functions appears to be novel at least in the case of modelling cost functions.

In an extensive Monte Carlo experiment, we find that the Copula Approximation performs much better, in terms of Bayes factors for model comparison, compared to HDMR and Multi-Layer Perceptrons (MLP). For the Copula Approximation, we provide a novel multivariate mixture-of-normals approximation. Given the good performance of the Copula Approximation in the Monte Carlo study, we undertake an empirical application to a sample of large US commercial banks, where the results are compared to those from a HDMR, an MLP, a translog, a minflex Laurent, and a Generalized Leontief (GL) function. The empirical application reveals that the HDMR and commonly used functional forms deliver different results compared to the Copula Approximation. Moreover, the latter receives significantly more support in the light of the data, based on predictive Bayes factors computed over left-out sub-samples of the data.

APPENDIX. Markov Chain Monte Carlo

The algorithm uses local information about both the gradient and the Hessian of the log-posterior conditional of $\boldsymbol{\theta}$ at the existing draw. A Metropolis test is again used for accepting the candidate so generated but the Gibbs sampling (GC) algorithm moves considerably faster relative to our naive scheme previously described. The GC algorithm starts at the first-stage GMM estimator and the MCMC runs until convergence. It has been argued (see Kumbhakar and Tsionas, 2016; Assaf et.al., 2018) that the performance of the GC algorithm is vastly superior to the standard Metropolis-Hastings (MH) algorithm and autocorrelations are much smaller.

Suppose $L(\boldsymbol{\theta}) = \log\, p\,(\boldsymbol{\theta}|\boldsymbol{X})$ is used to denote for simplicity the log posterior of $\boldsymbol{\theta}$, and $\boldsymbol{X}$ denotes the data and $\{\mathcal{U}_{it}, i = 1, \dots, n, t = 1, \dots, T\}$. The dimensionality of $\boldsymbol{\theta}$ is d_θ. Moreover, define the estimated covariance matrix:

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

which is the empirical counterpart of

$$\boldsymbol{G}_o(\boldsymbol{\theta}) = -\mathbb{E}_{Y|\boldsymbol{\theta}} \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \log\, p\,(\boldsymbol{X}|\boldsymbol{\theta}) \tag{A.2}$$

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

$$d\boldsymbol{\theta}(t) = \frac{1}{2}\tilde{\nabla}_{\boldsymbol{\theta}} L\{\boldsymbol{\theta}(t)\}dt + d\boldsymbol{B}(t) \tag{A.3}$$

where

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

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

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

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

$$\begin{aligned} \tilde{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_i^o + \frac{\varepsilon^2}{2}\{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}^o)\}_i + \frac{\varepsilon^2}{2}\sum_{j=1}^{d_\theta}\{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\}_{ij}\,\mathrm{tr}\left\{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\frac{\partial \boldsymbol{G}(\boldsymbol{a}^o)}{\partial\boldsymbol{\theta}_j}\right\} \\ -\varepsilon^2\sum_{j=1}^{d_\theta}\left\{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\frac{\partial\boldsymbol{G}(\boldsymbol{\theta}^o)}{\partial\boldsymbol{\theta}_j}\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\right\}_{ij} + \left\{\varepsilon\sqrt{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)}\boldsymbol{\xi}^o\right\}_i \\ = \boldsymbol{\mu}(\boldsymbol{\theta}^o,\varepsilon)_i + \left\{\varepsilon\sqrt{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)}\boldsymbol{\xi}^o\right\}_i \end{aligned}$$

where $\boldsymbol{\theta}^o$ is the current draw and ε is selected during the burn-in phase so that 20-30% of all candidates are, eventually, accepted. The proposal density is:

$$q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}^o) = \mathcal{N}_{K_\theta}\left(\tilde{\boldsymbol{\theta}},\, \varepsilon^2\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\right) \tag{A.6}$$

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

$$\min\left\{1, \frac{p(\tilde{\boldsymbol{\theta}}|\cdot,\boldsymbol{X})q(\boldsymbol{\theta}^o|\tilde{\boldsymbol{\theta}})}{p(\boldsymbol{\theta}^o|\cdot,\boldsymbol{X})q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}^o)}\right\}. \tag{A.7}$$

References

Alis, O., and Rabitz, H. (1999). General foundations of high dimensional model representations. Journal of Mathematical Chemistry, 25 (2), 197–233.

Assaf, A.G., and Tsionas, E.G. (2018). Bayes factors vs. P-values. Tourism Management, 67, 17–31.

Assaf, A.G., Li, G., Song,H., and Tsionas, E.G. (2018). Modeling and Forecasting Regional Tourism Demand Using the Bayesian Global Vector Autoregressive (BGVAR) Model. Journal of Travel Research, 58 (3), 383–397.

Badunenko, O., and Kumbhakar, S., C. (2017). Economies of scale, technical change and persistent and time-varying cost efficiency in Indian banking: Do ownership, regulation and heterogeneity matter?, European Journal of Operational Research, 260 (2), 789–803.

Barnett., WA, and Lee Y., W. (1985). The global properties of the minflex Laurent, generalized Leontief, and translog flexible functional forms. Econometrica, 53 (6), 1421–1437.

Barnett., WA, and Lee Y., W. (1987). The Laurent series approach to structural modeling. European Journal of Operations Research, 30 (3), 270–279.

Barnett W, Lee Y, W., and Wolfe, M. (1985). The three-dimensional global properties of the minflex Laurent, generalized Leontief, and translog flexible functional forms. Journal of Econometrics, 30 (1), 3–31.

Beccacece, F, Borgonovo, E, Buzzard, G, Cillo, A, and Zionts, S. (2015). Elicitation of multiattribute value functions through high dimensional model representations: Monotonicity and interactions. European Journal of Operational Research, 246 (2), 2–21.

Bos, J. W. B., Koetter, M., Kolari, J. W., and Kool, C. J. M. (2009). Effects of heterogeneity on bank efficiency scores. European Journal of Operational Research, 195 (1), 251–261.

Berger, A.N., and Mester, L. J. (1997), Inside the black box: What explains differences in the efficiencies of fmancial institutions? Journal of Banking and Finance, 21 (7), 895–947.

Berger AN, and Mester L. J. (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.

Cherubini, U., Luciano, E., Vecchiato, W. (2004). Copula Methods in Finance. Wiley, New York.

Christensen, L. R., Jorgenson, D. W., and Lau, L. J. (1971). Conjugate duality and the transcendental production function. Econometrica, 39, 206–218.

Ciresan, D.C., Meier, U., Gambardella, L.M., and Schmidhuber, J. (2010). Deep, big, simple neural nets for handwritten digit recognition. Neural Computation, 22 (12), 3207–3220.

Decoste, D., and Scholkopf, B. (2002). Training invariant support vector machines. Machine Learning, 46, 161–190.

Dellaportas, P., and Tsionas, M. G. (2019). Importance sampling from posterior distributions using copula-like approximations. Journal of Econometrics, 210, 45–57.

Desai, V.S. Crook, J.N. and Overstreet, G.A. (1996). A comparison of neural networks and linear scoring models in the credit union environment. European Journal of Operational Research, 95 (1), 24–37.

Diewert, W. E. (1971). An application of the Shephard Duality Theorem: A generalized leontief production function. Journal of Political Economy, 79 (3), 481–507.

Diewert, E., and Wales, J. T. (1987). Flexible functional forms and global curvature conditions. Econometrica, 55, 43–68.

Dong, Y., Firth, M., Hou, W., and Yang, W. (2016). Evaluating the performance of Chinese commercial banks: A comparative analysis of different types of banks. European Journal of Operational Research, 252 (1), 280–295.

Fan Y., and Patton A.J. (2014). Copulas in econometrics. Annual Review of Economics, 179–200.

Faugeras, O. P. (2013). Sklar's theorem derived using probabilistic continuation and two consistency results. Journal of Multivariate Analysis, 122, 271 – 277.

Feil B, Kucherenko S, and Shah N. (2009). Comparison of Monte Carlo and Quasi Monte Carlo sampling methods in high dimensional model representation. Wilmott Journal, 1 (4), 179–195.

Feng G, and Serletis, A. (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.

Fuss, M., and McFadden, D. (1978). Production Economics: A Dual Approach to Theory and Application,Vol.1. Amsterdam, Netherlands: North-Holland.

Galán, J. E. , Veiga, H. , and Wiper, M. P. (2015). Dynamic effects in inefficiency: evidence from the Colombian banking sector. European Journal of Operational Research, 240 (2), 562–571.

Gallant, A. (1981). On the bias in flexible functional forms and an essentially unbiased form: The Fourier flexible form. Journal of Econometrics 15 (2), 211–245.

Genest, C., J. and MacKay. (1986). The joy of copulas: Bivariate distributions with uniform marginals. American Statistician, 40 (4), 280–283.

Genius, M., Stefanou, S.E. and Tzouvelekas, V. (2012). Measuring productivity growth under factor non-substitution: An application to US steam-electric power generation utilities. European Journal of Operational Research, 220 (3), 844 –852.

Girolami, M., and 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 .

Hall, R.E. (1973). The specification of technology with several kinds of output. Journal of Political Economy, 81 (4), 878–892.

Hinton, G.E., and Salakhutdinov, R.R. (2006). Reducing the dimensionality of data with neural networks. Science 313, 504 – 507.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. Neural Networks 2 (5), 359–366.

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

Hughes J.P, and Mester L. J. (2013). Who said large banks don't experience scale economies? Evidence from arisk–return-driven cost function. Journal of Financial Intermediation, 22 (4), 559–585.

Hong, Y., J. Tu, and Zhou. G. (2007). Asymmetries in stock returns: Statistical tests and economic evaluation. Review of Financial Studies, 20 (5), 1547–1581.

Joe, H., (1997). Multivariate models and dependence concepts. Chapman & Hall/CRC.

Kakouris, I., and Rustem., B. (2014). Robust portfolio optimization with copulas. European Journal of Operational Research, 235, 28–37.

Karlis, D. and Meligkotsidou L. (2006). Finite multivariate Poisson mixtures with applications.

Journal of Statistical Planning Inference, 137, 1942 – 1960.

Koutsomanoli-Filippaki, A., and Mamatzakis, E. (2009). Performance and Merton-type default risk of listed banks in the EU: a panel VAR approach. Journal of Banking and Finance, 33 (11), 2050 – 2061.

Kristjanpoller, W., and Minutolo, M.C. (2018). A hybrid volatility forecasting framework integrating garch, artificial neural network, technical analysis and principal components analysis. Expert Systems with Applications, 109, 1–11.

Kumbhakar, S. C. (2013). Specification and estimation of multiple output technologies. European Journal of Operational Research, 231 (2), 465–473.

Kumbhakar, S. C. (2011). Estimation of production technology when the objective is to maximize return to the outlay. European Journal of Operational Research, 208 (2), 170–176.

Kumbhakar, S., C. and Tsionas E.G. (2016). The good, the bad and the technology: Endogeneity in environmental production models. Journal of Econometrics, 190 (2), 315– 327.

Li, T. and Rosenman, R. (2001). Estimating hospital costs with a generalized Leontief function. Health Economics, 10, 523 – 538.

Li, G., Wang, S.-W., Rabitz, H., Wang, S. and Jaffe, P. (2002). Global uncertainty assessments by high dimensional model representations (HDMR). Chemical Engineering Science, 57, 4445 – 4460.

Lovell, C. A. K. (1995). Econometric efficiency analysis: A policy-oriented review. European Journal of Operational Research, 80 (3), 452 – 461.

Lozano, Vivas, A. (1997). Profit efficiency for Spanish savings banks. European Journal of Operational Research, 98 (2), 381 – 394.

Maggi, B., and Guida, M. (2011). Modelling non-performing loans probability in the commercial banking system: efficiency and effectiveness related to credit risk in Italy. Empirical Economics 41 (2), 269–291.

Malikov, E., Kumbhakar, S. C., and Tsionas, E. G. (2016). 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, 31 (7), 1407–1429.

Mester, L. J. (1997). Measuring efficiency at US banks: Accounting for heterogeneity matters. Journal of Operational Research, 98 (2), 230–242.

Michaelides, P. G., Vouldis, A. T., and Tsionas, E. G. (2010). 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., and 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.

Nelsen, R.B. (2006). An introduction to copulas. Second Edition, Springer, U.S.A.

Olesen, O. B., and Petersen, N. C. (2016). Stochastic data envelopment analysis– A review. European Journal of Operational Research, 251 (1), 2–21.

Patton, A., J. (2012). A review of copula models for economic time series. Journal of Multivariate Analysis, 110, 4–18.

Poon, S.-H., Rockinger, M. and Tawn. J. (2004). Extreme value dependence in financial markets: Diagnostics, models, and financial implications. Review of Financial Studies, 17 (2), 581– 610.

Ray, S., and Das, A. (2010). Distribution of cost and profit efficiency: Evidence from Indian banking. European Journal of Operational Research, 201 (1), 297–307.

Rosenblatt, F. (1962). Principles of neurodynamics. New York: Spartan.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1986). Learning internal representations by error propagation. In: Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1: Foundations, 318–362. MIT Press, Cambridge.

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

Sklar, A. (1959). Fonctions de repartition à n dimensions et leurs marges. Publications del' Institut de statistique del' Université de Paris, 8, 229–231.

Sun, K., Kumbhakar, S. C., and Tveterås, R. (2015). Productivity and efficiency estimation: A semiparametric stochastic cost frontier approach. European Journal of Operational Research, 245 (1), 194–202.

Tecles, P.L., and Tabak, B.M. (2010). Determinants of bank efficiency: the case of Brazil. European Journal of Operational Research, 207 (3), 1587–1598.

Torri, G., Giacometti, R. and Paterlini S. (2018). Robust and sparse banking network estimation. European Journal of Operational Research, 270 (1), 51– 65.

Tsionas E.G. (2003). Combining DEA and stochastic frontier models: an empirical Bayes approach. European Journal of Operational Research, 147 (3), 499–510.

Tsionas, M.G. and Izzeldin, M. (2018). A novel model of costly technical efficiency, European Journal of Operational Research, 268 (2), 653 – 664.

Tsionas, E.G., and Mallick, S.K. (2018). A Bayesian semiparametric approach to stochastic frontiers and productivity. European Journal of Operational Research, 274 (1), 391–402.

Tzeremes, N G. (2015). Efficiency dynamics in Indian banking: A conditional directional distance approach. European Journal of Operational Research, 240 (3), 807–818.

Wanke, P., Barros, C.P., and Faria, J.R. (2015). Financial distress drivers in Brazilian banks: a dynamic slacks approach. European Journal of Operational Research, 240 (1), 258–268.

Wanke, P., Barros, C.P., and Emrouznejad, A. (2016). Assessing productive efficiency of banks using integrated Fuzzy-DEA and bootstrapping: A case of Mozambican banks. European Journal of Operational Research, 249 (1), 378–389.

Zhang, G.P. and Berardi. V.L. (2001). Time series forecasting with neural network ensembles: An application for exchange rate prediction, Journal of the Operational Research Society, 52, 652 – 664.