BIROn - Birkbeck Institutional Research Online

Tran, K.C. and Mamatzakis, Emmanuel C. and Tsionas, M.G. (2018) Why fully efficient banks matter? A nonparametric stochastic frontier approach in the presence of fully efficient banks. Empirical Economics , ISSN 0377-7332. (In Press)

Why Fully Efficient Banks Matter?
A Nonparametric Stochastic Frontier Approach in the Presence of Fully Efficient Banks

Kien C. Tran1,
Mike G. Tsionas2 &
Emmanuel Mamatzakis3

Abstract

A common assumption in the banking stochastic performance literature refers to the non-existence of fully efficient banks. This paper relaxes this strong assumption and proposes an alternative semiparametric zero-inefficient stochastic frontier models. Specifically, we consider a nonparametric specification of the frontier whilst maintaining the parametric specification of the probability of fully efficient banks. We suggest an iterative local maximum likelihood procedure that achieves the optimal convergence rates of both nonparametric frontier and the parameters contain in the probability of fully efficient banks. In an empirical application, and given the implications of not counting for fully efficient banks, we apply the proposed model and estimation procedure to a comprehensive global banking data set so as to derive new corrected measures of global bank performance and bank productivity growth. Overall, the results indicate that over time, in particular after the financial crisis, a shift of densities to lower efficiency scores and productivity growth takes primarily place in advanced economies and EU. The results also show that there is variability across regions, and the probability of fully efficient banks is mostly affected by bank specific variables that are related to bank's risk-taking, whereas country specific variables, such as inflation, also have an effect.

Keywords*:* Backfitting local maximum likelihood, Mixture models, Probability of fully efficient firms, Global banking system.

JEL: C13, C14, G20, G21.

Department of Economics, University of Lethbridge, 4401 University Drive W, Lethbridge, AB, T1K 7L1, Canada

Kien C. Tran

Lancaster University Management School, Lancaster, LA1 4YX, UK

Mike G. Tsionas

University of Sussex Business School, Jubilee Building, Brighton, BN1 9SL, UK

Emmanuel Mamatzakis

1. Introduction

One of the main assumptions in stochastic frontier analysis (e.g., Aigner, Lovell and Schmid, 1977; Meeusen and van den Brock, 1977) is that all firms are inefficient and their inefficiency is modeled through a continuous density. However, when some firms are fully efficient, a fact that cannot preclude on a prior grounds, applying stochastic frontier analysis was shown to have serious implications on the inefficiency estimates. Thus, to account for the possibility of fully efficient firms, Kumbhakar, Parmeter and Tsionas (2013), Rho and Schmidt (2015) propose a special class of two-component mixture model, for which they term "*zero-inefficiency stochastic frontier model*" (ZISF) that allows for the inefficiency term to have certain mass at zero with certain probability p and a continuous distribution with probability $1-p$. They further extend model to allow for the probability of fully efficient firm to depend on a set of covariates via a logit or a probit function. For a recent review of parametric ZISF models, see for example Parmeter and Kumbhkar (2014). Recently, Tran and Tsionas (2016) suggest a semiparametric version of the ZISF model by using nonparametric formulation of the probability function, and propose an iterative backfitting local maximum likelihood procedure to estimate the frontier parameters and the nonparametric function.

In this paper, we first propose an alternative semiparametric ZISF model, which is different from the one suggested by Tran and Tsionas (2016). Specifically, we consider a nonparametric specification of the frontier whilst maintaining the parametric specification (e.g., logit or probit function) of the probability of fully efficient firms. Unlike Tran and Tsionas (2016), by maintain the parametric assumption of the probability of fully efficient firm, there are no need for imposing local restrictions so as to ensure the estimated probability lies in the interval $[0,1]$. To estimate the unknown function of the frontier and the parameters of the probability of fully efficient firm, we modify the iterative backfitting local maximum likelihood procedure developed in Tran and Tsionas (2016), which is fairly simple to compute in practice. We also provide the necessary asymptotic properties of the modified proposed estimator.

Next, we apply the proposed model and estimation procedure to the global banking data set, classified per region across the world following IMF's World Economic Outlook classification to examine the productivity growth and efficiency across global banking system. Our application differentiates and contributes to the banking literature in several ways. Firstly, there are vast literature on bank productivity and efficiency, see for example, Allen and Rai (1996); Mester (1996); and Berger and Mester (2003); DeYoung and Hasan (1998); Feng and Serletis (2010); Fend and Zhang (2014) (for US banks); Berg et al. (1992); Alam (2001); Orea (2002); Cahnoto and Dermine (2003); Barros et al. (2009); Tortosa-Ausina et al. (2008); Delis et al. (2011) (for other countries). However, majority of these application are based on the approach of Data Envelopment Analysis (DEA) analysis. Such analysis provides the basis to estimate the Malmquist productivity index. When it comes to parametric measurement of productivity through stochastic frontier analysis, evidence is scarce (see Kumbhakar et al. (2001); Koutsomanoli et al. (2009); Assaf et al. (2011)). Thus, from the methodological and practical stand points, we provide in this paper, a novel nonparametric stochastic frontier approach to measure both bank efficiency and productivity, allowing for some banks to be fully efficient. Secondly, to the best of our knowledge, this is the first study that presents bank productivity and efficiency at a global level, aiming to examine cross-country variability, whilst controlling for the impact of various control variables, whether bank or country specific. Last but not least, we examine the effect of the credit crunch in 2008 and estimate what control variables affect the probability of having a fully efficient bank prior and ex-post the crisis. This is of the utmost importance in terms of bank performance, particularly over periods of high financial distress that could lead to a shift of the whole frontier.

Overall, the results indicated that over time, in particular after the financial crisis, a shift of densities to lower efficiency scores and productivity growth takes primarily place in advanced and EU banks. The results also show that there is variability across other regions, and the probability of fully efficiency banks is mostly affected by bank specific variables that are related to bank's risk-taking, whereas country specific variables, such as inflation, also have an effect. In addition, our results also indicate that the *'bad management hypothesis'* (Berger and DeYoung, 1997), '*agency cost hypothesis'* (Jensen and Meckling (1979)) and *'quite life hypothesis'* (Koetter and Noth (2013)) are in play with regards to bank efficiency, whereas risk-taking activities appear to dominate changes in efficiency and productivity growth in the eve of the credit crunch.

The rest of the paper is structured as follows. Section 2 develops the model and the estimation procedure. Also, in this section, limited Monte Carlo simulations are performed to examine the finite sample performance of the proposed estimators. Section 3 provides an empirical analysis of Global banking system. Concluding remarks are given in Section 4. The proofs of the Theorems are gathered in Appendix A, whilst extension of the proposed model to a fully localized (or fully nonparametric) is given in Appendix B.

2. The Model and Estimation Procedure

2.1 The Model

Suppose that we have a random sample $\{(Y_i, X_i, Z_i) : i = 1,\ldots,n\}$ from the population (X, Y, Z) where $Y_i \in \mathbb{R}$ is a scalar random variable representing output of firm i, $X_i \in \mathbb{R}^d$ is a vector of continuous regressors representing inputs of firm i, and $Z_i \in \mathbb{R}^r$ is a vector of continuous covariates which may or may not have common elements with X. Let x be a binary latent class variable, and assume that for $c = 0, 1$, x has a conditional discrete distribution $P(x = 0 \mid Z = z) = p(z)$ and $P(x = 1 \mid Z = z) = 1 - p(z)$. A nonparametric version of the zero-inefficiency stochastic frontier (NP-ZISF) model proposed by Kumbhakar et al. (2013) can be written as

$$Y_i = \begin{cases} m(X_i) + v_i & \text{with probability } p(Z_i) \\ m(X_i) + v_i - u_i & \text{with probability } 1 - p(Z_i) \end{cases}, \tag{1}$$

where $m(X_i)$ is the frontier function, $v_i \mid X_i = x \sim N(0, \sigma_v^2(x))$ and $u_i \mid X_i = x \sim \left|N(0, \sigma_u^2(x))\right|$. Conditioning on $X_i = x$, the functions $m(x)$, $\sigma_v^2(x)$ and $\sigma_u^2(x)$ are unknown but assumed to be smooth. Note that model (1) is special case of a two-component mixture model as well as latent class stochastic frontier models (e.g., Greene, 2005) with the (technology) function $m(x)$ being restricted to be the same for both regimes, and the composed error is $v_i - u_i(1 - I\{u_i = 0\})$ where $I\{A\}$ is an indicator function such that $I(A) = 1$ if A holds, and zero otherwise. Model (1) also contains several interesting features. First, when $p(z) = 1$, model (1) reduces to a nonparametric regression model. Second, when $p(z) = 0$, it becomes a nonparametric stochastic frontier model (e.g., Fan, Li and Weesink, (1996) and Kumbhakar et al. (2007)). Third, when $m(x)$ is linear in x and $\sigma_u^2(.) = \sigma_u^2$ and $\sigma_v^2(.) = \sigma_v^2$, it becomes a semiparametric ZISF model of Tran and Tsionas (2016). Finally, when $m(x)$ is linear in x and $p(.) = p$, $\sigma_u^2(.) = \sigma_u^2$ and $\sigma_v^2(.) = \sigma_v^2$, model (1) reduces to the parametric ZISF model of Kumbhakar et al. (2013). Consequently, model (1) can be viewed as a generalization of semi-parametric partially linear stochastic frontier regression models as well as the ZISF models. Thus, model (1) provides a general framework for ZISF models.

2.2 Identification

We now turn our attention to the model identification. Under the standard stochastic frontier framework regardless of parametric or nonparametric specification of the frontier, the parameter σ_u^2, the variance of u_i is identified through the moment restrictions on the composed errors $\varepsilon_i = v_i - u_i$, when u_i is left unspecified. However, when the inefficiency term, u_i is modelled in a flexible manner along with parametric specification the frontier, there are possible identification problems between the intercept and the inefficiency term. For more discussion on this identification issues, see for example, Griffin and Steel (2004). In the context of model (1), we have an additional parameter $p(.)$, which can be identified only if there are non-zero observations in each class. As Kumbhakar et al. (2013) and Rho and Schmidt (2015) point out, when $\sigma_u^2 \to 0$, $p(.)$ is not identified since the two classes become indistinguishable. Conversely, when $p(.) \to 1$ for a given z, σ_u^2 is not identified. In fact, when a data set contains little inefficiency, one might expect that σ_u^2 and $p(.)$ to be imprecisely estimated, since it is difficult to identify whether little inefficiency is due to $p(.)$ is close to 1 or σ_u^2 is close to zero. For the present discussion, we will assume that $\sigma_u^2 > 0$, and $0 < p(.) < 1$ so that all the parameters in model (1) are identified.

To complete the specification of the model, first given $Z = z$, we assume that, $p(z)$ takes a form of logistic function:

$$p(z) = \frac{\exp(z'\alpha)}{1 + \exp(z'\alpha)}, \tag{2}$$

so as to ensure that $0 < p(z) < 1$. Note that, in our setting, one could model $p(z)$ non-parametrically which makes model (1) fully nonparametric. However, as noted by Martins-Filho and Yao (2015), the main drawback of this approach is that, since all the parameters are localized, the rate of convergence of their estimator becomes slow when the number of conditioning variables get large (which frequently encounter in practice) implying that the accuracy of the asymptotic approximation can be poor (i.e., the curse of dimensionality problem). Appendix B provides a brief discussion as how to estimate model (1) when all the model's parameters are fully localized.

Next, let $f(Y, \theta(x))$ denote the conditional density of Y given $X = x$, $Z = z$ where $\theta(x) = (\alpha', \gamma(x)')'$ and $\gamma(x) = (m(x), \sigma^2(x), \lambda(x))'$. Given the distributional assumptions of v and u, the conditional pdf of Y given $X = x$ and $Z = z$ is

$$\begin{aligned} f(Y \mid \theta(x)) = {} & \left(\frac{p(z)}{\sigma_v(x)}\right)\phi\left(\frac{Y - m(x)}{\sigma_v(x)}\right) + \\ & (1 - p(z))\left[\frac{2}{\sigma(x)}\phi\left(\frac{Y - m(x)}{\sigma(x)}\right)\Phi\left(-(Y - m(x))\frac{\lambda(x)}{\sigma_v(x)}\right)\right] \end{aligned} \tag{3}$$

where $p(z)$ is defined in (2), $\sigma^2(x) = \sigma_u^2(x) + \sigma_v^2(x)$, $\lambda(x) = \sigma_u(x)/\sigma_v(x)$, $\phi(.)$ and $\Phi(.)$ are the probability density (pdf) and cumulative distribution functions (CDF) of a standard normal variable, respectively. It follows that the conditional log-likelihood is then given by

$$L_{1n}(\alpha, \gamma(x)) = \sum_{i=1}^{n} \log f(Y_i \mid \alpha, \gamma(x)). \tag{4}$$

2.3 Estimation: Backfitting Local Maximum Likelihood

From (4), we note that the vector $\theta(x)$ contains both finite dimensional as well as nonparametric functions which makes the direct maximization of (4) over $\theta(x)$ in an infinite-dimensional function space intractable and generally suffers from overfitting problem. To make (4) tractable in practice, we will employ local linear regression for model (1), albeit one could consider higher orders of local polynomials. However, general order of local polynomial fitting requires additional notational complexity, but the approach is the same. In local linear fitting, we first approximate $g(x)$ by taking the first-order Taylor series expansion of $g(x)$ at a given set point x_0. That is, for a given x_0 and x in the neighborhood of x_0,

$$g(x) \approx g_0(x_0) + \Gamma_1(x_0)(x - x_0), \tag{5}$$

where $g_0(x_0)$ is a (3×1) vector and $\Gamma_1(x_0)$ is a $(3 \times d)$ matrix of the first-order derivatives. Next, we define the kernel function:

$$K_H(X_i, x_0) = |H|^{-1} K(H^{-1}(X_i - x_0)), \tag{6}$$

where H is a bandwidth matrix which we assume to be positive definite and symmetric. In practice, in (6) we could use a multivariate product kernel,

$$K(u) = \prod_{j=1}^{d} k(u_j),$$

where $k(.)$ is a symmetric univariate probability density function. Then the corresponding conditional *local* log-likelihood function for data $\{(Y_i, X_i, Z_i) : i = 1, \ldots, n\}$ can be written as

$$L_{2n}(\alpha, g_0(x_0), \Gamma_1(x_0)) = \sum_{i=1}^{n} \{\log f(Y_i; \alpha, g_0(x_0) + \Gamma_1(x_0)(X_i - x_0))\} K_H(X_i - x_0). \tag{7}$$

Thus, the conditional local log-likelihood depends on x. Note that however, the global parameter α does not depend on x, and by maximizing (7), α will be estimated *locally* and hence it does not possess the usual parametric $\sqrt{n}$-consistency. To preserve the $\sqrt{n}$-consistency estimator of

α, we use a backfitting approach similar to Tran and Tsionas (2016), which motivated by Huang and Yao (2012). To do this, let $\tilde{g}(x_0) = \{\tilde{m}(x_0), \tilde{\sigma}^2(x_0), \tilde{\lambda}(x_0)\}$ and $\tilde{\alpha}(x_0)$ be the maximizer of the local log-likelihood function (7), then the initial local linear estimators of $g(x)$ and $\alpha(x)$ are given by $\tilde{g}(x_0) = \tilde{g}_0(x_0)$ and $\tilde{\alpha} = \tilde{\alpha}(x_0)$. Now given the initial estimator $\tilde{g}(x_0)$, the parameter vector α can be estimated *globally* by maximizing the following *global* log-likelihood function where we replace $g(x)$ with its initial estimate $\tilde{g}(x_0)$ in (4):

$$L_{3n}(\alpha, \tilde{g}(x_i)) = \sum_{i=1}^{n} \log f(Y_i \mid \alpha, \tilde{g}(x_i)), \tag{8}$$

Let $\hat{\alpha}$ be the solution of maximizing (8). In section 3 below, we will show that, under certain regularity conditions $\hat{\alpha}$ will retain its $\sqrt{n}$-consistency property. Given the estimates of $\hat{\alpha}$, $g(x)$ can be estimated by maximizing the following conditional *local* log-likelihood function:

$$L_{4n}(\hat{\alpha}, g_0(x_0), \Gamma_1(x_0)) = \sum_{i=1}^{n} \{\log f(Y_i; \hat{\alpha}, g_0(x_0) + \Gamma_1(x_0)(X_i - x_0))\} K_H(X_i - x_0), \tag{9}$$

Let $\hat{g}_0(x_0)$ and $\hat{\Gamma}_1(x_0)$ be the maximizer of (9), then the local linear estimator of $g(x)$ is given by $\hat{g}(x) = \hat{g}_0(x)$. Finally, $\hat{\alpha}$ and $\hat{g}(x)$ can be further be improved by iterating until convergence. As noted by Tran and Tsionas (2106), convergence is typically fast and requires only two or three iterations. The final estimates of $\hat{\alpha}$ and $\hat{g}(x)$ will be denoted as backfitting local maximum likelihood (BLML). The final estimate of $p(z)$ can be obtained via $\hat{p}(z) = \frac{\exp(z'\hat{\alpha})}{1 + \exp(z'\hat{\alpha})}$.

We summarize the above estimation procedure with the following computational algorithm:

Step 1: For each z_i, $i = 1, \ldots, n$, in the sample, maximize the conditional *local* log-likelihood (7) to obtain the estimate of $\tilde{g}(x_i)$. Note that if the sample size n is large the maximisation could be performed on a random subsample N_s, where $N_s << n$ so as to reduce the computational burden.

Step 2: From step 1, conditional on $\tilde{g}(x_i)$, maximize the *global* log-likelihood function (8) to obtain $\hat{a}$.

Step 3: Conditional on $\hat{a}$ from step 2, maximize the conditional local log-likelihood function (9) to obtain $\hat{g}(x_i)$.

Step 4: Using $\hat{g}(x_i)$ repeat step 2 and then step 3 until the estimate of $\hat{a}$ converges.

Note that to implement the estimation algorithm described above, specifications of the kernel function $K(.)$ as well as bandwidth matrix H are required. For the kernel function, we use a product of univariate kernel where Epanechnikov or Gaussian function is a popular choice for each kernel. As for the bandwidth selection, we follow Kumbhakar et al. (2007) and use a d-dimensional vector of bandwidth $h = (h_1, \ldots, h_d)'$ such that $h = h_b s_X n^{-1/5}$ where h_b is a scalar, and $s_X = (s_{X_1}, \ldots, s_{X_d})'$ is the vector of empirical standard deviations of the d components of X. This choice of bandwidth vector is adjusted for different scales of the regressors and different sample sizes. Then data driven methods such as cross-validation (CV) can be used (see for example Li and Racine (2007)) to evaluate a grid of values for h_b. In our context, we use a likelihood version of CV, which is given by

$$CV(h_b) = \frac{1}{n} \sum_{i=1}^{n} \log f(Y_i; \hat{a}^{(i)}, \hat{g}^{(i)}(x_i) \mid x, z), \quad (10)$$

where $\hat{a}^{(i)}$ and $\hat{g}^{(i)}(x_i)$ are the leave-one-out version of the backfitting local MLE described above. However, it is important to note that, in semiparametric modeling, undersmoothing conditions (see Theorem 1 below) are typically required in order to obtain $\sqrt{n}$ - consistency for the global parameters. The optimal bandwidth $\hat{h} = \hat{h}_b s_X n^{-1/5}$ selected by CV will be in the order of $n^{-1/5}$ which does not satisfy the required undersmoothing conditions. However, a reasonable adjusted bandwidth, which suggested by Li and Liang (2008) that satisfies the undersmoothing condition, can be used, and it is given by $\tilde{h} = \hat{h} \times n^{-2/15} = O(n^{-1/3})$. We will apply this adjusted bandwidth in our simulations and empirical application below. Finally, note that our proposed approach above can be easily modified and extended to accommodate other models as well that allow for the distribution of u_i to depend on a set of covariate either parametrically or nonparametrically without affecting the estimation algorithm.

2.4 Estimation of Bank -Specific Inefficiency:

Follow the discussion of Kumbhakar et al. (2013), we can similarly consider several approaches to estimate firm-specific inefficiency. The first approach is based on the popular estimator of Jondrow et al. (1982) where under our setting, the conditional density of u given $e(x)$ is

$$f(u \mid e(x)) = \begin{cases} 0 & \text{with probability } p(z) \\ N_+(m_*(x), s_*^2(x)) & \text{with probability } (1 - p(z)) \end{cases},$$

Where $N_+(.)$ denotes the truncated normal, $m_*(x) = -e(x)s_u^2(x) / s^2(x)$ and $s_*^2(x) = s_u^2(x)s_v^2(x) / s^2(x)$. Thus, the conditional mean of u given $e(x) = Y - m(x)$ is:

$$E(u \mid e(x)) = (1 - p(z)) \frac{s(x)l(x)}{1 + l^2(x)} \left[\frac{f(l(x)e(x) / s(x))}{F(-l(x)e(x) / s(x))} - \frac{l(x)e(x)}{s(x)} \right]. \quad (11)$$

A point estimator of individual inefficiency score could be obtained by replacing the unknown parameters in (7) by their estimates discussed above, and $e(x)$ by $\hat{e}(x) = Y - \hat{m}(x)$.

Another approach is to construct the posterior estimates of inefficiency $\tilde{u}_i$. To do this, let p_i^* denotes the "posterior" estimate of the probability of being fully efficient where

$$p_i^* = \frac{(\hat{p}(z) / \hat{s}_v(x)) f(\hat{e}_i(x) / \hat{s}_v(x))}{(\hat{p}(z) / \hat{s}_v(x))(f(\hat{e}_i(x) / \hat{s}_v(x)) + (1 - \hat{p}(z))(2 / \hat{s}(x)) f(\hat{e}_i(x) / \hat{s}(x)) F(-\hat{e}_i(x) / \hat{s}(x))} \quad (12)$$

Then the "posterior" estimate of inefficiency can be defined as $\tilde{u}_i = (1 - p_i^*)\hat{u}_i$ where $\hat{u}_i$ is the estimate of inefficiency based on (11).

2.5 Asymptotic Theory

In this section, we derive the sampling property of the proposed backfitting local MLE $\hat{\alpha}$ and $\hat{\gamma}(x) = (\hat{\beta}'(x), \hat{\sigma}^2(x), \hat{\lambda}(x))'$. In particular, we will show that the backfitting estimator $\hat{\alpha}$ is $\sqrt{n}$-consistent and follows an asymptotic normal distribution. In addition, we also provide the asymptotic bias and variance of the estimator $\hat{\gamma}(x)$, and show that asymptotically, it has smaller variance compared to $\tilde{\gamma}(x)$. To this end, let us define the following additional notations.

Let $\theta(x) = (\alpha', \gamma(z))'$ and $\ell(\theta(x), z, y) = \log f(y \mid \theta(x), z)$. Define $q_\theta(\theta(x), z, y) = \dfrac{\partial \ell(\theta(x), z, y)}{\partial \theta}$,

$q_{\theta\theta}(\theta(x), z, y) = \dfrac{\partial^2 \ell(\theta(x), z, y)}{\partial \theta \partial \theta'}$ and the terms q_α, q_γ, $q_{\alpha\alpha}$, $q_{\alpha\gamma}$ and $q_{\gamma\gamma}$ can be defined similarly. In addition, let $\Psi(w \mid x) = E[q_\gamma(\theta(x), z, y) \mid x = w]$,

$$I_{\theta\theta}(x) = -E[q_{\theta\theta}(\theta(x), z, y) \mid x] = \begin{bmatrix} I_{\alpha\alpha}(x) & I_{\alpha\gamma}(x) \\ I_{\alpha\gamma}(x) & I_{\gamma\gamma}(x) \end{bmatrix}$$

where

$$I_{\alpha\alpha}(x) = -E[q_{\alpha\alpha}(\theta(x), z, y) \mid x]$$

$$I_{\gamma\gamma}(x) = -E[q_{\gamma\gamma}(\theta(x), z, y) \mid x]$$

$$I_{\alpha\gamma}(x) = -E[q_{\alpha\gamma}(\theta(x), z, y) \mid x]$$

Finally, let $\mu_j = (\int u^j K(u) du) I_d$, $\kappa_j = (\int u^j K^2(u) du) I_d$ and $|H| = h_1 h_2 \ldots h_d$. We make the following assumptions:

Assumption 1: The sample $\{(X_i, Y_i, Z_i),\ i = 1,\ldots,n\}$ is independently and identically distributed from the joint density $f(x,y,z)$, which has continuous first derivative and positive in is support. The support for X, denoted by χ, is a compact subset of $\mathbb{R}^d$ and $f(X) > 0$ for all $X \in \chi$.

Assumption 2: The unknown functions $g(x) = (m(x), \sigma^2(x), \lambda(x))'$ are twice partially continuously differentiable in its argument.

Assumption 3: The matrixes $I_{\theta\theta}(x)$ and $I_{\alpha\alpha}$ are positive definite.

Assumption 4: The kernel density function $K(.)$ is symmetric, continuous and has bounded support.

Assumption 5: For some $\zeta < 1 - \rho^{-1}$, $n^{2\zeta-1}|H| \to \infty$ and $E(X^{2\rho}) < \infty$.

All the above assumptions are relatively mild and have been used in the mixture models and local likelihood estimation literature. Given the above assumptions, we now ready to state our main results in the following theorems.

Theorem 1: Under Assumptions 1-5 and in addition, $n|H|^4 \to 0$ *and* $n|H|^2 \log(|H|^{-1}) \to \infty$*, we have*

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{D} N(0, A^{-1}\Sigma A^{-1}),$$

where $A = E\{I_{\alpha\alpha}(x)\}$ *and* $\Sigma = Var\left\{\frac{\partial l(\alpha, \theta(x), z, y)}{\partial \alpha} - I_{\alpha\gamma}(x)\delta(x,y,z)\right\}$ *with* $\delta(x,y,z)$ *is the first* $(r \times r)$ *sub-matrix of* $I_{\theta\theta}^{-1}(x)q_\theta(\theta(x), z, y)$.

Theorem 2: Under Assumptions 1-5 and in addition, as $n \to \infty$, $|H| \to 0$*, and* $n|H| \to \infty$ *we have*

$$(n|H|)^{1/2}\{\hat{g}(x) - g(x) - B(x) + O_p(\textstyle\sum_{i=1}^{d} h_i^2)\} \xrightarrow{D} N\{0, \kappa_0 f^{-1}(x) I_{\gamma\gamma}^{-1}\},$$

where $B(x) = \frac{1}{2}\mu_2 |H|^2 I_{\gamma\gamma}^{-1}(z)\Psi''(x \mid x)$.

The proofs of Theorems 1 and 2 are given in Appendix A. The proofs are straightforward extension of the proofs of Theorems 1 and 2 in Tran and Tsionas (2016) to the multivariate case, and therefore we only provide the key steps of the proofs. Note that, the result from Theorem 2 shows that, as for common semiparametric model, the estimate of a has no effect on the first-order asymptotic since the rate of convergence of $\hat{g}(x)$ is slower than that of $\sqrt{n}$. Consequently, it is fairly straightforward to see that $\hat{g}(x)$ is more efficient than the initial estimate of $\tilde{g}(x)$.

2.6 Monte-Carlo Simulations

In this section, we conduct some simulations to study the finite sample performance of the proposed estimator. To this end, we consider the following data generating process (DGP) for the specification of $m(x_i)$, $s_u^2(x)$ and $s_v^2(x)$:

$$m(x_1,x_2) = 1 + x_1 + x_2 + 0.5x_1^2 + 0.5x_2^2 + x_1x_2,$$

$$s_v^2(x_1,x_2) = 0.5\exp(0.2x_1 + 0.5x_2),$$

$$s_u^2(x_1,x_2) = 0.5\exp(0.5x_1 + 0.2x_2),$$

$$p(z) = \exp(0.5z) / [1 + \exp(0.5z)]$$

The covariates $x = (x_1,x_2)$ and z are generated in dependently from uniform distribution on $[0,1]$. The random error term v is generated as $N(0,s_v^2(x))$ and the one-sided error u is generated as $\left|N(0,s_u^2(x)\right|$. For all our simulations, we set $l = (1, 2.5, 5\}$, and let the sample sizes vary over $n = 1000$ and $n = 2000$. For each experimental design, 1000 replications are performed.

For our approach, we use the Gaussian kernel function and the bandwidth is chosen according to $\tilde{h} = \hat{h} \times n^{-1/5}$ where $\hat{h}$ is the optimal bandwidth based on CV approach discussed earlier in Section 2.3. To assess the performance of the estimators of the unknown functions

$m(x_i)$, $\sigma_v^2(x)$, and $\sigma_u^2(x)$, we consider the mean average square errors (MASE) for each experimental design:

$$MASE = \frac{1}{1000}\sum_{r=1}^{1000}\left\{\frac{1}{n}\sum_{j=1}^{n}[\hat{\xi}_r(x_j) - \xi_r(x_j)]^2\right\},$$

where $\hat{\xi}(.) = \hat{m}(.)$, $\hat{\sigma}_v^2(.)$ or $\hat{\sigma}_u^2(.)$, and $\{x_{ji} : j = 1,2;\ i = 1,\ldots,N\}$ are the set of evenly space grid points distributed on the support of $x = (x_1, x_2)$.

To assess the performance of the estimator of the unknown parameter in the probability function, we use the means squared errors (MSE):

$$MSE = \frac{1}{1000}\sum_{r=1}^{1000}(\hat{\alpha}_r - \alpha)^2$$

We use a bootstrap procedure to estimate the standard errors, and construct pointwise confidence intervals for the unknown functions as well as the unknown parameters of the probability function. To do this, for a given x_i and z_i, generate the bootstrap sample, Y_i^* from a given distribution of Y specified in (1) with $\{m(.), \sigma_v^2(.), \sigma_u^2(.), \alpha\}$ are replaced by their estimates. By applying or propose estimation procedure for each of the bootstrap samples, we obtain the standard errors and confidence intervals.

Finally, in addition to the assessment of the above properties, we also examine the average biases, standard deviations and MSEs of technical inefficiency and returns to scale measures. For comparison purposes, we also include these results for the parametric ZISF model of Kumbhakar et al. (2013) in which the frontier is estimated by:

$$m(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

Table 1 displays the simulation results for the estimated MASE of $\hat{\xi}(x_j)$ and the estimated MSE of $\hat{\alpha}$ for various values of λ. From Table 1, first we observe that as the sample size increases, both the estimated MSE for parameter estimates $\hat{\alpha}$, and MASE reduces. Second, we also observe that as the sample size doubles, the estimated MSE of $\hat{\alpha}$ reduces to about half of the original values; this is consistent with the fact that the back-fitting local ML estimator of $\hat{\alpha}$ is $\sqrt{n}$-consistent as predicted by Theorem 1.

Table 1: MASE of $(\hat{m}(x), \hat{s}_v^2(x), \hat{s}_u^2(x))$ and MSE of $\hat{a}$

	$n = 1000$ $l = 1$	$n = 2000$ $l = 1$	$n = 1000$ $l = 2.5$	$n = 2000$ $l = 2.5$	$n = 1000$ $l = 5$	$n = 2000$ $l = 5$
	MASE	**MASE**	**MASE**	**MASE**	**MASE**	**MASE**
$m(.)$	0.150	0.129	0.135	0.118	0.085	0.071
$s_v^2(.)$	0.140	0.114	0.144	0.100	0.093	0.074
$s_u^2(.)$	0.129	0.097	0.155	0.104	0.044	0.035
	MSE	**MSE**	**MSE**	**MSE**	**MSE**	**MSE**
a	0.009	0.008	0.008	0.007	0.005	0.003

Note: Authors' estimations. Mean square error (MSE), mean average square errors (MASE).

We next examine the accuracy of the standard error estimation via a bootstrap approach. Table 2 summarizes the performance of the bootstrap approach for standard errors of the estimated functions $(\hat{m}(x), \hat{s}_v^2(x), \hat{s}_u^2(x))$ evaluated at $x = \{0.1, 0.2, ..., 0.9\}$, for two different samples and two different bandwidths which correspond to under-smoothing $\tilde{h} = \hat{h} \times n^{-2/15}$ and appropriate amount $\hat{h}$. In the table, the standard deviation of 1000 estimates are denoted by STD which can be viewed as the true standard errors, whilst the average bootstrap standard errors are denoted SE along with their standard deviations are given the parentheses. The SEs are calculated as the average of 1000 estimated standard errors. The coverage probabilities for all the parameters are given the last column and they are obtained based on the estimated standard errors. The results from the table 3 show that the suggested bootstrap procedure approximates the true standard deviations quite well and the coverage probabilities are close to the nominal levels for almost all cases.

Table 2: Bootstrap Standard Errors, Standard Deviations and Coverage Probabilities.

Parameter	STD	SE(STD)	95% Coverage
$n = 1000, h = 0.08$			

a	0.024	0.026(0.005)	94.8
$n = 1000,\ h = 0.16$			
a	0.029	0.028(0.006)	93.9
$n = 2000,\ h = 0.07$			
a	0.016	0.017(0.003)	94.9
$n = 2000,\ h = 0.14$			
a	0.018	0.019(0.004)	94.5

Note: Estimations based on 1000 estimated standard errors using bootstrap. STD = standard deviations of estimated parameters; SE = estimated standard errors using bootstrap procedure.

Note that the bootstrap procedure also allows us to compute the point-wise coverage probabilities for the probability functions. Table 3 provides the 95% coverage probabilities of $m(.)$, $s_v^2(.)$ and $s_u^2(.)$ for a set of evenly space grid points distributed on the support of x. In the table, the row labeled with $m_{(\hat{a})}(x)$, $s^2_{v(\hat{a})}(x)$ and $s^2_{u(\hat{a})}(x)$ gives the results using the proposed approach, whilst $m_{(a)}(x)$, $s^2_{v(a)}(x)$ and $s^2_{u(a)}(x)$ gives the results assuming a were known. For most cases, the coverage probabilities are close to the nominal level. However, the coverage levels are slightly low for points 0.1, 0.2 and 0.3 when right amount smoothing is used. This is consistent with the expectation that under-smoothing is required.

Table 3: The Pointwise Coverage Probabilities for $\{m(x), s_v^2(x), s_u^2(x)\}$

x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
				$n = 1000,\ h = 0.16$					
$m_{(\hat{a})}(x)$	0.90	0.92	0.92	0.94	0.95	0.96	0.95	0.94	0.94
$m_{(a)}(x)$	0.92	0.93	0.93	0.95	0.95	0.96	0.95	0.95	0.94
$s^2_{v(\hat{a})}(x)$	0.89	0.89	0.91	0.92	0.94	0.95	0.95	0.95	0.93
$s^2_{v(a)}(x)$	0.91	0.92	0.92	0.95	0.95	0.95	0.97	0.95	0.96
$s^2_{u(\hat{a})}(x)$	0.84	0.88	0.90	0.91	0.94	0.95	0.95	0.95	0.95
$s^2_{u(a)}(x)$	0.89	0.91	0.93	0.95	0.95	0.95	0.95	0.94	0.92
				$n = 2000,\ h = 0.08$					
$m_{(\hat{a})}(x)$	0.91	0.93	0.94	0.94	0.94	0.94	0.95	0.95	0.95
$m_{(a)}(x)$	0.93	0.94	0.95	0.95	0.95	0.96	0.96	0.95	0.95
$s^2_{v(\hat{a})}(x)$	0.90	0.92	0.92	0.92	0.93	0.94	0.95	0.95	0.94
$s^2_{v(a)}(x)$	0.92	0.94	0.93	0.95	0.95	0.95	0.95	0.95	0.95
$s^2_{u(\hat{a})}(x)$	0.93	0.93	0.93	0.94	0.94	0.93	0.93	0.94	0.94
$s^2_{u(a)}(x)$	0.93	0.94	0.95	0.95	0.95	0.95	0.96	0.95	0.95

Note: $m_{(\hat{a})}(x)$: when $\hat{a}$ is estimated and $m_{(a)}(x)$: when a is assumed to be known. $s^2_{v(.)}(x)$ and $s^2_{u(.)}(x)$ are defined similarly.

3. Global Banking Analysis

3.1. The data set

To empirically test our proposed methodology, we opt for a global bank sample, so as to provide comprehensive bank efficiency measures in presence of variability across economies. After removing errors and inconsistencies, we end up with an unbalanced dataset that includes 17399 observations for 31 advanced countries, 7130 observations for 35 emerging countries, and 2471 observations for 40 developing countries. The classification of country-groups is based on IMF World Economic Outlook April 2015. All the bank-specific financial variables are obtained from Bankscope database, and represent values in thousand euros.[1] Data for country-level variables are collected from the World Bank Indicators database. This sample represents the majority of financial institutions at a global scale.

Moreover, we opt for a bank cost function as:

$$TC_{it} = \begin{cases} f(P_{it}, Y_{it}, N_{it}) + v_{it} & \text{with probablity } p(Z_{it}) \\ f(P_{it}, Y_{it}, N_{it}) + v_{it} + u_{it} & \text{with probability } 1 - p(Z_{it}) \end{cases}, \qquad (13)$$

where TC_{it} the total cost for firm (bank) i at year t, P_{it} is a vector of input prices, Y_{it} is a vector of outputs of the bank, N_{it} a vector of fixed netputs while Z_{it} is a vector of country-specific environmental variables, v_{it} represents random errors that are assumed to be i.i.d. and have $N(0, s_v^2)$ while u_{it} represents non-negative inefficiency effects that are assumed to be independently but not identically distributed.

Inputs, input prices and outputs are chosen using the intermediation approach and follow Koutsomanoli-Filippaki and Mamatzakis (2009) and Tanna et al. (2011). Based on intermediation banks transform deposits to loans and securities or other earning assets. To this end, the cost function includes two outputs: net loans and other earning assets. The inputs are financial capital (deposits and short-term funding), labour (personnel expenses) and physical capital (fixed assets). The price of financial capital is the interest expenses on deposits divided by total interest bearing borrowed funds, the price of labour is the ratio of personnel expenses over total assets, while the price of physical capital is the ratio of overhead expenses (excluding personnel expenses) to fixed assets. Total bank cost is then calculated as the sum of overheads, such as personnel and administrative expenses, interest, fee and commission expenses. Furthermore, we include equity as a quasi-fixed net put (Berger and Mester, 2003, Koutsomanoli-Filippaki and Mamatzakis, 2009). This is so because we would like to capture the impact from an alternative source of funding on the bank cost structure. If such impacts are ignored then this might cause bias in measuring efficiency, in particular for banks with high equity capital. More equity capital would take into

[1] We exclude banks for which: (i) we had less than three observations over time; (ii) we had no information on the country-level control variables; (iii) we had no information of nonperforming loans.

account that bank management leans towards risk aversion. In addition, we include nonperforming loans (*NPL*) as a negative quasi-fixed input (Hughes and Mester, 2010) to consider risk-taking activities. Note that we also include bank fixed assets to account for physical capital (Berger and Mester 2003).

The summary statistics of these variables are provided in Table 4 for each country-group according to the classification of IMF World Economic Outlook. Interestingly, we notice that the average amount of nonperforming loans in advanced economies' banking industries is almost twice that in emerging economies and eight times that in developing economies.

Table 4. Descriptive statistics of bank variables.

	Advanced economies	**Emerging economies**	**Developing economies**
Variables	**Mean**	**Mean**	**Mean**
Bank outputs and input prices			
Total assets	17951329	9659393	1255046
Total costs	644465	441730	82203
Net loans	9036183	5152601	626854
Other earning assets	7354650	3796769	449156
Price of fund	2.4624	8.9291	5.7072
Price of physical capital	201.4653	415.9367	140.6929
Price of labour	1.1191	2.5210	2.1607
Nonperforming loans	336033	179375	48030
Equity	1032482	754589	133758
Banks specific and control variables			
Z-score	0.6965	0.8081	0.8443
Capital ratio	8.3162	14.7165	13.0949
Fees	0.4040	1.1228	1.2440
Liquidity ratio	15.1238	26.3849	23.4960
Securities	30.0283	40.3673	32.3383
GDP per capita	10.5394	8.3914	7.6975
Inflation	1.2208	10.4568	7.6531
Population density	242.0769	93.6112	226.5159
Market size	28.5054	25.5037	19.4081

Notes: The Table reports the average values of variables used for estimation in each group of economies. Total assets; total costs = total interest expenses + overheads; net loans = gross loans – nonperforming loans; other earning assets; nonperforming loans; equity are reported in thousand USD. Price of fund = total interest expenses/total customer deposits; price of physical capital = other operating expenses/fixed assets; price of labour = personnel expenses/total assets. Z-score= (1+ROE)/ (Standard Deviation of ROE); Size= natural logarithm of total assets; Capital ratio = equity over total assets; Liquidity ratio= liquid assets over total assets; Investment Banking Fees= net fees, commission and trading income over total assets; Securities/TA= total securities over total asset. As country variables we employ: GDP per capita; Inflation;

Population density is the number of people per square kilometer; Market size= value of total shares traded on the stock market exchange. The grouping of advanced, emerging and developing economies follow the IMF, World Economic Outlook, 2014.

In Table 4 and the analysis thereafter we employ some bank specific control variables. Given that during the period of our sample there have been episodes of high risk, we take into bank-specific risk in the estimation of the efficiency scores. To this end, we opt for the z-score as a bank specific measure insolvency risk. This is defined as $z\text{-}score = (1 + ROE) / s_{ROE}$, where ROE is the return on equity and s_{ROE} is the estimate of standard deviation of ROE (as in Koutsomanoli-Filippaki and Mamatzakis, 2009; Delis and Staikouras, 2011, Staikouras et al., 2008). In addition, to take into account of liquidity risk we employ the ratio of liquid assets over total assets.[2] Lastly, we also use the ratio of equity over total assets to take into account capital risk (Koutsomanoli-Filippaki and Mamatzakis, 2009). High capital ratio would imply low capital risk, i.e. equity is a buffer against financial instability. Table 4 includes some descriptive statistics of the three measures of bank risk employed in our analysis. Perhaps not surprisingly, given the financial crisis in 2008, banks in the advanced economies, as z-score at 0.69, face higher risk compared to emerging and developing, 0.8081 and 0.8443 respectively. Descriptive statistics also show that banks in emerging and developing countries are more capitalized and have more liquidity than banks in advanced economies. In addition, we take macroeconomic environment account, opting for GDP per capita and inflation as proxies for the dynamism and the macroeconomic stability of each country. We also include population density and market size to capture size effects of the banking industry.

3.2 Bank efficiency in the presence of fully efficient banks

Table 5 reports bank efficiency for each country-group. There is some variability in efficiency across the world, notably in Middle East and Sub-Sahara Africa. Surprisingly, there is also variability in bank performance as measured by bank efficiency among economies in EU. This is so despite the mandatory convergence process, including in financial markets that economies

[2] Liquid assets are the sum of trading assets, loans and advances with maturity less than three months (Altunbas and Marques (2008)). Liquidity ratio reports bank's liquid assets. If the ratio takes low values would imply high liquidity risk.

have to go through prior to their accession to the EU. Clearly when it comes to bank efficiency we do not observe convergence in EU. However, regarding the economies that form the euro zone, there variability in efficiency is less pronounced, whilst for some economies, i.e. Greece, Slovakia, there is low level of bank efficiency. Economies in Latin America and the Caribbean show a rather low level of efficiency at 0.74, as well as economies in Sub-Sahara Africa at 0.72. Economies in Asia/Pacific and Common Wealth and Independent States report efficiency scores of around 0.78, whereas for the former countries there is some variability.

Table 5: Global bank efficiency in the presence of fully efficiency banks.

Advanced Economies outside Europe					
Australia	0.80	Japan	0.87	Singapore	0.82
Canada	0.84	Korea	0.77	Switzerland	0.92
Hong Kong	0.83	New Zealand	0.76	Taiwan	0.87
Iceland	0.77	Norway	0.82	USA	0.85
Israel	0.85	San Marino	0.76		
		Average 0.82			
EU					
Austria	0.91	Germany	0.84	Poland	0.75
Belgium	0.81	Greece	0.70	Portugal	0.79
Bulgaria	0.78	Hungary	0.81	Romania	0.72
Cyprus	0.80	Ireland	0.83	Slovakia	0.70
Czech	0.80	Italy	0.88	Slovenia	0.74
Denmark	0.88	Lithuania	0.73	Latvia	0.76
Estonia	0.75	Luxembourg	0.70	Sweden	0.83
Finland	0.77	Malta	0.75	Spain	0.82
France	0.82	Netherlands	0.81	UK	0.87
		Average 0.79			
Europe, except EU					
Albania	0.78	Croatia	0.83	Serbia	0.77
Andorra	0.80	FYROM	0.79	Turkey	0.82
Bosnia and Herzegovina	0.77				
Average	***0.79***				
Latin America and the Caribbean					
Argentina	0.78	Colombia	0.79	Jamaica	0.77
Bahamas	0.75	Costa Rica	0.75	Panama	0.70
Bermuda	0.80	Dominican Rep.	0.72	Peru	0.74
Bolivia	0.71	Ecuador	0.72	Trinidad & Tobago	0.71
Brazil	0.75	El Salvador	0.76	Uruguay	0.74
Chile	0.81	Honduras	0.74	Venezuela	0.71
		Average 0.74			
Asia/Pacific					
Bangladesh	0.72	Malaysia	0.84	Taiwan	0.84
Cambodia	0.70	Nepal	0.73	Thailand	0.80
China	0.75	Pakistan	0.80	Vietnam	0.76
India	0.83	Philippines	0.83		
Indonesia	0.84	Sri Lanka	0.80		
		Average 0.78			
Middle East, North Africa					
Bahrain	0.75	Kuwait	0.80	Qatar	0.73
Egypt	0.65	Lebanon	0.81	Saudi Arabia	0.73
Jordan	0.77	Oman	0.76	UAE	0.78
		Average 0.75			
Commonwealth of Independent States					
Armenia	0.75	Georgia	0.77	Russian	0.81
Azerbaijan	0.79	Kazakhstan	0.83	Ukraine	0.84
Belarus	0.82	Moldova Rep.	0.73		
		Average 0.79			
Sub-Saharan Africa					

Angola	0.75	Mauritius	0.77	South Africa	0.85
Benin	0.74	Mozambique	0.77	Swaziland	0.72
Botswana	0.70	Namibia	0.70	Tanzania	0.68
Ethiopia	0.71	Nigeria	0.79	Uganda	0.75
Ghana	0.65	Senegal	0.71	Zambia	0.64
Kenya	0.75	Senegal	0.77	Zambia	0.85
		Average 0.72			

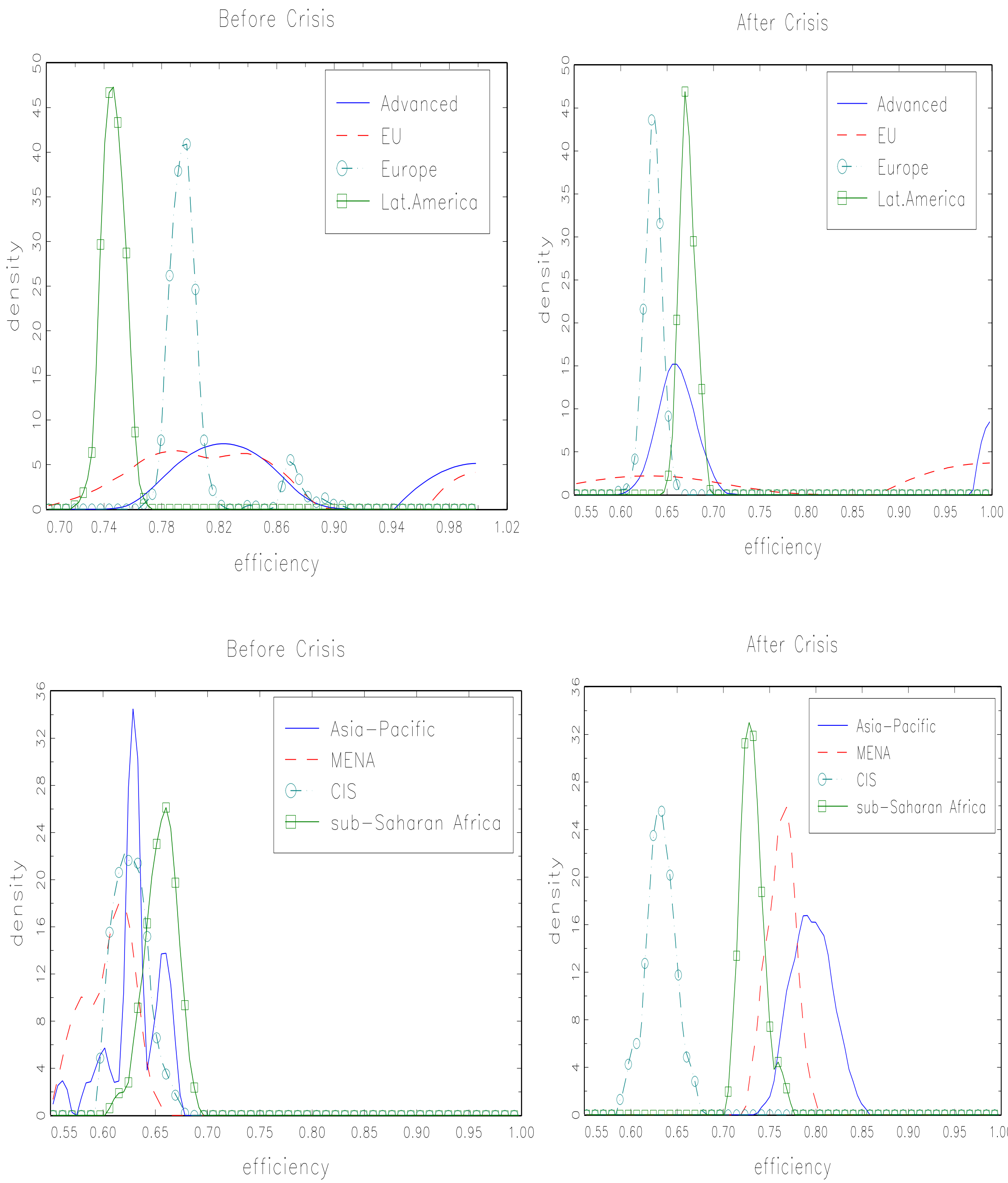
Note: The Table reports average bank efficiency for each country according to geographic region. The classification is based on IMF World Economic Outlook April 2014.

3.3 Densities of bank efficiency.

One of the advantages of the proposed methodology is that it allows deriving information for the density function of bank efficiency whilst taking into account that there is the possibility of having a fully efficient bank presence. In the previous section, we report that there is considerable variation in efficiency scores across the world, but also within selected group of countries, most notably the EU. One could consider many underlying reasons for such variability; given the extended time period of our sample we shall highlight the importance of the financial crisis in 2008. It is undoubtedly the case that bank efficiency changes over time due to the impact of the financial crisis. We capture this in the present analysis, by presenting densities of bank efficiency over time that is prior to and ex post the crisis. Figure 1 presents density of bank efficiency before and after the crisis for all the country groups we identify in the previous section. To facilitate the presentation, we present densities of bank efficiency for two wider groups at a time; first group presents banks in Advanced, EU, Europe and Latina America; whilst the second groups includes banks in Asia Pacific, Middle East and North Africa, Common Wealth and Independent States, and last Sub-Sahara Africa.

Figure 1 shows that efficiency scores have some presence to one (see Advanced and EU countries after the crisis), indicating that indeed there are some fully efficient banks in the sample. A standard stochastic frontier analysis would miss this point. Before the crisis, for the group countries in Advanced and EU efficiency scores take values from 0.7 to 1 with an average to 0.8, though there is evidence of a bipolar density for both groups with a long right. These results are of interest as it proves that the common assumption, in the literature of bank efficiency, of half normal distribution in efficiency is very restrictive, as it does not represent the true density. They also show that despite average efficiency is around 0.8, there are fully efficient banks. Alas, this picture dramatically changes after the crisis where the values spread from 0.55 to 1 and average efficiency is around 0.65. Also, it is worth noting that in EU the density after the crisis becomes more platycurtic in comparison to before the crisis, whereas for the advance economies the density of efficiency displays higher kurtosis. Banks in Europe (except EU countries) and Latin America exhibit high kurtosis in both periods, but there is a shift in densities to the left towards lower efficiency scores and in addition banks in Latin America overpass the efficiency scores of banks located in Europe. Interestingly, Figure 1 (see second line of diagrams) show that the crisis has not harm bank efficiency in countries in Asia Pacific, Middle East and North Africa and last Sub-Sahara Africa. In fact, densities for those countries shift to higher efficiency scores. For the banks Common Wealth and Independent States though there is slight deterioration in efficiency after the crisis.

Figure 1. Technical efficiency distributions; prior and ex post the credit crisis.

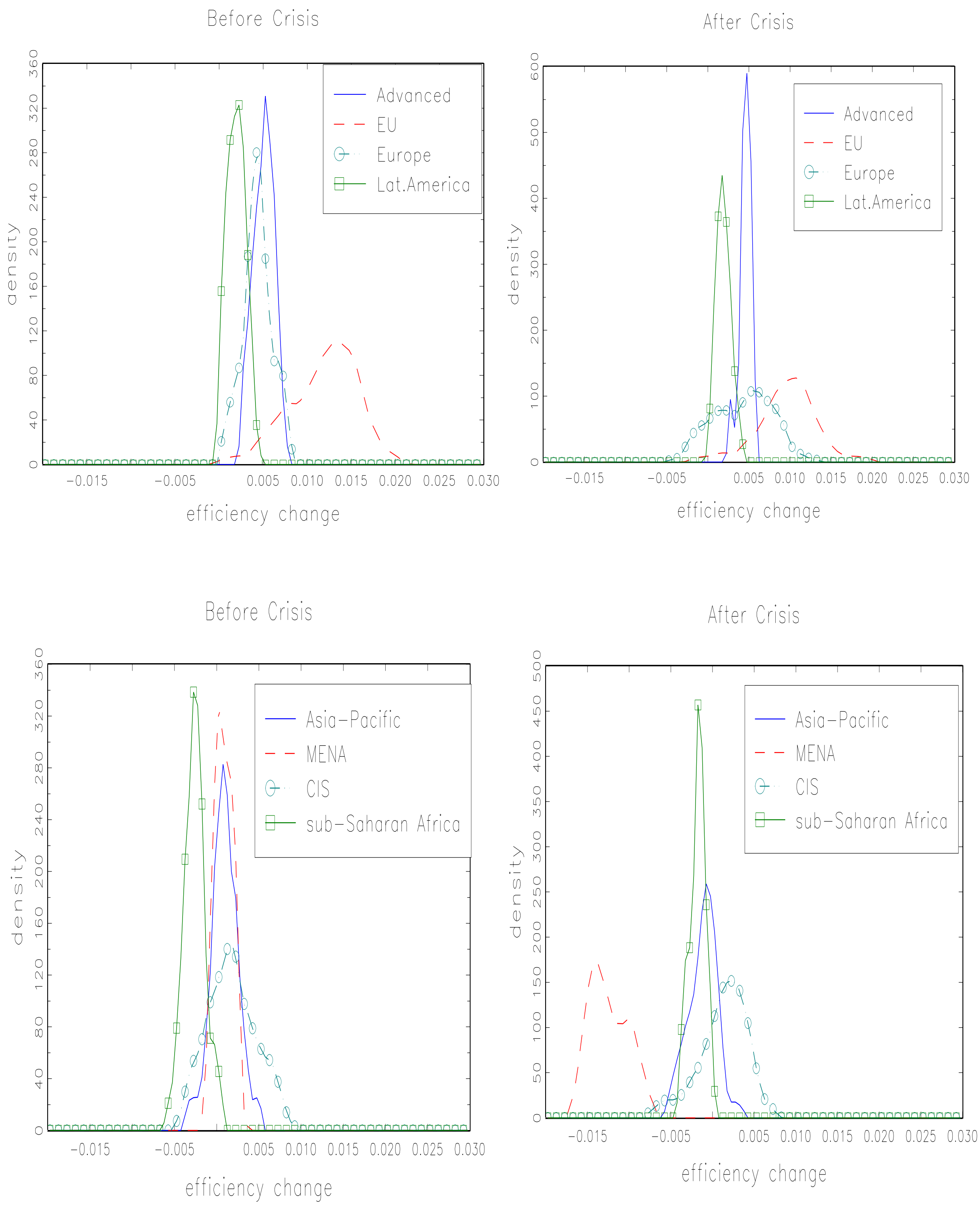


Note: To facilitate the presentation we present densities for two sets of countries; namely Advanced, EU, Europe (except EU) and Latin America (Lat. America); whilst the second set includes banks in Asia Pacific, Middle East and North Africa, (MENA), Common Wealth and Independent States (CIS), and last Sub-Sahara Africa.

To enrich the information regarding the density of efficiency when fully efficient banks are present we present next the density of changes in efficiency that captures the underlying dynamics around the financial crisis. As expected there is some variability of changes in efficiency before and after the crisis. During the second period, bank efficiency changes for Advanced and EU countries shows higher kurtosis towards zero. One of the main concerns that have been raised since the credit crunch is the low degree of alertness of financial systems prior to the crisis (Allen and Carletti (2010), Brunnermeier (2009), Covitz et al. (2013)). Following our modeling, our results show that signs of the crisis could have been identified well in advance and thereby allowing a better response to. In terms of the second grouping of countries, changes in bank efficiency in banks in Middle East and North Africa appear to move towards negative values after the crisis. To some extend similar pattern is observed for banks in Asia and Pacific.

These results clearly emphasize that there is striking heterogeneity across most of regions. This disparity in changes of efficiency testifies the complexity of financial world, so much so after the financial crisis.

Figure 2. Technical efficiency change distributions; prior and ex post the credit crisis.



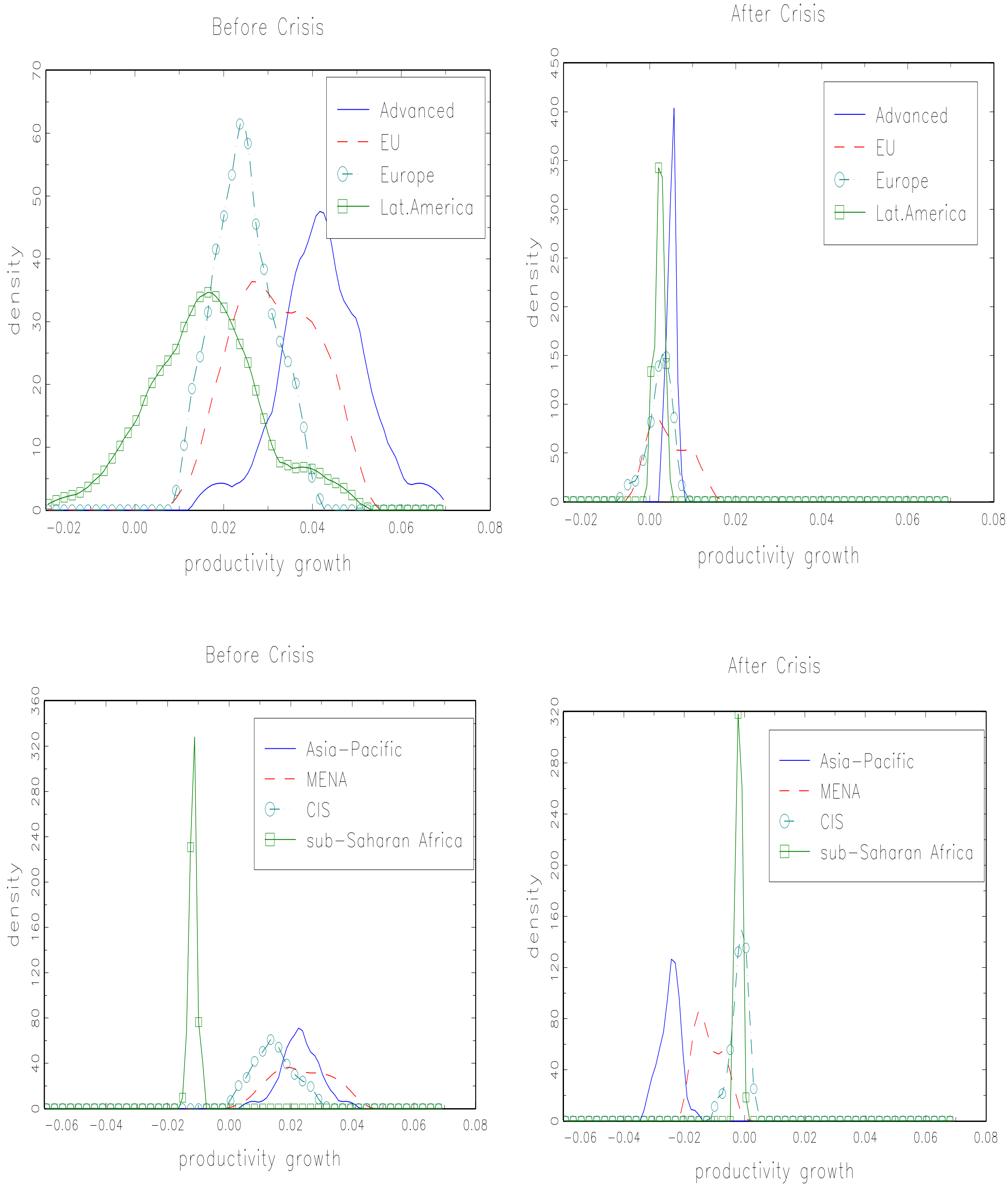
Note: To facilitate the presentation we present densities for two sets of countries; namely Advanced, EU, Europe (except EU) and Latin America (Lat. America); whilst the second set includes banks in Asia Pacific, Middle East and North Africa, (MENA), Common Wealth and Independent States (CIS), and last Sub-Sahara Africa.

3.4 Densities of productivity growth.

The flexibility of our modeling allows deriving the density, in addition to efficiency change presented in the previous, of efficiency change. This in turn, would allow to estimate the density of productivity growth, which is technical change plus efficiency change. This is the first time that bank productivity growth is estimated at global level, having accommodated of presence of fully efficient bank.[3] Figure 3 presents the density of productivity growth for the regions we identify as above. The Figure 3 clearly shows that the financial crisis has been very detrimental for bank productivity growth across the world, but the sub-Sahara Africa, as it noticeably shifts to the left hand side towards lower productivity growth level. This shift is as dramatic as in the case of efficiency, but even more so. Moreover, prior to the crisis productivity growth of banks in Advanced, EU, Europe and Latin America countries exhibit lower kurtosis compared to after the financial crisis when it appears to converge to lower level of productivity growth whilst densities are leptokurtic. The crisis clearly has changed the steady state of banks productivity growth to lower levels than before the crisis, though variability is also lower. Similar pattern, but less dramatic, is observed for bank productivity growth in countries in Asia Pacific, Middle East and North Africa and Common Wealth and Independent States. Banks in Sub-Sahara Africa register an improvement in their productivity growth towards zero after the crisis, compared to negative values prior to the crisis.

[3] Some bank productivity studies exist, but focus mostly on a single country, e.g. (Koetter and Noth, 2013; Kumbhakar and Sarkar, 2003; Martín-Oliver et al., 2013; Barros et al. 2009; and Assaf et al. 2011) or for a certain group of countries, i.e. in EU (Koutsomanoli and Mamatzakis, 2009; Delis et al., 2011; Fiordelisi and Molyneux, 2010).

Figure 3. Productivity growth distributions; prior and ex post the credit crisis.

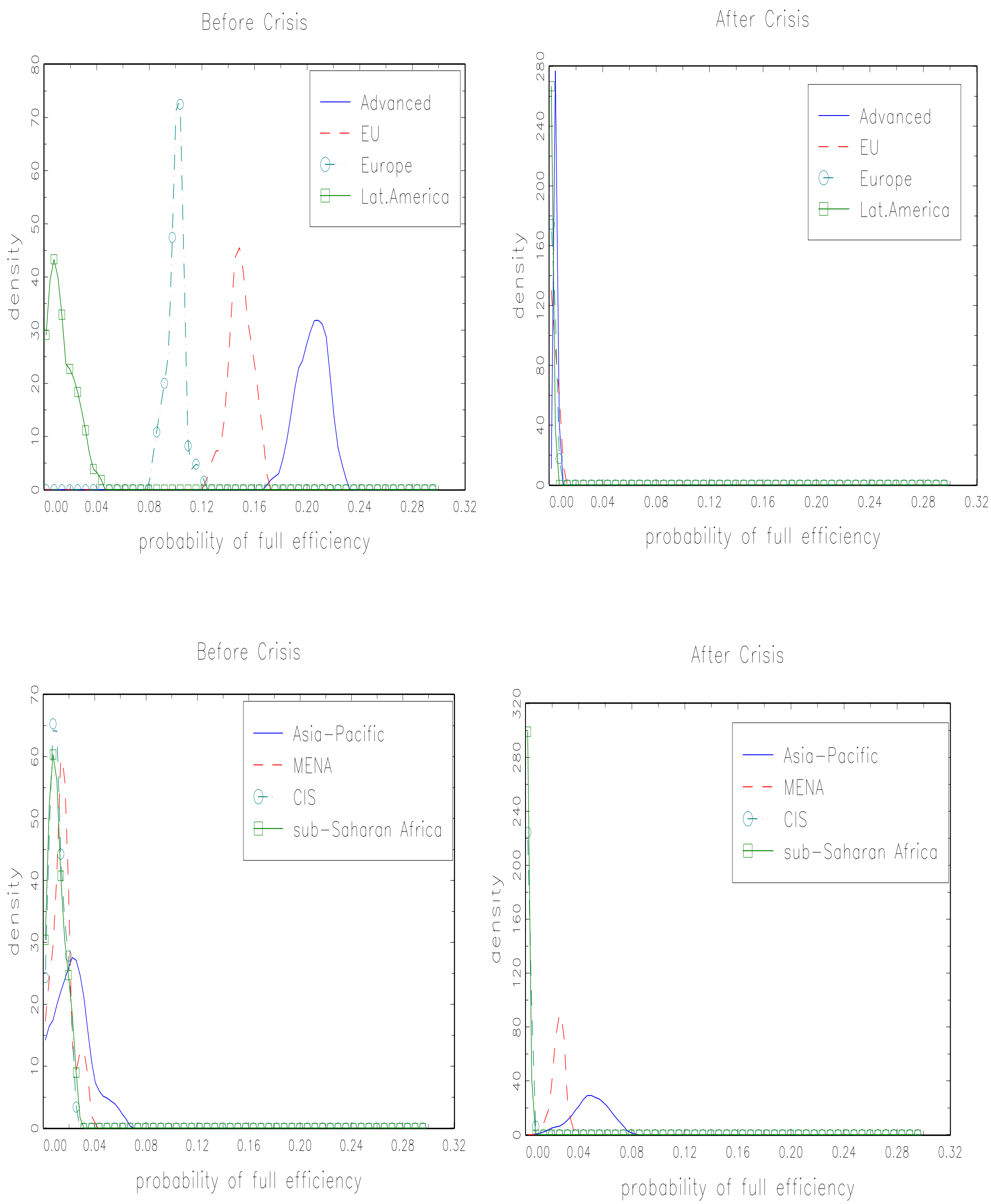


Note: To facilitate the presentation we present densities for two sets of countries; namely Advanced, EU, Europe (except EU) and Latin America (Lat. America); whilst the second set includes banks in Asia Pacific, Middle East and North Africa, (MENA), Common Wealth and Independent States (CIS), and last Sub-Sahara Africa.

3.5 Probability functions of global bank efficiency.

Our modeling allows estimating probability function of full efficient bank. Figure 4 presents the density functions of such probability functions at 100 points between 0 and 1. Once more results are acutely revealing of the impact of financial crisis on bank efficiency worldwide, and in particular in Advanced and EU countries. Prior to the crisis banks in advanced economies exhibit the higher probability of full efficiency around 20%, followed by banks in EU, Europe (excluding EU countries) and last by Latin America. Alas, since the crisis the probability of full efficiency collapses to zero for all banks in this group of countries. This is also true for banks in Common Wealth and Independent States (CIS), and Sub-Sahara Africa. For banks located in Asia Pacific and Middle East and North Africa countries the probability of fully efficient banks appears to hold but at very low levels and close to zero. Overall, these density functions demonstrate dramatically that the financial crisis has substantially reduced the probability of a bank, worldwide, to achieve full efficiency. It is the first time that such result comes to light as most studies focus on the level of bank efficiency post the crisis. Herein we argue that focusing on the level of bank efficiency is rather myopic as the main aftermath of the financial crisis has fundamentally shifted bank performance away from fully efficiency frontier. We observe a shift of the frontier to lower levels than a movement along the frontier as it is frequently quoted.

Figure 4. Probability functions of full efficiency; prior and ex post the credit crisis.



Note: To facilitate the presentation we present densities for two sets of countries; namely Advanced, EU, Europe (except EU) and Latin America (Lat. America); whilst the second set includes banks in Asia Pacific, Middle East and North Africa, (MENA), Common Wealth and Independent States (CIS), and last Sub-Sahara Africa.

3.6 The impact of control variables on global bank efficiency.

Having derived bank efficiency, taking into account full efficient banks, at country and global level would be of interest to examine their underlying association with key bank and country specific covariates. Table 5 presents marginal effect on bank efficiency for some key bank and country specific variables. As expected there exists a negative relationship between bank efficiency and Z-score that is in line with *"bad management hypothesis"* (Berger and DeYoung 1997, Koutsomanoli and Mamatzakis 2009). Berger and DeYoung, (1997) and Koutsomanoli and Mamatzakis, (2009) argue that the higher z-score, lower default risk, comes at a cost as it would result to divert disproportionally more resources from day-to-day activities to screening and monitoring operations. Our results show that this would eventually lead to lower efficiency. Liquidity ratio also negatively affects bank efficiency similarly to the impact of impact to that of Z-score. This has been particularly true during credit crunch in 2008 given conditions of low liquidity and depletion of capital that saw banks with lower liquidity to outperformed banks higher liquidity. Along these lines it is hardly surprising that capital ratio asserts a statistically significant negative impact on bank efficiency. This results bring into light a somewhat forgotten hypothesis the *"agency cost hypothesis"* (Jensen and Meckling (1979)), arguing that an increase in equity over total assets would raise agency costs for shareholders. Under certain conditions bank managers might increase the risk-taking activities as banks increase their debt through capital injections (Grossman and Hart (1982); Williams (1987)). Such practices would harm bank efficiency.

The fee ratio (see Table 5) negatively affects banks efficiency across all countries in the global sample. It appears that raising non-interest income would result to lower efficiency as the former is more volatile, and as such more risky, than interest income (Demirguc-Kunt and Huizinga (2010)). Diversification related to benefits from bank income generating appears not to be present in contrast to Merciera et al. (2007). On the other hand, the impact of securities over total assets ratio on efficiency is positive for all banks in the sample.

Table 5: Marginal effects of efficiency.

	Adv.	EU	EUR	LAT.AM.	As-P.	MENA	CIS	s.-SA
Z-score	-0.27**	-0.23**	-0.35**	-0.55**	-0.29***	-0.71**	-0.87**	-0.92**
	(0.012)	(0.014)	(0.010)	(0.022)	(0.017)	(0.032)	(0.044)	(0.017)
Capital ratio	-0.224**	-0.201**	-0.167**	-0.445**	-0.442**	-0.680**	-0.717**	-0.886**
	(0.023)	(0.017)	(0.022)	(0.013)	(0.032)	(0.012)	(0.022)	(0.002)

Fees	-0.054** (0.017)	-0.061** (0.012)	-0.044 (0.023)	-0.078** (0.001)	-0.068** (0.027)	-0.077** (0.016)	-0.44** (0.003)	-0.76** (0.005)
Liquidity ratio	-0.185** (0.015)	-0.173** (0.017)	-0.221** (0.011)	-0.477** (0.022)	-0.551** (0.021)	-0.354** (0.004)	-0.653** (0.002)	-0.880** (0.001)
Securities	-0.242** (0.026)	-0.250** (0.025)	-0.273** (0.014)	-0.551** (0.015)	-0.396** (0.016)	-0.448** (0.003)	-0.715** (0.003)	-0.897** (0.002)
GDP per capita	-0.017 (0.022)	-0.016 (0.013)	-0.022* (0.010)	-0.033* (0.015)	-0.044** (0.005)	-0.022 (0.017)	-0.553** (0.005)	-0.847** (0.012)
Inflation	0.012** (0.006)	0.009** (0.002)	0.013** (0.004)	0.171** (0.007)	0.025** (0.005)	0.018** (0.006)	0.165** (0.004)	0.014 (0.028)
Population density	-0.024** (0.002)	-0.016** (0.005)	-0.020** (0.007)	-0.018** (0.003)	-0.0031 (0.0044)	0.0044 (0.0151)	-0.091** (0.001)	-0.465** (0.004)
Market size	-0.017** (0.003)	-0.022** (0.004)	-0.032** (0.001)	-0.067** (0.014)	-0.044** (0.012)	-0.065** (0.019)	-0.155** (0.012)	-0.853** (0.032)
Trend	-0.0012 (0.0014)	-0.0020 (0.0013)	-0.0017 (0.0022)	0.0022 (0.0034)	-0.005 (0.003)	0.004 (0.015)	-0.017** (0.002)	-0.0012 (0.0156)

Notes: The table provides the marginal effects of efficiency with respect to the bank and country specific control variables. Standard errors are reported in parentheses. Z-score= (1+ROE)/(Standard Deviation of ROE); Capital ratio = equity over total assets; Liquidity ratio= liquid assets over total assets; Fees= net fees, commission and trading income over total assets; Securities= total securities over total assets. Country specific variables: GDP per capita; Inflation; Population density is the number of people per square klm; Market size= value of total shares traded on the stock market exchange. Trend captures time. Adv refers to Advanced countries, EU, EUR to Europe (except EU), LATAM to Latin America, As-P to Asia Pacific, MENA to Middle East and North Africa, CIS to Common Wealth and Independent States, and last s.-SA to Sub-Sahara Africa. ** refers to significance at 1%, * refers to significance to 5%.

With respect to the country-level control variables, we find that GDP per capita assert a positive and significant impact on bank efficiency in all country-groups, insinuating that increases of GDP per capita could improve efficiency as operating expenses improve (Dietsch and Lozano-Vivas (2000)). Inflation, on the other hand, has a negative impact on bank efficiency as it would increase uncertainty related to i.e. salaries and wages and thus increase operating costs and as such reduce bank efficiency Molyneux and Thornton (1992). Population density and market size increases bank efficiency as both would result to cost savings in operations (Dietsch et al. (2000)). Lastly it is striking that the trend has negative impact on bank efficiency, but it is hardly significant across all regions.

Table 6 presents average elastiticities of probability of full efficiency, $p(z)$, with respect to the bank and country specific control variables. Results are broadly in line with results in Table 5 in the extent that bank specific but also country specific variables are reported to be of importance for the probability of full efficiency. Note that all variables, apart from inflation and trend, increase the probability of full efficiency with economic and statistical significance to be higher for the bank specific variables, in particular those variables that are related to risk such as z-score and liquidity ratio. Improving the bank specific risk profile appears to increase the probability of fully bank efficiency. Alas, country specific uncertainty, such as inflation, would reduce the probability of fully efficiency.

Table 6: Marginal effects of probability of full efficiency, $p(z)$

	Adv.	EU	EUR	Lat.Am.	As-Pac.	MENA	CIS	s.-SA
Z-score	0.154**	0.132**	0.043*	0.055**	0.043**	0.252**	0.414**	0.551**
	(0.065)	(0.003)	(0.021)	(0.012)	(0.006)	(0.012)	(0.021)	(0.002)
Capital ratio	0.085	0.043**	0.035**	0.043**	0.022**	0.173	0.276**	0.445**
	(0.122)	(0.003)	(0.005)	(0.017)	(0.005)	(0.125)	(0.015)	(0.003)
Fees	0.044	0.017**	0.022	0.032**	0.018**	0.022	0.155**	0.327**
	(0.057)	(0.003)	(0.016)	(0.007)	(0.003)	(0.036)	(0.005)	(0.005)
Liquidity ratio	0.032*	0.022**	0.016**	0.017**	0.015**	0.027	0.128**	0.225**
	(0.015)	(0.001)	(0.002)	(0.002)	(0.005)	(0.019)	(0.021)	(0.004)
Securities	0.025	0.017	0.025**	0.014**	0.010**	0.022	0.120**	0.128**
	(0.017)	(0.015)	(0.004)	(0.004)	(0.004)	(0.032)	(0.015)	(0.004)
GDP per capita	0.017**	0.020	0.017	0.005	0.022	0.014**	0.224**	0.205
	(0.005)	(0.031)	(0.025)	(0.031)	(0.019)	(0.005)	(0.005)	(0.016)
Inflation	-0.005**	-0.017	-0.032	-0.055**	-0.044**	-0.024	-0.045**	-0.032
	(0.001)	(0.033)	(0.044)	(0.017)	(0.005)	(0.015)	(0.007)	(0.027)
Population density	0.0015	0.003	0.0017	0.007	0.017	0.014	0.0071	0.227**
	(0.022)	(0.005)	(0.025)	(0.006)	(0.014)	(0.022)	(0.0013)	(0.015)
Market size	0.022**	0.004	0.003	0.004	0.016**	0.005	0.0110**	0.322**
	(0.003)	(0.007)	(0.002)	(0.022)	(0.006)	(0.007)	(0.002)	(0.013)
Trend	-0.0014	-0.0012	-0.0013	-0.0016	-0.0022*	-0.0017	-0.035**	-0.018**
	(0.0022)	(0.0044)	(0.0032)	(0.0155)	(0.0010)	(0.0232)	(0.01)	(0.001)

Notes: The table provides the average elasticities of probability of full efficiency, $p(z)$, with respect to the bank and country specific control variables. Standard errors are reported in parentheses. Z-score= (1+ROE)/(Standard Deviation of ROE); Capital ratio = equity over total assets; Liquidity ratio= liquid assets over total assets; Fees= net fees, commission and trading income over total assets; Securities= total securities over total assets. Country specific variables: GDP per capita; Inflation; Population density is the number of people per square klm; Market size= value of total shares traded on the stock market exchange. Trend captures time. Adv refers to Advanced countries, EU, EUR to Europe (except EU), LAT.AM. to Latin America, As-P. to Asia Pacific, MENA to Middle East and North Africa, CIS to Common Wealth and Independent States, and last s.-SA to Sub-Sahara Africa.

4. Conclusions

This paper first provides an alternative semiparametric approach for estimating the ZISF model by allowing for the frontier to have an unknown smooth function of explanatory variables whilst maintaining the parametric assumption on the probability of fully efficient firms. In particular, we suggest a modified version of the iterative backfitting local maximum likelihood estimation developed in Tran and Tsionas (2016). We show that the proposed estimator achieves the optimal convergence rates for both parameters of the probability of fully efficient firm and the nonparametric function of the frontier. We provide asymptotic properties of the proposed estimator. The finite sample performances of the proposed estimator are examined via Monte Carlo simulations.

Next, we use the proposed method to examine the productivity growth and efficiency of the global banking system. Overall, our analysis demonstrate that the financial crisis has substantially reduced the probability of a bank, worldwide, to achieve full efficiency. It is the first time that such result comes to light as most studies focus on the level of bank efficiency post the crisis. As such previous literature focusing on the level of bank efficiency rather than the change of efficiency in the presence of fully efficient banks is rather myopic. We argue that the main aftermath of the financial crisis is that bank performance has shifted away from fully efficiency frontier, whilst also productivity growth shifted to lower values. This is a downward shift of the whole frontier rather than movement along the frontier. As policy implication we suggest that improving bank's risk profile would increase the probability of fully bank efficiency, as reducing macroeconomic uncertainty, such as inflation, would also do.

Finally, we did not consider hypothesis testing of parametric vs. nonparametric frontier and/or whether all banks are fully inefficient/efficient in this paper because they are beyond the scope of this paper. However, these topics are of interest in their own rights and deserve attention for future research.

References

Aigner, D.J., Lovell, C.A.K., & Schmidt P. (1977). Formulation and estimation of stochastic frontier production models. *Journal of Econometrics*, 6 (1), 21-27.

Alam, I. M. S. (2001). A nonparametric approach for assessing productivity dynamics of large US banks. *Journal of Money, Credit and Banking*, 121-139.

Allen, L., & Rai, A. (1996). Operational efficiency in banking: An international comparison. *Journal of Banking & Finance*, 20(4), 655-672.

Amel, D., Barnes, C., Panetta, F., & Salleo, C. (2004). Consolidation and efficiency in the financial sector: a review of the international evidence. *Journal of Banking & Finance*, 28(10), 2493-2519.

Atkinson, S. E., & Dorfman, J. H. (2005). Bayesian measurement of productivity and efficiency in the presence of undesirable outputs: crediting electric utilities for reducing air pollution. *Journal of Econometrics*, 126(2), 445-468.

Barth, J. R., Caprio Jr, G., & Levine, R. (2004). Bank regulation and supervision: what works best? *Journal of Financial Intermediation*, 13(2), 205-248.

Bekaert, G., & Harvey, C. R. (2003). Emerging markets finance. *Journal of Empirical Finance*, 10(1), 3-55.

Berg, S. A., Førsund, F. R., & Jansen, E. S. (1992). Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980-89. *The Scandinavian Journal of Economics*, S211-S228.

Berger, A. N., & DeYoung, R. (1997). Problem loans and cost efficiency in commercial banks. *Journal of Banking & Finance*, 21(6), 849-870.

Berger, A. N., & 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.

Boucinha, M., Ribeiro, N., & Weyman-Jones, T. (2013). An assessment of Portuguese banks' efficiency and productivity towards euro area participation. *Journal of Productivity Analysis*, 39(2), 177-190.

Charnes, A., W. Cooper, & E., Rhodes (1978). Measuring the efficiency of decision-making units, *European Journal of Operational Research*, vol. 2, 429–444.

Chaffai, M. E., Dietsch, M., & Lozano-Vivas, A. (2001). Technological and environmental differences in the European banking industries. *Journal of Financial Services Research*, 19(2-3), 147-162.

Delis, M. D., Molyneux, P., & Pasiouras, F. (2011). Regulations and productivity growth in banking: Evidence from transition economies. *Journal of Money, Credit and Banking*, 43(4), 735-764.

Demirgüç-Kunt, A., & Huizinga, H. (1999). Determinants of commercial bank interest margins and profitability: some international evidence. *The World Bank Economic Review*, 13(2), 379-408.

Demirgüç-Kunt, A., & Huizinga, H. (2010). Bank activity and funding strategies: The impact on risk and returns. *Journal of Financial Economics*, 98(3), 626-650.

Demirguc-Kunt, A., Laeven, L., & Levine, R. (2003). Regulations, market structure, institutions, and the cost of financial intermediation. *National Bureau of Economic Research*, No. W9890.

Dietsch, M., & Lozano-Vivas, A. (2000). How the environment determines banking efficiency: A comparison between French and Spanish industries. *Journal of Banking & Finance*, 24(6), 985-1004.

Feng, G., & Serletis, A. (2010). Efficiency, technical change, and returns to scale in large US banks: Panel data evidence from an output distance function satisfying theoretical regularity. *Journal of Banking & Finance*, 34(1), 127-138.

Feng, G., & Zhang, X. (2014). Returns to scale at large banks in the US: A random coefficient stochastic frontier approach. *Journal of Banking & Finance*, 39, 135-145.

Fukuyama, H. (1995) Measuring efficiency and productivity growth in Japanese banking: a nonparametric frontier approach. *Applied Financial Economics*, 5, 95-107.

Gorton, G., & Rosen, R. (1995). Corporate control, portfolio choice, and the decline of banking. *The Journal of Finance*, 50(5), 1377-1420.

Griffin, J. & Steel, M.F.J. (2004). Semiparametric Bayesian Inference for Stochastic Frontier Models. *Journal of Econometrics*, 123, 121-152.

Grossman, S. J., & Hart, O. D. (1982). Corporate financial structure and managerial incentives. In The economics of information and uncertainty (pp. 107-140). University of Chicago Press.

Herd, R., & Dougherty, S., (2007). Growth prospects in China and India compared. *The European Journal of Comparative Economics*, 4, 65-89.

Huang, M. & Yao, W. (2012). Mixture of regression model with varying mixing proportions: A semiparametric approach. *Journal of the American Statistical Association*, 107, 711-724.

Huang, M., Li, R. & Wang, S. (2013). Nonparametric mixture regression models. *Journal of the American Statistical Association*, 108, 929-941.

Hume, M., & Sentance, A. (2009). The global credit boom: Challenges for macroeconomics and policy. *Journal of International Money and Finance*, 28(8), 1426-1461.

Jensen, M. C., & Meckling, W. H. (1979). Theory of the firm: Managerial behavior, agency costs, and ownership structure (pp. 163-231). Springer Netherlands.

Koutsomanoli-Filippaki, A., 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, 2050–2061.

Kumbhakar, S. C., Lozano-Vivas, A., Lovell, C. K., & Hasan, I. (2001). The effects of deregulation on the performance of financial institutions: the case of Spanish savings banks. *Journal of Money, Credit and Banking*, 101-120.

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

Kumbhakar, S. C., Parmeter, C. F. & Tsionas, E.G. (2013). A zero inefficiency stochastic frontier model. *Journal of Econometrics*, 172, 66-76.

Kwan, S. H. (2003). Operating performance of banks among Asian economies: An international and time series comparison. *Journal of Banking & Finance*, 27(3), 471-489.

Li, R. & Liang, H. (2008). Variable selection in semiparametric modeling. *Annals of Statistics*, 36, 261-286.

Li, Q. & Racine, J. (2007). Nonparametric econometrics. Princeton, NJ: Princeton University Press.

Martins-Filho, C. & Yao, F. (2015). Semiparametric stochastic frontier estimation via profile likelihood. *Econometric Reviews* (forthcoming).

Meeusen, W., & van den Broeck, J. (1997). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18 (2), 435-444.

Merciera, S., Schaeck K., and Wolfe S. (2007). Small European banks: Benefits from diversification? *Journal of Banking and Finance*, 31, pages 1975–98.

Molyneux, P., & Thornton, J. (1992). Determinants of European bank profitability: a note. *Journal of Banking & Finance*, 16(6), 1173-1178.

Myers, S. C. (1977). Determinants of corporate borrowing. *Journal of Financial Economics*, 5(2), 147-175.

Olson, D., & Zoubi, T. A. (2011). Efficiency and bank profitability in MENA countries. *Emerging Markets Review*, 12(2), 94-110.

Orea, L. (2002). Parametric decomposition of a generalized Malmquist productivity index. *Journal of Productivity Analysis*, 18(1), 5-22.

Parmeter, C. & Kumbhakar, S.C. (2014).

Pastor, J., Perez, F., & Quesada, J. (1997). Efficiency analysis in banking firms: An international comparison. *European Journal of Operational Research*, 98(2), 395-407.

Rho, S. & Schmidt, P. (2015). Are all firms inefficient? *Journal of Productivity Analysis*, 43, 327-349.

Sealey, C., Lindley, J., (1977) Inputs, Outputs, and a Theory of Production and Cost at Depository Financial Institutions. *The Journal of Finance*, 32, 1251-1266.

Staikouras, C., Mamatzakis, E., & Koutsomanoli-Filippaki, A. (2008). Cost efficiency of the banking industry in the South Eastern European region. *Journal of International Financial Markets, Institutions and Money*, 18(5), 483-497.

Tirtiroglu, D., Daniels, K. N., & Tirtiroglu, E. (2005). Deregulation, intensity of competition, industry evolution, and the productivity growth of US commercial banks. *Journal of Money, Credit, and Banking*, 37(2), 339-360.

Tortosa-Ausina, E., Grifell-Tatjé, E., Armero, C., & Conesa, D. (2008). Sensitivity analysis of efficiency and Malmquist productivity indices: An application to Spanish savings banks. *European Journal of Operational Research*, 184(3), 1062-1084.

Tran, K.C. & Tsionas, E. G. (2016). Zero-inefficiency stochastic frontier models with varying mixing proportion: A semiparametric approach. *European Journal of Operational Research*, 249, 1113-1123.

Wheelock, D. C., & Wilson, P. W. (2000). Why do banks disappear? The determinants of US bank failures and acquisitions. *Review of Economics and Statistics*, 82(1), 127-138.

Appendix A: Proofs of the theorems

Let $\tilde{\gamma}(.) = \{\tilde{\mu}(.), \tilde{\sigma}^2(.), \tilde{\lambda}(.)\}'$. Also let $\gamma(.) = \{\mu(.), \sigma^2(.), \lambda(.)\}'$ and α denote the true values.

Proof of Theorem 1: The proof of this theorem follows similarly to the proof of Theorem 1 of Tran and Tsionas (2016) and Huang and Yao (2012). Thus, we only outline the key steps of the proof.

To derive the asymptotic properties of $\hat{\alpha}$, we first let

$$\hat{\alpha}^* = \sqrt{n}(\hat{\alpha} - \alpha),$$

$$\ell(\tilde{\gamma}(X_i), \alpha, Z_i, Y_i) = \log f(Y_i \mid \tilde{\gamma}(X_i), \alpha, Z_i)$$

$$\ell(\tilde{\gamma}(X_i), \hat{\alpha} + n^{-1/2}\alpha^*, Z_i, Y_i) = \log f(Y_i \mid \tilde{\gamma}(x_i), \hat{\alpha} + n^{-1/2}\alpha^*, Z_i)$$

Then $\hat{\alpha}^*$ is the maximize of

$$L_n(\alpha^*) = \sum_{i=1}^{n} \{\ell(\tilde{\gamma}(X_i), \alpha + n^{-1/2}\alpha^*, Z_i, Y_i) - \ell(\tilde{\gamma}(X_i), \alpha, Z_i, Y_i)\} \quad \text{(A.1)}$$

By using a Taylor series expansion and after some calculation, yields

$$L_n(\alpha^*) = A_n\alpha^* + \frac{1}{2}\alpha^{*\prime}B_n\alpha^* + o_p(1) \quad \text{(A.2)}$$

where

$$A_n = n^{-1/2}\sum_{i=1}^{n} \frac{\partial \ell(\tilde{\gamma}(X_i), \alpha, Z_i, Y_i)}{\partial \alpha}$$

$$B_n = n^{-1}\sum_{i=1}^{n} \frac{\partial^2 \ell(\tilde{\gamma}(X_i), \alpha, Z_i, Y_i)}{\partial \alpha \partial \alpha'}$$

Next we evaluate the terms A_n and B_n. First, expanding A_n around $\gamma(X_i)$, we obtain

$$A_n = n^{-1/2}\sum_{i=1}^{n}\frac{\partial \ell(g(X_i),\alpha,Z_i,Y_i)}{\partial \alpha} + n^{-1/2}\sum_{i=1}^{n}\frac{\partial^2 \ell(g(X_i),\alpha,Z_i,Y_i)}{\partial \alpha \partial \gamma'}[\tilde{g}(X_i) - g(X_i)]$$
$$+ O_p(n^{-1/2}\,\|\tilde{g}(.) - g(.)\|_\infty^2)$$

$$= n^{-1/2}\sum_{i=1}^{n}\frac{\partial \ell(g(X_i),\alpha,Z_i,Y_i)}{\partial \alpha} + D_{1n} + O_p(n^{-1/2}\,\|\tilde{g}(.) - g(.)\|_\infty^2)$$

where the definition of D_{1n} should be apparent. Following Tran and Tsionas (2016), it can be shown that

$$A_n = n^{-1/2}\sum_{i=1}^{n}\left\{\frac{\partial \ell(g(X_i),\alpha,Z_i,Y_i)}{\partial \alpha} - I_{\alpha\gamma}(X_i)\delta(X_i,Y_i,Z_i)\right\} + o_p(1) \tag{A.3}$$

where $\delta(X,Y,Z)$ is the first $r \times r$ submatrix of $I_{\theta\theta}^{-1}(X)\theta_{\theta}(\theta(X),Z,Y)$. Similarly, for B_n, it can be shown that

$$B_n = -E[I_{\alpha\alpha}(X)] + o_p(1) = B + o_p(1) \tag{A.4}$$

Thus, from (A.2) in conjunction with (A.4) and an application of quadratic approximation lemma (see for example Fan and Gijbel (1996, p. 210)), leads to

$$\hat{\alpha}^* = B^{-1}A_n + o_p(1) \tag{A.5}$$

if A_n is a sequence of stochastically bounded vectors. Consequently, the asymptotic normality of $\hat{\alpha}^*$ follows from that of A_n. Note that since A_n is the sum of i.i.d. random vectors, it suffices to compute the mean and covariance matrix of A_n and evoke the Central Limit Theorem. To this end, from (A.3), we have

$$E(A_n) = n^{1/2}E\left\{\frac{\partial \ell(g(X),\alpha,Z,Y)}{\partial \alpha} - I_{\alpha\gamma}(X)\delta(X,Y,Z)\right\} \tag{A.6}$$

The expectation of each element of the first term on the right hand side can be shown to be equal to 0 and further calculation shows that $E\,\{I_{\alpha\gamma}(X)\delta(X,Y,Z)\} = 0$. Thus $E(A_n) = 0$. The variance of A_n is $Var(A_n) = Var\left\{\frac{\partial \ell(g(x),\alpha,Z,y)}{\partial \alpha} - I_{\alpha\gamma}(X)\delta(X,Y,Z)\right\} = \Sigma$. By the Central Limit Theorem, we obtain the desired result.

Proof of Theorem 2: Recall that, given the estimate of $\hat{a}$, $\hat{g}(x)$ maximizes (7). By redefining appropriate notations:

$h(x_0, X) = g_0(x_0) + \Gamma_1(x_0)(X - x_0)$,

$g^* = (n|H|)^{1/2}\{g - g_0(x_0), |H|(g' - \Gamma_1(x_0))\}'$,

then the proof follows directly from the proof of theorem 2 of Tran and Tsionas (2016). Thus, we omit it here for brevity.

Appendix B: Fully Localised Model

The discussion in Section 2 has been limited to the case where the probability of fully efficient firm $p(z)$ is assumed to have a logistic function. In this Appendix, we extend the model to allow for nonparametric function $p(z)$. We will consider two cases. In the first case, we assume that $Z = X$ and show how to estimate this model as well as discuss the asymptotic properties of the local MLE. In the second case where in general $Z \neq X$, we will briefly discuss only the estimation procedure but not the asymptotic properties since they are more complicated and beyond the scope of this paper.

Case 1: When $Z = X$

In this case, we first redefine the vector function $q(x) = (p(x), g(x)')'$ and for a given set point x_0 and x in the neighbourhood of x, we approximate the function $q(x)$ by a linear function similar to (5),

$$q(x_0) \approx q_0(x_0) + \Theta_1(x_0)(x - x_0),$$

where $q_0(x_0)$ is a (4×1) vector and $\Phi_1(x_0)$ is a $(4 \times d)$ matrix of the first-order derivatives. Then the conditional local log-likelihood function is:

$$L_{5n}(\theta_0(x_0),\Theta_1(x_0)) = \sum_{i=1}^{n}\{\log f(Y_i;\theta_0(x_0)+\Theta_1(x_0)(X_i - x_0))\}K_H(X_i - x_0), \quad \text{(B.1)}$$

where the kernel function $K_H(X_i - x_0)$ is defined as before. Let $\hat{\theta}_0(x_0)$ denote the local maximizer of (B.1). Then the local MLE of $\theta(x)$ is given by $\hat{\theta}(x) = \hat{\theta}_0(x_0)$. To obtain the asymptotic property of $\hat{\theta}(.)$, we modify the following notations:

$$q_1(\theta(x),Y) = \partial L_5(\theta(x),Y)/\partial\theta, \quad q_2(\theta(x),Y) = \partial^2 L_5(\theta(x),Y)/\partial\theta\partial\theta',$$

$$I(x) = -E\{q_2(\theta(X),Y) \mid X = x\}, \text{ and } \Psi(u \mid x) = \int_Y q_1(\theta(x),Y)f(Y \mid \theta(u))dY$$

Assumptions:

B1: The support for X, denoted by $\mathcal{X}$, is compact subset of $\mathbb{R}^d$. Furthermore, the marginal density $f(x)$ of X is twice continuously differentiable and positive for $x \in \mathcal{X}$.

B2: The unknown function $\theta(x)$ have continuous second derivatives and in addition, $\sigma^2(x) > 0$ and $0 < p(x) < 1$ hold for all $x \in \mathcal{X}$.

B3: There exists a function $M(y)$, with $E[M(y)] < \infty$ such that for all Y, and all $\theta \in nbhd$ of $\theta(x)$, $|\partial L_5(\theta,Y)/\partial\theta_j\partial\theta_k\partial\theta_l| < M(y)$.

B4: The following conditions hold for all i and j:

$$E\{|\partial L_5(\theta(x),Y)/\partial\theta_j|^3\} < \infty, \quad E\{(\partial^2 L_5(\theta(x),Y)/\partial\theta_i\partial\theta_j)^2\} < \infty.$$

B5: The kernel function $K(.)$ has bounded support and satisfies:

$$(\int K(u)du)I_d = 1, \quad (\int uK(u)du)I_d = 0, \quad (\int u^2K(u)du)I_d < \infty,$$

$$(\int K^2(u)du)I_d < \infty, \quad (\int |K(u)|^3 du)I_d < \infty.$$

B6: $|H| \to 0$, $n|H| \to \infty$, and $n|H|^5 = O(1)$ as $n \to \infty$.

Proposition 1: Suppose that conditions (B1)-(B6) hold. Then it follows that

$$(n\,|H|)^{1/2}\{\hat{\theta}(x) - \theta(x) - B(x) + o(|H|^2)\} \xrightarrow{D} N(0, \kappa_0 f^{-1}(x) I_{\theta\theta}^{-1}),$$

where $B(x) = \frac{1}{2}\mu_2 \,|H|^2\, I_{\theta\theta}^{-1}(z)\Psi''(x\,|\,x)$ with κ_0 and μ_2 are defined as in Section 2.

The proof of Proposition 1 is a straightforward extension of the proof of Theorem 2 in Huang, Li and Wang (2013) to the multivariate case, and hence will be omitted.

<u>Case 2</u>: When $Z \neq X$

For this case the local MLE estimation is similar to case 1, albeit it is more complicated. To see this, let us once again redefine the vector function $\theta(z,x) = (p(z), g(x)')'$, then for a given set points z_0 and x_0, approximate $\theta(z,x)$ linearly as before. Also, define the kernel function for z as $W_A(Z_i, z_0) = |A|^{-1} W(A^{-1}(Z_i - z_0))$ where $W(v) = \prod_{j=1}^{r} w(v_j)$ with $w(.)$ is a univariate probability function, A is a bandwidth matrix and $|A| = a_1 a_2 \ldots a_r$. Then the modified conditional local log-likelihood function can be written as:

$$L_{6n}(\theta_0(z_0,x_0), \Theta_1(z_0,x_0)) = \sum_{i=1}^{n}\{\log f(Y_i; \theta_0(z_0,x_0) + \Theta_1(z_0,x_0)(Z_i - z_0)(X_i - x_0))\} W_{A_1}(Z_i - z_0) K_{H_1}(X_i - x_0), \qquad \text{(B.2)}$$

Let $\theta^*(z_0,x_0)$ be the maximizer of (14) where $\theta^*(z_0,x_0) = (p^*(z_0,x_0), g(z_0,x_0)')'$, then the local MLE of $\theta(.,.) = (p(.,.), g(.,.)')'$ is given by $\tilde{p}(z,x) = p^*(.,.)$ and $\tilde{g}(z,x) = g^*(.,.)$. Note that however, since the $p(z)$ do not depend on x and $g(x)$ do not depend on z, the improved estimators of $p(z)$ and $g(x)$ can be obtained using integrated backfitting approach. Thus, given the estimates $\tilde{p}(z,x)$ and $\tilde{g}(z,x)$, the initial estimates of $p(z)$ and $g(x)$ (up to additive constants) are given by

$$\tilde{p}(z) = \int \tilde{p}(z,x) f_X(x)dx$$
,
$$\tilde{g}(x) = \int \tilde{g}(z,x) f_Z(z)dz$$

where $f_X(x)$ and $f_Z(z)$ are marginal densities of X and Z, respectively. Now given the initial estimator of $\tilde{p}(z)$, For every fixed set points x_0 within the closed support of X, the improved estimator of $g(x_0)$ is defined as $\hat{g}(x_0) = \hat{g}_0(x_0) = \hat{g}_0$ where $\hat{g}_0$ is the first minimizer of the following plug-in conditional local log-likelihood function:

$$L_{7n}(\tilde{p}(z_i), g_0(x_0), \Gamma_1(x_0)) = \sum_{i=1}^{n} \{\log f(Y_i; \tilde{p}(z_i), g_0(x_0) + \Gamma_1(x_0)(X_i - x_0))\} K_{H_2}(X_i - x_0).$$
(B.3)

Given the estimates of $\hat{g}(x_i)$, we can obtain the improved estimator of $p(z_i)$, denote by $\hat{p}(z_0) = \hat{p}_0(z_0) = \hat{p}_0$ where $\hat{p}_0$ is the first maximizer the following plug-in conditional local log-likelihood function:

$$L_{8n}(p(z_0), \hat{g}(x_i)) = \sum_{i=1}^{n} \{\log f(Y_i; \hat{g}(x_i), p_0(z_0) + \Pi_1(z_0)(Z_i - z_0))\} W_{A_2}(Z_i - z_0). \quad \text{(B.4)}$$