

Zero-Inefficiency Stochastic Frontier Models with Varying Mixing Proportion: A Semiparametric Approach

Kien C. Tran*

Department of Economics
University of Lethbridge
4401 University Drive W
Lethbridge, Alberta
T1K 3M4 CANADA

and

Mike G. Tsionas

Department of Economics
Lancaster University Management School
LA1 4YX U.K.

Abstract

In this paper, we propose a semiparametric version of the zero-inefficiency stochastic frontier model of Kumbhakar, Parmeter and Tsionas (2013) by allowing for the proportion of firms that are fully efficient to depend on a set of covariates via unknown smooth function. We propose a (iterative) backfitting local maximum likelihood estimation procedure that achieves the optimal convergence rates of both frontier parameters and the nonparametric function of the probability of being efficient. We derive the asymptotic bias and variance of the proposed estimator and establish its asymptotic normality. In addition, we discuss how to test for parametric specification of the proportion of firms that are fully efficient as well as how to test for the presence of fully inefficient firms, based on the sieve likelihood ratio statistics. The finite sample behaviors of the proposed estimation procedure and tests are examined using Monte Carlo simulations. An empirical application is further presented to demonstrate the usefulness of the proposed methodology.

Keywords: Zero-inefficiency, Varying proportion, Semiparametric approach, Backfitting local maximum likelihood, Sieve likelihood ratio statistics.

*Corresponding author. Emails: kien.tran@uleth.ca and tsionas@aub.gr. An earlier version of this paper was presented at the North American Productivity Workshop (NAPW) in Ottawa, June 4-7, 2014. The authors would like to thank the Editor, I. Bomze and two anonymous referees for constructive comments and suggestions that substantially improved an earlier version of this paper.

1. Introduction

One drawback or restrictive assumption of estimating productivity and efficiency through stochastic frontier analysis originally proposed by Aigner et al. (1977), Meeusen and van den Broeck (1977), (see also, Ondrich and Ruggiero (2001)) was recently pointed out by Kumbhakar, Parmeter and Tsionas (2013, KPT hereafter) and Rho and Schmidt (2015, RS hereafter). The assumption that, a priori, all firms are inefficient and their inefficiency is modelled through a continuous density was shown to have considerable implications. When some firms are, in fact, fully efficient, a fact that we cannot preclude on prior grounds, applying stochastic frontier analysis with the familiar distributions (half-normal, exponential etc.) results in biased estimates of inefficiency.

To overcome this draw back, KPT (and independently by RS) proposed a new model for which they call “zero-inefficiency stochastic frontier (ZISF) model, that allows for the inefficiency term to have mass at zero with certain probability, π and a continuous distribution, with probability, $1 - \pi$. In essence, their model takes a special form of the latent class model considered by, among others, Ivaldi et al. (1995), Caudill (2003), Orea and Kumbhakar (2004) and Greene (2005). The interesting feature of the proposed model is that only non-existence and existence of inefficiency differs but not the frontier itself. KPT and RS also extend the ZISF model to allow for π to depend on a set of covariates via a logit or a probit function. Estimation of the model parameters can be carried out by using either standard maximum likelihood or E-M algorithm (see RS).

In this paper we use a non-parametric formulation for the probability as a function of covariates, $\pi(\cdot)$ which does not impose restrictive assumptions on what determines full efficiency. The issue is important as misspecification of the parametric form of probability has implication for estimating technical efficiency and, more specifically, which firms are fully efficient. Although functional forms for production or cost functions are more or less established in applied studies, this is not so for the functional form of the probability of firms being fully efficient, $\pi(\cdot)$. This is quite important since the functional form of $E(y | X)$ depends on the functional form of $\pi(\cdot)$ and the covariates.

To accommodate for the unknown probability of firms being efficient function in the estimation, we develop an iterative backfitting local maximum likelihood procedure which is

fairly simple to compute in practice. We also derive the necessary asymptotic theory of the proposed estimator. Specifically, we derive the asymptotic bias and variance of the proposed estimator and established its asymptotic normality. In addition, we discuss how to test for parametric specification of the probability function of firms that are fully efficient as well as how to test for the presence of fully efficient firms, based on the bootstrap sieve likelihood ratio statistics (Fan et al. (2001)).

We use both Monte Carlo experiments and real-world data from U.S banks to illustrate the applicability of the new model, and compare the results with the standard stochastic frontier models as well as the model proposed by KPT where the probability is a parametric function of covariates. Our Monte Carlo results indicate that the proposed estimation methods as well as the bootstrap sieve likelihood ratio statistics perform well in samples of the size typically used in applied econometric studies.

The rest of the article is organized as follows. Section 2 introduces the semiparametric zero-inefficiency stochastic frontier model. Section 3 derives the backfitting local maximum likelihood estimator and discusses construction of inefficiency scores. Section 4 establishes the asymptotic properties of the proposed estimator. Hypothesis testing for the parametric specification of probability of firms being fully efficient as well as testing for the presence of fully inefficient firms are discussed in Section 5. Monte Carlo simulations are presented in Section 6, while Section 7 provides an empirical application to the U.S. banking industry. Section 8 provides concluding remarks. Proofs of the theorems are gathered in the Appendix.

2. The Model

We consider the following semiparametric version of the zero-inefficiency stochastic frontier (SP-ZISF) model of KPT:

$$y_i = \begin{cases} x_i' \beta + v_i & \text{with probability } \pi(z_i) \\ x_i' \beta + v_i + su_i & \text{with probability } 1 - \pi(z_i) \end{cases}, \quad (1)$$

where y_i is a scalar representing output of firm i , x_i is a $d \times 1$ vector of inputs, v_i is random noise, u_i is one-sided random variable representing technical inefficiency, $s = +1$ for cost

frontier and $s = -1$ for production frontier, $\pi(\cdot)$ is an unknown smooth function representing the proportion of firms that are fully efficient and z_i is a $q \times 1$ vector of covariates which influence whether a firm is inefficient or not; and the z_i may or may not be a subset of x_i . Note that in (1) the technology is the same for both regimes, and the composed error is $v_i - u_i(1 - 1\{u_i = 0\})$ where $1\{\cdot\}$ is an indicator function and $P(1\{u_i = 0\}) = \pi(z_i)$. For illustration purpose, we focus mainly on the production frontier. Cost frontier can be handled in the same way by replacing the negative sign on u_i by a positive sign. In addition, to simplify our discussion, we consider univariate z . Extension to multivariate z is straightforward but at the expenses of increasing notational complexity and the “curse of dimensionality” problem.

Identification Issues:

Under the standard stochastic frontier framework, there is no identification issue arise since the parameter σ_u^2 , the variance of u_i is identified through the moment restrictions on the composed errors $\varepsilon_i = v_i - u_i$. However, in the context of model (1), we have an additional parameter $\pi(\cdot)$ which can be identified only if there are non-zero observations in each class. In addition, as KPT and RS point out, when $\sigma_u^2 \rightarrow 0$, $\pi(\cdot)$ is not identified since the two classes become indistinguishable. Conversely, when $\pi(\cdot) \rightarrow 1$ 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 $\pi(\cdot)$ to be imprecisely estimated, since it is difficult to identify whether little inefficiency is due to $\pi(\cdot)$ is close to 1 or σ_u^2 is close to zero. However, this identification issue is more relevant to the testing problem of all firms are efficient (or inefficient). We will return to the discussion of this hypothesis testing as well as other hypothesis testing problems in the later section. For the present discussion, we will assume that $\sigma_u^2 > 0$ and $0 < \pi(\cdot) < 1$ so that all the parameters in model (1) are identified.

To complete the specification of the model, let $f(z)$ and $f(y | x, z)$ denote the marginal density of z and the conditional density of y given x and z , respectively. In addition, we assume

throughout the paper that $f(y | x, z)$ is known and belongs to a class of parametric densities with parameter $\theta \in \Theta \subset \mathfrak{R}^k$ where k a positive integer and the function $\pi(z) : \mathfrak{R}^q \rightarrow [0, 1]$ is a smooth function which is twice continuously differentiable.

3. Estimation

3.1 Backfitting Local Maximum Likelihood Procedure

To make specific assumption regarding the conditional distribution of $f(y | x, z)$, we follow the standard practice and assume that $v_i | x, z \sim i.i.d. N(0, \sigma_v^2)$ and $u_i | x, z \sim i.i.d. N(0, \sigma_u^2)$, albeit other distributions such as exponential, truncated normal or gamma can also be considered for u_i . The conditional probability density function of $\varepsilon_i = v_i - u_i$ is given by

$$f(\varepsilon | x, z) = \left(\frac{\pi(z)}{\sigma_v} \right) \phi \left(\frac{\varepsilon}{\sigma_v} \right) + (1 - \pi(z)) \left[\frac{2}{\sigma} \phi \left(\frac{\varepsilon}{\sigma} \right) \Phi \left(-\varepsilon \frac{\lambda}{\sigma} \right) \right], \quad (2)$$

where $\sigma^2 = \sigma_u^2 + \sigma_v^2$, $\lambda = \sigma_u / \sigma_v$, $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density (pdf) and cumulative distribution functions (CDF) of a standard normal variable, respectively. To avoid the non-negativity restrictions we make use of the following transformation: $\lambda = \exp(\bar{\lambda}) = \tilde{\lambda}$ and $\sigma^2 = \exp(\bar{\sigma}^2) = \tilde{\sigma}^2$. Let $\theta = (\beta', \tilde{\lambda}, \tilde{\sigma}^2)'$ then it follows that the conditional pdf of y given x and z is

$$f(y | x, z) = \left(\frac{\pi(z)}{\tilde{\sigma}_v} \right) \phi \left(\frac{y - x'\beta}{\tilde{\sigma}_v} \right) + (1 - \pi(z)) \left[\frac{2}{\tilde{\sigma}} \phi \left(\frac{y - x'\beta}{\tilde{\sigma}} \right) \Phi \left(-(y - x'\beta) \frac{\tilde{\lambda}}{\tilde{\sigma}} \right) \right], \quad (3)$$

and conditional log-likelihood is then given by

$$L^*(\pi(z_i), \theta) = \sum_{i=1}^n \log f(y_i; \pi(z_i), \theta | x, z). \quad (4)$$

From (4), it is clear that if $\pi(z)$ is known and belong a class of parametric function with finite dimensional parameter vector, then standard maximum likelihood (ML) estimator can be

obtained by maxing (4) as discussed in KPT. However, $\pi(z)$ is generally unknown in practice rendering the standard MLE infeasible. To make the MLE operational, we approximate the unknown function $\pi(z)$ locally by a linear function, albeit in practice, one might wish to consider higher orders of local polynomials for $\pi(z)$. For a given value of z_0 , and z in the neighborhood of z_0 , a Taylor series expansion of $\pi(z_i)$ at z_0 gives

$$\pi(z_i) \approx \pi(z_0) + \pi'(z_0)(z_i - z_0) = a(z_0) + b(z_0)(z_i - z_0),$$

where $\pi'(z_0)$ is the derivative of $\pi(z_i)$ evaluated at z_0 . Then, the conditional *local* log-likelihood function associated with (4) can be written as

$$L_1(a(z_0), b(z_0), \theta) = \sum_{i=1}^n \{\log f(y_i; a(z_0), b(z_0), \theta | x, z)\} K_h(z_i - z_0), \quad (5)$$

where $K_h(\xi) = h^{-1}K(\xi / h)$ and $K(\cdot)$ is a kernel function and h is the appropriate bandwidth. Thus the conditional local log-likelihood depends on z . However, since the parameter θ does not depend on z , we suggest the following backfitting procedure which motivated by Huang and Yao (2012) for estimating semiparametric mixture regression models. Specifically, for a given value of z_0 , we first estimate $\pi(\cdot)$ locally by maximizing (5) with respect to a , b and θ . Let \tilde{a} , \tilde{b} and $\tilde{\theta}$ be the solution to the maximization problem of (5), that is¹

$$(\tilde{a}, \tilde{b}, \tilde{\theta}) = \arg \max_{(a, b, \theta)} L_1(a, b, \theta).$$

Then $\tilde{\pi}(z_0) = \tilde{a}(z_0)$ and $\tilde{\theta}(z_0) = \tilde{\theta}$. Now note that the global parameter vector θ do not depend on z and since θ is estimated locally, it does not possess the usual parametric \sqrt{n} -consistency. To preserve the \sqrt{n} -consistency and to improve the efficiency, given the estimate of $\tilde{\pi}(z_0)$, the

¹ In practice, the estimation is performed at a set of given z_0 values. A simple approach is to set $z_0 = z_1, \dots, z_0 = z_n$ which yields a set of $\tilde{\pi}(z_0)$ values.

parameter vector θ can be estimated globally by maximizing the following (global) log-likelihood function where we replace $\pi(z_0)$ with its estimate $\tilde{\pi}(z_0)$ in (4)

$$L_2(\theta) = \sum_{i=1}^n \log f(y_i; \tilde{\pi}(z_i), \theta | x, z). \quad (6)$$

Let $\hat{\theta}$ be the solution of maximizing (6). In the next section, we will show that, under certain regularity conditions, $\hat{\theta}$ retains its \sqrt{n} -consistency property. Given the estimates of $\hat{\theta}$ and to improve efficiency, the function $\pi(z_0)$ can be obtained by maximizing the local likelihood function

$$L_3(a(z_0), b(z_0), \hat{\theta}) = \sum_{i=1}^n \{\log f(y_i; a(z_0), b(z_0), \hat{\theta} | x, z_0)\} K_h(z_i - z_0). \quad (7)$$

Let $\hat{\pi}(\cdot) = \hat{a}(\cdot)$ be the solution of maximizing (7). Finally, $\hat{\theta}$ and $\hat{\pi}(z)$ can further be improved by iterating until convergence. We will denote the final $\hat{\pi}(z)$ and $\hat{\theta}$ as iterative backfitting local MLE.

We summarize the above backfitting local ML estimation procedure with the following computational algorithm:

Step 1: For each z_i , $i = 1, \dots, n$, in the sample, maximize the conditional *local* log-likelihood (5) to obtain the estimate of $\tilde{\pi}(z_i)$. Note that if the sample size n is large, (5) could be performed on a random subsample N_s where $N_s \ll n$ to reduce the computational burden. Also, to ensure that the estimates of $\pi(\cdot)$ fall within the interval $[0, 1]$, we reparameterizing the local linear parameters using logistic function.

Step 2: From step 1, conditional on $\tilde{\pi}(z_i)$, maximize the conditional *global* log-likelihood function (6) to obtain $\hat{\theta}$.

Step 3: Conditional on $\hat{\theta}$ from step 2, maximize the conditional local log-likelihood function (7) to obtain $\hat{\pi}(z_i)$.

Step 4: Using $\hat{\pi}(z_i)$ repeat step 2 and then step 3 until the estimate of $\hat{\theta}$ converges.

Remark: First, in practice, one could stop at step 3 to reduce the computational burden. However, iteration between step 2 and step 3 until convergence is highly recommended. Based on our limited experience, convergence is typically fast as it requires only a few iterations. Second, Step 1 requires specifications of the kernel function $K_h(\cdot)$ as well as bandwidth h . For the kernel function, an Epanechnikov or Gaussian function is a popular choice. As for the bandwidth selection, data driven methods such as cross-validation (CV) can be used (see for example Li and Racine (2007)). In our context, we use a likelihood version of CV which is given by

$$CV(h) = \frac{1}{n} \sum_{i=1}^n \log f(y_i; \hat{\pi}^{(i)}(z_i), \hat{\theta}^{(i)} | x, z), \quad (8)$$

where $\hat{\pi}^{(i)}(z_i)$ and $\hat{\theta}^{(i)}$ are the leave-one-out version of the backfitting local MLE described above. Third, 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} 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, the iterative backfitting local MLE described in this section uses direct maximization of the log likelihood functions (5), (6) and (7). An alternative approach is to use EM algorithm procedure. The main advantage of EM algorithm is that it is numerically stable and possesses the ascent property in the sense that when the sample size is large enough, each iteration raises the likelihood value (Greene (2012), and Huang and Yao (2012)). However, the main drawback of

EM algorithm is that it requires extensive computation, especially for the model considers in this paper, and the convergence can be very slow. Huang and Yao (2012) give detail implementations of EM algorithms for normal mixture model which is very similar to our model. Interest readers are referred to their paper for more details.

3.2 Estimation of Firm-Specific Inefficiency:

Follow the discussion of KPT, 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 ε is

$$f(u | \varepsilon) = \begin{cases} 0 & \text{with probability } \pi(z) \\ N_+(\mu_*, \sigma_*^2) & \text{with probability } (1 - \pi(z)) \end{cases},$$

where $N_+(\cdot)$ denotes the truncated normal, $\mu_* = -\varepsilon\sigma_u^2 / \sigma^2$ and $\sigma_*^2 = \sigma_u^2\sigma_v^2 / \sigma^2$. Thus, the conditional mean of u given $\varepsilon = y - x'\beta$ is:

$$E(u | \varepsilon, z) = (1 - \pi(z)) \frac{\sigma\lambda}{1 + \lambda^2} \left[\frac{\phi(-\lambda\varepsilon / \sigma)}{\Phi(-\lambda\varepsilon / \sigma)} - \frac{\lambda\varepsilon}{\sigma} \right]. \quad (9)$$

A point estimator of individual inefficiency score could be obtained by replacing the unknown parameters in (9) by their estimates and ε by $\hat{\varepsilon}(x) = y - x'\hat{\beta}$.

The second approach is to use the modal estimator which defined as

$$M(u | \varepsilon, z) = \frac{df(u | \varepsilon, z)}{du} = 0, \quad (10)$$

and (10) is known to have a zero at the value of $u = \mu_*$ whenever $\varepsilon < 0$, and zero otherwise.

Hence, multiplying μ_* by $(1 - \pi(z))$ yields the modal estimator.

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

$$p_i^*(z) = \frac{(\hat{\pi}(z) / \hat{\sigma}_v) \phi(\hat{\varepsilon}_i \hat{\lambda}_i / \hat{\sigma}_v)}{(\hat{\pi}(z) / \hat{\sigma}_v) \phi(\hat{\varepsilon}_i \hat{\lambda}_i / \hat{\sigma}_v) + (1 - \hat{\pi}(z)) (2 / \hat{\sigma}) \phi(\hat{\varepsilon}_i \hat{\lambda}_i / \hat{\sigma}) \Phi(-\hat{\varepsilon}_i \hat{\lambda}_i / \hat{\sigma})}. \quad (11)$$

Then the posterior estimate of inefficiency can be defined as $\tilde{u}_i = (1 - p_i^*(z)) \hat{u}_i$ where \hat{u}_i is the estimated of inefficiency based on (9) or (10). KPT provide an intuitive explanation for why the estimator given in (11) would be particularly helpful for researchers and regulators in the merger case to determine the probability of a specific firm or a group firms in the industry is to being fully efficient.

4. Asymptotic Properties

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

Let $\gamma(z) = (\pi(z), \theta)'$ and $\ell(\gamma(z), x, y) = \log f(y | x, \gamma(z))$. Define $q_\gamma(\gamma(z), x, y) = \frac{\partial \ell(\gamma(z), x, y)}{\partial \gamma}$, $q_{\gamma\gamma}(\gamma(z), x, y) = \frac{\partial^2 \ell(\gamma(z), x, y)}{\partial \gamma \partial \gamma'}$ and the terms q_θ , q_π , $q_{\theta\theta}$, $q_{\theta\pi}$ and $q_{\pi\pi}$ can be defined similarly.

In addition, let $\psi(w | z) = E[q_\pi(\gamma(z), x, y) | z = w]$,

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

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

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

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

Finally, let $\mu_j = \int u^j K(u) du$ and $\kappa_j = \int u^j K^2(u) du$. We make the following assumptions:

Assumption 1: The sample $\{(x_i, y_i, z_i), i = 1, \dots, n\}$ is independently and identically distributed from the joint density $f(x, y, z)$ which has continuous first derivative and positive in its support. The support for z , denoted by \mathbb{Z} , is a compact subset of \mathfrak{R} and $f(z) > 0$ for all $z \in \mathbb{Z}$.

Assumption 2: The unknown function $\pi(z)$ is twice continuously differentiable in its argument. Furthermore, $\pi(z) > 0$ hold for all $z \in \mathbb{Z}$.

Assumption 3: The matrix $I_{\gamma\gamma}(z)$ and $I_{\theta\theta}$ are positive definite.

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

Assumption 5: For some $\zeta < 1 - r^{-1}$, $n^{2\zeta-1}h \rightarrow \infty$ and $E(z^{2r}) < \infty$.

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, $nh^4 \rightarrow 0$ and $nh^2 \log(1/h) \rightarrow \infty$, we have

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

where $A = E\{I_{\theta\theta}(z)\}$ and $\Sigma = Var\left\{\frac{\partial \ell(\pi(z), \theta, x, y)}{\partial \theta} - I_{\theta\pi}(z)d(x, y, z)\right\}$ with $d(x, y, z)$ is the first element of $I_{\gamma\gamma}^{-1}(z)q_{\gamma}(\gamma(z), x, y)$.

Theorem 2: Under Assumptions 1-5 and in addition, as $n \rightarrow \infty$, $h \rightarrow 0$ and, $nh \rightarrow \infty$ we have

$$\sqrt{nh}\{\hat{\pi}(z) - \pi(z) - B(z) + o_p(h^2)\} \xrightarrow{D} N\{0, \kappa_0 f^{-1}(z)I_{\pi\pi}^{-1}\},$$

where $B(z) = \frac{1}{2}\mu_2 h^2 I_{\pi\pi}^{-1}(z)\psi''(z | z)$.

The proofs of Theorem 1 and 2 are given in Appendix A. Note that, the result from Theorem 2 shows that, as for common semiparametric model, the estimate of θ has no effect on the first-order asymptotic since the rate of convergence of $\hat{\pi}(z)$ is slower than that of \sqrt{n} . Consequently, it is fairly straightforward to see that $\hat{\pi}(z)$ is more efficient than the initial estimate of $\tilde{\pi}(z)$.

5. Hypothesis Testing

Given the structure of model (1), it is of great interest to ask whether the probability of a firm being efficient takes a specific parametric form such as those suggested in KPT or RS. This question leads to the following hypothesis testing problem:

$$H_0 : \pi(z_i) = h(z_i, \delta), \quad (12)$$

where $h(z_i, \delta)$ is a specific parametric function and δ is a vector of unknown parameters. For example, as in KPT and RS, one can assume $h(z_i, \delta) = \exp(z_i' \delta) / [1 + \exp(z_i' \delta)]$ or $h(z_i, \delta) = \Phi(z_i' \delta)$ where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variate. Under the null hypothesis, model (1) reduces to the parametric zero-inefficiency stochastic frontier model considered by KPT and RS. However, under the alternative hypothesis, model (1) is a semiparametric model and hence the number of parameters under the alternative is undefined. One useful approach to test for the above null hypothesis is to use sieve likelihood ratio (hereafter SLR) statistics suggested by Fan et al. (2001), and it is given by:

$$T = 2\{L^*(H_1) - L^*(H_0)\}, \quad (13)$$

where $L^*(H_0)$ and $L^*(H_1)$ denote the log-likelihood function computed under the null and the alternative hypothesis, respectively. Fan et al. (2001) show that the SLR statistics are asymptotically distribution free of nuisance parameters and follow $\chi_{b_n}^2$ distributions (for a sequence $b_n \rightarrow \infty$) under the null hypothesis (i.e., Wilks phenomenon) for testing a number of useful hypotheses for a variety of useful models such as nonparametric regression, varying

coefficient and generalized varying coefficient models. However, since model (1) belongs to the class of semiparametric mixture models, the asymptotic null distribution of the SLR may or may not follow χ_b^2 distribution. Thus, one approach is to derive the asymptotic distribution of T . Alternatively, we can use the conditional bootstrap procedure suggested by Cai et al. (2000) to approximate the asymptotic null distribution. The conditional bootstrap can be conducted as follows. Let $\{\bar{\beta}, \bar{\sigma}^2, \bar{\lambda}, \bar{\delta}\}$ be the MLE under the null hypothesis. Given x_i , generate a bootstrap sample, y_i^* from a given distribution of y specified in (1) with $\{\pi(\cdot), \beta, \sigma^2, \lambda\}$ are replaced by their MLE estimates $\{\bar{\beta}, \bar{\sigma}^2, \bar{\lambda}, \bar{\delta}\}$. For each bootstrap sample, calculate the test statistic T^* in (13), and use the distribution T^* as an approximation to the distribution of T .

It is important to note that, the conditional bootstrap described above is valid only if the asymptotic null distribution is independent of nuisance parameter $\pi(\cdot)$ (i.e., Wilk's phenomenon). We investigate the Wilk's phenomenon via Monte Carlo simulation below. Our simulation results indicate that, indeed Wilk's type of phenomenon continue to hold for the model consider in this paper.

Another interesting question that arises is whether all firms are inefficient. This question leads to the following testing hypothesis: $H_0 : \pi(z) = 0$ for all z . Under null hypothesis of $H_0 : \pi(z) = 0$, model (1) reduces to a standard stochastic frontier and this is simply a special case of the testing problem of constancy of $\pi(\cdot)$ which take on a specific value of 0. Thus, in principle, a simple modification of sieve likelihood ratio statistics in (13) can be used to test the null. However, since the value of 0 lies on the boundary of the parameter space of π , the asymptotic null distribution of the test statistics is no longer a χ^2 distribution. Thus, one approach is to derive the asymptotic distribution of the test statistics under this null hypothesis along the line of Andrews (2001), which is very complicated, given the semiparametric nature of the alternative. In addition, it is beyond the scope of this paper and we will leave it for future research.

Alternatively, since the null hypothesis of $H_0 : \pi(z) = 0$ is a special case of $H_0 : \pi(z) = \pi$, the conditional bootstrap described earlier can be used to approximate the asymptotic distribution of the test statistics, provided that the Wilk's type of phenomenon

continues to hold. Our Monte Carlo results below show that the test which is based on the conditional bootstrap has approximately correct sizes.

Finally, note that we did not pursue the hypothesis testing problem of $H_0 : \pi(z) = 1$ for all z (i.e., all firms are efficient) simply because under this null hypothesis, there is a technical problem related to the fact that σ_u^2 is not identified, which invalidates the conditional bootstrap procedure, albeit the SLR test would remain valid. In this case, there is a need for deriving the asymptotic distribution of the test statistics, and this is beyond the scope of this paper. Further investigation for this case would be interesting and useful for future research.

6. Monte Carlo Simulations

In this section, we conduct some simulations to study the finite sample performance of the proposed estimator and test statistics. To this end, we consider the following data generating process (DGP):

$$y_i = \begin{cases} 1 + x_i + v_i & \text{with probability } \pi(z_i) \\ 1 + x_i + v_i - u_i & \text{with probability } 1 - \pi(z_i) \end{cases},$$

where $\pi(z_i) = 0.05 + 0.6 \sin(\pi z_i)$. We generate z_i from a uniform distribution on $[0, 1]$ and the x_i is generated from a $N(0, 1)$. The random error term v_i is generated as $N(0, 0.5)$ and the one-sided error u_i is generated as $|N(0, 0.5\lambda)|$. For all of our simulations, we set $\lambda = \{1, 2.5, 5\}$, and let the sample sizes vary over $n = 2500$ or $n = 5000$. For each experimental design, 1,000 replications are performed.

We use the Gaussian kernel function and the bandwidth is chosen according to $\tilde{h} = \hat{h} \times n^{-2/15}$ where \hat{h} is the optimal bandwidth based on CV approach previously discussed in Section 3.1.

We measure the performance of the estimate of the probability of firms being fully efficient function $\pi(z)$ by computing the mean average squared errors (MASE):

$$MASE = \frac{1}{1000} \sum_{r=1}^{1000} \left\{ \frac{1}{n} \sum_{j=1}^n [\hat{\pi}_r(z_j) - \pi_r(z_j)]^2 \right\}.$$

The performance of the estimates of the production parameters is measured by the mean squared errors (MSE)

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

where $\hat{\theta} = \hat{\beta}$, $\hat{\sigma}^2$ or $\hat{\lambda}$. The simulations were performed on the mainframe using FORTRAN 77 using G77 compiler of GNU.

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

Table 2 reports the empirical sizes of the bootstrap SLR statistics. From Table 2, we see that, there are little sizes distortions indicating the conditional bootstrap provides a good approximation for the asymptotic distribution of the SLR statistics.

Table 3 summarizes the performance of the bootstrap approach for standard errors of estimate of parameters for two different samples, and three different bandwidths which correspond to under-smoothing ($\tilde{h} = \hat{h} \times n^{-2/15}$), appropriate amount (\hat{h}) and over-smoothing ($2\hat{h}$). 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 SE 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.

Note that the bootstrap procedure also allows us to compute the point-wise coverage probabilities for the probability functions. Table 4 provides the 95% coverage probabilities of

$\pi(z)$ for a set of evenly spaced grid points distributed on the support of z . In the table, the row labelled with $\pi(\hat{\theta})$ gives the results using the proposed approach, whilst $\pi(\theta)$ provides the results assuming θ were known. For most cases, the coverage probabilities are close to the nominal level, especially when under-smoothing or appropriate smoothing is used. For the case of over-smoothing, the results are somewhat less satisfactory. Moreover, the coverage levels are slightly low for point 0.4 and slightly high for points 0.8 and 0.9

Next, we investigate whether Wilk's type of phenomenon holds for the proposed model. Under the null hypothesis of (12), we assume the probability of efficient firm takes a specific parametric form. The DGP is the same as above except now we generate $\pi(z)$ as logistic function or standard normal CDF with parameter δ . For each function, we fixed the value of $\lambda = 2.5$ and set 3 different values of $\delta = \{-1, 0, 1\}$, and use nonparametric kernel density estimation to compute the unconditional (asymptotic) null distribution of SLR statistics with $n = 2500$ and $h = 0.06$ via 500 replications². The resulting densities are plotted in solid lines in Figures 1.a to 1.c. As can be seen from these plots, the resulting densities are very close indicating that the asymptotic distribution of the SLR statistics are not sensitive to the choice of function $\pi(z)$. This suggests that Wilk's type of results continue to hold for our model.

Finally, to validate the conditional bootstrap approach, for each assumed function, we select 3 typical samples generated from the 3 different values of δ , and compute the conditional null distribution based on its 500 bootstrap samples. The resulting densities are depicted as dotted curves in the same Figures 1.a-1.c. From these figures, we can observe that the proposed conditional bootstrap approach performed quite well to approximate the asymptotic null distribution.

7. Empirical Application

There exists a vast literature on measuring productivity and efficiency for the banking sectors in various countries (see for example, Tzeremes (2015), Galán et al. (2015), Sathye (2003) and the articles in Volume 98, Issue 2 (1997) of the *European Journal of Operational Research*, just to name a few). However, all these applications typically do not allow for the presence of fully

² We also conduct simulations using other bandwidths $h = 0.12$ and $h = 0.24$. The results are very similar, and hence we do not report them here but available from the authors upon request.

efficiency banks, and hence these results could potentially be misleading if in fact there are efficient banks in the sample.

In this section, we provide an application of the U.S. banking sectors to illustrate the usefulness and merit of our proposed model and approach. The data we use are taken from Koetter et al. (2012) which consist of large number of individual U.S. commercial banks from Reports of Condition and Income of the Federal Reserve System³. The data contain annual year-end from all U.S. insured banks between 1976 and 2007. After controlling for outliers and missing observations, the final sample use in the estimation consists of 342,868 observations.

Following convention in the competition and efficiency literature, the regressors used in our model are logs of three input prices: price of fixed assets (w_1), cost of labor (w_2) and purchased funds costs (w_3), levels of two outputs: loans (y_1) and federal funds sold and securities purchased (y_2), a time trend (t) and the log of total assets to control for size effects (z). In addition, to be in line with the intermediation approach, it is assumed that banks transform various saving of consumers and firms into loans and investment, and seek to minimize costs. Thus, the dependent variable is total operating costs implying a cost frontier approach is employed.

Note that our proposed model and approach is designed for cross-section data, and since we are using panel data, we need to make some assumptions regarding the temporal behavior of the technical inefficiency and random noise. Following KPT, first we include a time variable in the SF function to allow for technical change or shift in the frontier; and second, for simplicity we assume that both u and v are independently and identically distributed.

We employ the translog specification for the cost frontier which can be written as:

$$\begin{aligned} \ln C_{it} = & \beta_0 + \sum_{j=1}^2 \beta_j \ln y_{j,it} + \sum_{j=1}^2 \gamma_j \ln w_{j,it} + \frac{1}{2} \sum_{l=1}^2 \sum_{k=1}^2 \beta_{lk} \ln y_{l,it} \ln y_{k,it} + \frac{1}{2} \sum_{l=1}^2 \sum_{k=1}^2 \gamma_{lk} \ln w_{l,it} \ln w_{k,it} \\ & + \sum_{l=1}^2 \sum_{k=1}^2 \alpha_{lk} \ln y_{l,it} \ln w_{k,it} + \alpha_t t + \alpha_{tt} t^2 + \sum_{l=1}^2 \delta_{lk} t \ln y_{l,it} + \sum_{l=1}^2 \theta_{lk} t \ln w_{l,it} + v_{it} + u_{it}, \end{aligned}$$

³ See Koetter et al. (2012) for the issues involved as well as details construction of the data set. The data are available also from Restrepo-Tobon and Kumbhakar (2014).

where we assume $u_{it} \sim i.i.d. |N(0, \sigma_u^2)|$ and $v_{it} \sim i.i.d. N(0, \sigma_v^2)$. We impose the usual symmetry conditions in which we assume $\beta_{lk} = \beta_{kl}$ and $\alpha_{lk} = \alpha_{kl}$. In addition, we normalized the cost and input prices by one input price (here we use w_3) to ensure that the linear homogeneous restriction of the cost function with respect to input prices hold.

For comparison purpose, we also estimate a half-normal standard scholastic frontier model (HN-SFM) and a zero-inefficiency stochastic frontier (ZISF) model of KPT, which assumes the underlying probability of firms being fully efficient follows the logit specification of z_i .

Note that, since the estimated parameters of the translog frontier do not have any direct economic interpretation, for brevity, we do not report all the parameter estimates here but these are available from the authors upon request. Instead, we summarize the results for the estimated returns to scale (RTS), technical change (TC) as well as report the results that associated with the estimated probability function and the estimated technical inefficiencies.

For a cost function, RTS measures the proportional increase in costs due to an increase in all outputs, that is, RTS can be defined as the reciprocal of $\sum_j (\partial \ln C / \partial y_j)$. Thus, if RTS is larger than one then a proportional increase in all outputs will lead to a less than proportional increase in cost, implying that the scale operation is below optimum, and hence there are benefits from expansion (i.e., economies of scale). The opposite holds true when RTS is less than one. For TC, it is defined as the rate of change in cost over time, *ceteris paribus*, i.e., $\partial \ln C / \partial t$. Therefore, a negative value of TC suggests a reduction in cost overtime, implying technical progress, and a positive value of TC shows a technical regress, *ceteris paribus*.

As previously mentioned in the Introduction, it is important to recognize that the prominent feature of the ZISF model is that the frontier itself does not vary across the two classes of firms but only the existence or non-existence of inefficiency differs. Thus, we would expect that the estimated RTS and TC of the three models would not differ significantly. Indeed, our results indicated that the estimated RTS and TC are very similar across all models. For RTS, the estimated values ranging from 0.85 to 1.3 with the mean value of 1.11 and standard deviation of 0.22; while the estimated values of TC ranging from -0.091 to 0.019 with the mean value of -0.016 and standard deviation of 0.0068. These results indicate that most of the banks experienced economies of scale as well as technical progress.

We now turn our attention to the results of the estimated probability function, and the estimated technical efficiencies. To present results for the probability function, we normalize the log of total assets as $z_{it}^* = (z_{it} - z_{\max}) / (z_{\max} - z_{\min})$ where z_{it} is the log of total assets. The function is then evaluated at 100 points between 0 and 1, and presented (along with two standard error bands) in Figure 2. From Figure 2, we observe that banks attaining full efficiency are, most likely, concentrated near the 20% asset quantile where the probability peaks at about 80% while 95% confidence intervals indicate that the probability can be as high as one. Larger banks seem to be inefficient with large probability, although there is some evidence of marginal behavior near the 79-85% quantiles. Zero inefficiency seems to prevail for most banks roughly below the median (normalized) assets. These results are consistent with the hypothesis that larger banks are inefficient due, for example, to the “quiet life hypothesis” (Koetter and Vins (2008) and Koetter et al. (2012)), albeit other reasons could also be responsible for inefficient larger banks and nearly efficient smaller banks.

The estimated technical inefficiency distributions are displayed in Figure 3. It can be seen from Figure 3 that, the SFM based on the half-normal specification yields inefficiency results that display virtually no mass at zero, indicating that no banks are fully efficient. The inefficiency scores lie in the range of 0.5% to 14% with the mean value of 6.5% and the standard deviation of 0.021. In contrast, results from ZISF indicate that there is some mass at zero with a long right tail in the inefficiency distribution. This suggests that, albeit there are some fully efficiency banks, inefficiency can be as high as 20% for some other banks. The inefficiency scores lie in a wider range than in the case of half-normal SFM, ranging from 0% to 20% with the mean value of 4.5% and the standard deviation of 0.032. The semi-parametric specification places even more mass at zero, and the inefficiency distribution is much tighter than both half-normal SFM and ZISF. The inefficiency scores lies between 0% to 10%, with the mean value of 2.3% and the standard deviation of 0.012. Thus, from the results in Figure 3, we can see that parametric models (HN-SFM or ZISF) deliver very different inefficiency distributions compared to the semi-parametric specification.

To determine which specification is more appropriate for the data considered, we use the SLR test discussed in Section 5 to test for the hypotheses of (i) $H_0 : \pi(z) = e^{\alpha z} / (1 + e^{\alpha z})$ (ZISF model) and (ii) $H_0 : \pi(z) = 0$ (HN-SFM), based on the conditional bootstrap critical

values. Our SLR tests produce the conditional bootstrap p -values of 0.0236 for testing (i) and 0.0073 for testing (ii), suggesting that both null hypotheses are rejected at a 5% significant level in favor of the nonparametric specification of the probability function. Consequently, for this particular data set, our results provide evidence that a flexible specification of the probability function is critical and, in particular, material in terms of inefficiency estimation.

8. Concluding Remarks

In this paper, we propose semiparametric approach for estimating the zero-inefficiency stochastic frontier model (e.g., KPT (2013) and RS (2015)) by allowing for the proportion of firms that are fully efficient to depend on a set of covariates via unknown smooth function. In particular, we propose an iterative backfitting local maximum likelihood estimation procedure that achieves the optimal convergence rates of both frontier parameters and the nonparametric function of the probability of firms being efficient. We derive the asymptotic bias and variance of the proposed estimator and establish its asymptotic normality. In addition, we discuss how to test for parametric specification of the proportion of firms that are fully efficient as well as how to test for the presence of fully inefficient firms, based on the conditional bootstrap sieve likelihood ratio statistics. The finite sample behaviors of the proposed estimation procedure and tests are examined using Monte Carlo simulations. We apply the proposed method to data on a large number of individual U.S. commercial banks to examine the effects of total assets on the probability of banks being efficient as well as technical inefficiency measurements overall. Our analysis indicated that flexible specification of the probability function of banks being efficient is critical in efficiency estimation.

Note that the estimation approach proposed in this paper can also be easily modified and extended to other models as well that allow for the distribution of u_i to depend on covariate z_i either parametrically (e.g., Reifschneider and Stevenson (1991), Caudill and Ford (1993) and Caudill et al. (1995)) or nonparametrically. For example, if it is assumed that $\sigma^2(z_i) = \exp(z_i' \alpha)$, then by simply redefining the finite dimensional parameter vector θ , the estimation algorithm proceeds as discussed. On the hand, if we assume $\sigma^2(z_i)$ to be an unknown smoothing function,

then by approximating this function locally at a point z_0 and by modifying the local log likelihood function in (4), the estimation algorithm remains unaffected.

Finally, it would be interesting to extend the current model to full nonparametric setting that includes both continuous and categorical variables in the frontier as well as in the probability function.

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.
- Andrews, D.W.K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69 (3), 683-734.
- Cai, Z., Fan, J., & Li, R. (2000). Efficient estimation and inference for varying-coefficient models. *Journal of American Statistical Association*, 95, 888-902.
- Caudill, S.B., & Ford, J.M. (1993). Biases in frontier estimation due to heteroskedasticity. *Economics Letters*, 41, 17-20.
- Caudill, S.B., Ford, J.M., & Gropper, D.M. (1993). Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. *Journal of Business Economic Statistics*, 13, 105-111.
- Fan, J., & Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.
- Fan, J., & Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, 11, 1031-1057.
- Fan, J., Zhang C., & Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Annals of Statistics*, 29, 153-193.
- Galán, J.E., Veiga, H. and Wiper, M.P. (2015). Dynamics effects in efficiency: Evidence from Colombian banking sector. *European Journal of Operational and Research*, 240, 562-571.
- Greene, W. (2012). *Econometric Analysis*. (5th ed.). New York: Prentice Hall.
- Huang, M., & Yao, W. (2012). Mixture of regression model with varying mixing proportions: a semiparametric approach. *Journal of American Statistical Association*, 107, 711-724.
- Ivaldi, M., Monier-Dilhan, S. and Simioni, M. (1995). Stochastic production frontiers and panel data: A latent variable framework. *European Journal of Operational Research*, 80(3), 534-547.
- Jondrow, J., Lovell, C.A.K., Materov, I.S., & Schmidt P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19 (2/3), 233-238.
- Koetter, M., & Vins, O. (2008). The quiet life hypothesis in banking: evidence from German savings banks. *Working paper*.

- Koetter, M., Kolari, J., & Spierdijk, L. (2012). Enjoying the quiet life under deregulation?: Evidence from adjusted Lerner indices for U.S. banks. *Review of Economics and Statistics*, 94(2), 462-480.
- Kumbhakar, S., Parmeter, C.F., & Tsionas, E.G. (2013). A zero inefficiency stochastic frontier model. *Journal of Econometrics*, 172, 66-76.
- 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.
- Meeusen, W., & van den Broeck, J. (1997). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18 (2), 435-444.
- Ondrich, J. and Ruggiero, J. (2001). Efficiency measurement and stochastic frontier model. *European Journal of Operational and Research*, 129, 432-442.
- Restrepo-Tobon, D., & Kumbhakar, S.C. (2014). Enjoying the quiet life under deregulation? Not quite. *Journal of Applied Econometrics*, 29 (2), 333-343.
- Reifschneider, D., & Stevenson, R. (1991). Systematic departures from the Frontier: a framework for the analysis of firm inefficiency. *International Economic Review*, 32, 715, 723.
- Rho, S., & Schmidt, P. (2015). Are all firms inefficient? *Journal of Productivity Analysis*, 43, 327-349.
- Sathye, M. (2003). Efficiency of banks in a developing economy: The case of India. *European Journal of Operational and Research*, 148, 662-671.
- Tzeremes, N. (2015). Efficiency dynamics in Indian banking: A conditional directional distance approach. *European Journal of Operational and Research*, 240, 807-818.

Appendix A: Proofs of the Theorems.

We first introduce some additional notations. Let $\tilde{\pi}^* = \sqrt{nh}(\tilde{\pi} - \pi)$, $\tilde{\beta}^* = \sqrt{nh}(\tilde{\beta} - \beta)$, $\tilde{\lambda}^* = \sqrt{nh}(\tilde{\lambda} - \lambda)$, $\tilde{\sigma}^{2*} = \sqrt{nh}(\tilde{\sigma}^2 - \sigma^2)$ where π , β , λ and σ^2 are the true values. Also, let $\tilde{\theta}^* = (\tilde{\beta}^{*'}, \tilde{\lambda}^*, \tilde{\sigma}^{2*})'$ and $\tilde{\gamma}^* = (\tilde{\pi}^*(\cdot), \tilde{\theta}^{*'})'$

Proof of Theorem 1: The proof of this theorem follows similarly to that of Huang and Yao (2012). We show the key steps of the proof.

To derive the asymptotic properties of $\hat{\theta}$, we first examine the asymptotic behavior of $\tilde{\gamma} = (\tilde{\pi}, \tilde{\theta}')$ which is the local MLE of (4). Let

$$\begin{aligned}\hat{\theta}^* &= \sqrt{n}(\hat{\theta} - \theta), \\ \ell(\tilde{\pi}(z_i), \theta, x_i, y_i) &= \log f(y_i | \tilde{\pi}(z_i), \theta) \\ \ell(\tilde{\pi}(z_i), \hat{\theta} + n^{-1/2}\tilde{\theta}, x_i, y_i) &= \log f(y_i | \tilde{\pi}(z_i), \hat{\theta} + n^{-1/2}\tilde{\theta})\end{aligned}$$

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

$$L_n(\theta^*) = \sum_{i=1}^n \{\ell(\tilde{\pi}(z_i), \hat{\theta} + n^{-1/2}\tilde{\theta}, x_i, y_i) - \ell(\tilde{\pi}(z_i), \theta, x_i, y_i)\} \quad (\text{A.1})$$

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

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

where

$$\begin{aligned}A_n &= n^{-1/2} \sum_{i=1}^n \frac{\partial \ell(\tilde{\pi}(z_i), \theta, x_i, y_i)}{\partial \theta} \\ B_n &= n^{-1/2} \sum_{i=1}^n \frac{\partial^2 \ell(\tilde{\pi}(z_i), \theta, x_i, y_i)}{\partial \theta \partial \theta'}\end{aligned}$$

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

$$\begin{aligned}
A_n &= n^{-1/2} \sum_{i=1}^n \frac{\partial \ell(\pi(z_i), \theta, x_i, y_i)}{\partial \theta} + n^{-1/2} \sum_{i=1}^n \frac{\partial^2 \ell(\pi(z_i), \theta, x_i, y_i)}{\partial \theta \partial \pi} [\tilde{\pi}(z_i) - \pi(z_i)] + O_p(n^{-1/2} \|\tilde{\pi}(\cdot) - \pi(\cdot)\|_\infty^2) \\
&= n^{-1/2} \sum_{i=1}^n \frac{\partial \ell(\pi(z_i), \theta, x_i, y_i)}{\partial \theta} + D_{1n} + O_p(n^{-1/2} \|\tilde{\pi}(\cdot) - \pi(\cdot)\|_\infty^2)
\end{aligned}$$

where the definition of D_{1n} should be apparent. Now, applying Lemma A.1 of Fan and Huang (2005), we have

$$\tilde{\gamma}(z_i) - \gamma(z_i) = n^{-1} f^{-1}(z_i) I_{\gamma\gamma}^{-1}(z_i) \sum_{j=1}^n \frac{\partial \ell(\gamma(z_j), x_j, y_j)}{\partial \gamma} K_h(z_j - z_i) + O_p(\delta_n)$$

where $\delta_n = n^{-1/2} h^{3/2} + (nh)^{-1} \sqrt{\log(1/h)}$. Under the condition $nh^2 / \log(1/h) \rightarrow \infty$, we have $O_p(n^{1/2} \delta_n) = o_p(1)$. Furthermore, since $\pi(z_i) - \pi(z_j) = O((z_i - z_j)^2)$ and $K(\cdot)$ is symmetric about 0, we have

$$\begin{aligned}
D_{1n} &= n^{-3/2} \sum_{j=1}^n \sum_{i=1}^n \frac{\partial^2 \ell(\pi(z_i), \theta, x_i, y_i)}{\partial \theta \partial \pi} f^{-1}(z_i) d(x_j, y_j, z_j) K_h(z_i - z_j) + O_p(n^{1/2} h^2) \\
&= D_{2n} + O_p(n^{1/2} h^2)
\end{aligned}$$

where $d(x_j, y_j, z_j)$ is the first element of $I_{\gamma\gamma}^{-1}(z_j) q_\gamma(\gamma(z_j), x_j, y_j)$ and the definition of D_{2n} should

be apparent. Let $D_{3n} = -n^{-1/2} \sum_{j=1}^n I_{\theta\pi}(z_j) d(x_j, y_j, z_j)$, then it can be shown that

$D_{2n} - D_{3n} \xrightarrow{p} 0$. Hence, under the condition $nh^4 \rightarrow 0$, we have

$$A_n = n^{-1/2} \sum_{i=1}^n \left\{ \frac{\partial \ell(\pi(z_i), \theta, x_i, y_i)}{\partial \theta} - I_{\theta\pi}(z_i) d(x_i, y_i, z_i) \right\} + o_p(1) \quad (\text{A.3})$$

For B_n , it can be shown that

$$B_n = -E[I_{\theta\theta}(x)] + o_p(1) = B + o_p(1) \quad (\text{A.4})$$

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

$$\hat{\theta}^* = B^{-1} A_n + o_p(1) \quad (\text{A.5})$$

if A_n is a sequence of stochastically bounded vectors. Consequently, the asymptotic normality of $\hat{\theta}^*$ 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

$$E(A_n) = n^{1/2} E \left\{ \frac{\partial \ell(\pi(z), \theta, x, y)}{\partial \theta} - I_{\theta\pi}(z) d(x, y, z) \right\} \quad (\text{A.6})$$

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 $E I_{\theta\pi}(z) d(x, y, z) = 0$. Thus $E(A_n) = 0$. The variance of A_n is $\text{Var}(A_n) = \text{Var} \left\{ \frac{\partial \ell(\pi(z), \theta, x, y)}{\partial \theta} - I_{\theta\pi}(z) d(x, y, z) \right\} = \Sigma$. By the Central Limit Theorem, we obtain the desired result. \square

Proof of Theorem 2: Recall that, given the estimate of $\hat{\theta}$, $\hat{\pi}(z)$ maximizes (5). Let $\eta(z_o, z) = a(z_o) + a'(z_o)(z - z_o)$ and $\pi^* = (nh)^{1/2} \{\pi - a(z_o), h(\pi' - a'(z_o))\}'$, then $\hat{\pi}^*$ maximizes

$$L_n^*(\pi^*) = \sum_{i=1}^n \{ \ell(\eta(z_o, z_i) + (nh)^{-1/2} \pi^{*'} w_i, \hat{\theta}, x_i, y_i) - \ell(\eta(z_o, z_i), \hat{\theta}, x_i, y_i) \} K_h(z_i - z_o)$$

where $w_i = (1, (z_i - z_o) / h)'$. Using Taylor expansion of $\ell(\cdot)$ and after some calculation, we have

$$L_n^*(\pi^*) = \hat{\Delta}_n' \pi^* + \frac{1}{2} \pi^{*'} \hat{\Gamma}_n \pi^* + o_p(1) \quad (\text{A.7})$$

where

$$\hat{\Delta}_n = (nh)^{-1/2} \sum_{i=1}^n \frac{\partial \ell(\eta(z_i, z_o), \hat{\theta}, x_i, y_i)}{\partial \eta} w_i K_h(z_i - z_o)$$

$$\hat{\Gamma}_n = (nh)^{-1} \sum_{i=1}^n \frac{\partial^2 \ell(\eta(z_i, z_o), \hat{\theta}, x_i, y_i)}{\partial \eta \partial \eta'} w_i w_i' K_h(z_i - z_o)$$

By the SLLN, Assumption 3 and Lemma A.1 of Fan and Huang (2005), it can be shown that

$$E(\hat{\Gamma}_n) \rightarrow -f(z_0) \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \otimes I_\pi = -\hat{\Gamma}$$

and $\text{var}\{(\Gamma_n)_{ij}\} = O((nh)^{-1})$ implying that $\hat{\Gamma}_n = -\hat{\Gamma} + o_p(1)$. Thus (A.7) can be written as

$$L_n^*(\pi^*) = \hat{\Delta}_n' \pi^* + \frac{1}{2} \pi^{*'} \hat{\Gamma} \pi^* + o_p(1) \quad (\text{A.8})$$

Using the quadratic approximation lemma, yields

$$\hat{\pi}^* = \hat{\Gamma}^{-1} \hat{\Delta}_n + o_p(1) \quad (\text{A.9})$$

if $\hat{\Delta}_n$ is a sequence of stochastically bounded random vectors. An expansion of $\hat{\Delta}_n$ leads to

$$\hat{\Delta}_n = (nh)^{-1/2} \sum_{i=1}^n \frac{\partial \ell(\eta(z_i, z_0), \theta, x_i, y_i)}{\partial \eta} w_i K_h(z_i - z_0) + G_n + o_p(1) = \Delta_n + G_n + o_p(1)$$

where

$$G_n = (nh)^{-1/2} \sum_{i=1}^n \frac{\partial^2 \ell(\eta(z_i, z_0), \theta, x_i, y_i)}{\partial \eta \partial \theta} w_i (\hat{\theta} - \theta) K_h(z_i - z_0)$$

Since $\sqrt{n}(\hat{\theta} - \theta) = O_p(1)$, it can be shown that $G_n = -h^{-1/2} I'_{\eta\theta}(z) f(z) = o_p(1)$ where $I_{\eta\theta}(z) = -E[q_{\eta\theta}(\eta(z), \theta, x, y) | z]$. Thus, (A.9) becomes

$$\hat{\pi}^* = \hat{\Gamma}^{-1} \Delta_n + o_p(1) \quad (\text{A.10})$$

The asymptotic normality of $\hat{\pi}^*$ follows from that of Δ_n so it suffices to calculate the mean and variance of Δ_n . Since $K(\cdot)$ is symmetric and bounded, we have

$$\begin{aligned} E(\Delta_n) &= n(nh)^{-1/2} E \left\{ \frac{\partial \ell(\eta(z, z_0), \theta, x, y)}{\partial \eta} w K_h(z - z_0) \right\} \\ &= \frac{(nh)^{1/2} f(z_0)}{2} \begin{pmatrix} \mu_2 \\ 0 \end{pmatrix} \otimes \hat{\Gamma}(z_0) \psi''(z_0) \{1 + o_p(1)\} \end{aligned}$$

and

$$\begin{aligned}
\text{Var}(\Delta_n) &= h^{-1} E \left\{ \left(\frac{\partial \ell(\eta(z, z_0), \theta, x, y)}{\partial \eta} \right)^2 w w' K_h^2(z - z_0) \right\} \\
&= f(z_0) \begin{pmatrix} \kappa_0 & \kappa_1 \\ \kappa_1 & \kappa_2 \end{pmatrix} \otimes \hat{\Gamma}(z_0) \{1 + o_p(1)\}
\end{aligned}$$

Let $e_1 = (1 \ 0)'$ denote a (2×1) unit vector, then

$$\{e_1' \text{Var}(\Delta_n) e_1\}^{-1} \{e_1' \Delta_n - e_1' E(\Delta_n)\} \xrightarrow{D} N(0, 1)$$

by using standard argument. \square

Table 1: MSE of $(\hat{\beta}, \hat{\sigma}^2, \hat{\lambda})$ and MASE of $\hat{\pi}(\cdot)$

$n = 2500, \lambda = 1.0$		$n = 5000, \lambda = 1.0$
MSE		MSE
β	0.0028	0.0014
σ^2	0.0057	0.0022
λ	0.1011	0.0051
MASE		MASE
$\pi(\cdot)$	0.1357	0.0072

$n = 2500, \lambda = 2.5$		$n = 5000, \lambda = 2.5$
MSE		MSE
β	0.0021	0.0011
σ^2	0.0051	0.0023
λ	0.0097	0.0044
MASE		MASE
$\pi(\cdot)$	0.1019	0.0059

$n = 2500, \lambda = 5.0$		$n = 5000, \lambda = 5.0$
MSE		MSE
β	0.0018	0.0010
σ^2	0.0049	0.0025
λ	0.0085	0.0044
MASE		MASE
$\pi(\cdot)$	0.0095	0.0044

Table 2: Empirical Sizes of the Bootstrap SLR Statistics
 $(\lambda = 2.5)$

$n = 2,500$				$n = 5,000$		
$H_0 : \pi(z) = e^z / (1 + e^z)$				$H_0 : \pi(z) = e^z / (1 + e^z)$		
	1%	5%	10%	1%	5%	10%
Emp. Sizes	0.0117	0.0493	0.1105	0.0101	0.0502	0.099
$n = 2,500$				$n = 5,000$		
$H_0 : \pi(z) = 0$				$H_0 : \pi(z) = 0$		
	1%	5%	10%	1%	5%	10%
Emp. Sizes	0.0112	0.0489	0.1112	0.0104	0.0503	0.108

Table 3: Standard Deviations, Standard Errors and Coverage Probabilities

Parameter	STD	SE(STD)	95% Coverage
<i>n</i> = 2500, <i>h</i> = 0.06			
β_1	0.027	0.029(0.006)	94.6%
σ^2	0.015	0.014(0.004)	95.2%
λ	0.022	0.021(0.002)	93.2%
<i>n</i> = 2500, <i>h</i> = 0.12			
β_1	0.028	0.027(0.006)	94.7%
σ^2	0.017	0.018(0.0004)	95.2%
λ	0.023	0.025(0.002)	93.5%
<i>n</i> = 2500, <i>h</i> = 0.24			
β_1	0.019	0.020(0.005)	94.8%
σ^2	0.009	0.008(0.004)	95.4%
λ	0.013	0.013(0.002)	93.5%
<i>n</i> = 5000, <i>h</i> = 0.05			
β_1	0.018	0.017(0.003)	95.0%
σ^2	0.009	0.010(0.002)	95.3%
λ	0.012	0.011(0.001)	95.0%
<i>n</i> = 5000, <i>h</i> = 0.10			
β_1	0.019	0.018(0.003)	94.7%
σ^2	0.010	0.010(0.002)	95.1%
λ	0.011	0.011(0.001)	94.2%
<i>n</i> = 5000, <i>h</i> = 0.20			
β_1	0.020	0.020(0.003)	93.1%
σ^2	0.011	0.010(0.003)	94.0%
λ	0.012	0.013(0.001)	92.2%

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

Table 4: The Pointwise Coverage Probabilities for $\pi(z)$

z	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$n = 2500, h = 0.06$									
$\pi(\hat{\theta})$	94.6%	94.7%	94.8%	90.9%	95.4%	95.2%	94.9%	98.9%	97.7%
$\pi(\theta)$	94.1%	94.3%	94.5%	90.7%	95.0%	95.0%	94.8%	98.0%	97.1%
$n = 2500, h = 0.12$									
$\pi(\hat{\theta})$	94.6%	94.8%	94.9%	94.9%	95.6%	95.6%	94.9%	98.9%	97.8%
$\pi(\theta)$	94.1%	94.3%	94.5%	94.7%	95.0%	95.0%	94.8%	98.0%	97.1%
$n = 2500, h = 0.24$									
$\pi(\hat{\theta})$	91.6%	94.8%	94.9%	90.1%	95.7%	95.7%	94.8%	98.8%	90.7%
$\pi(\theta)$	92.1%	94.3%	94.5%	90.5%	95.0%	95.0%	94.8%	96.0%	92.1%
$n = 5000, h = 0.05$									
$\pi(\hat{\theta})$	95.5%	95.7%	95.5%	92.4%	95.4%	95.1%	95.1%	98.6%	96.1%
$\pi(\theta)$	95.2%	95.3%	95.2%	92.1%	95.0%	95.0%	95.0%	98.1%	96.7%
$n = 5000, h = 0.10$									
$\pi(\hat{\theta})$	95.5%	95.7%	95.5%	95.4%	95.4%	95.1%	95.1%	98.0%	96.1%
$\pi(\theta)$	95.7%	95.8%	95.2%	95.1%	95.0%	95.0%	95.0%	97.8%	96.5%
$n = 5000, h = 0.20$									
$\pi(\hat{\theta})$	91.5%	95.7%	95.5%	89.9%	95.4%	95.1%	95.1%	98.5%	90.1%
$\pi(\theta)$	92.7%	95.8%	95.2%	91.1%	95.0%	95.0%	95.0%	98.2%	92.0%

Note: $\pi(\hat{\theta})$: when $\hat{\theta}$ is estimated and $\pi(\theta)$: when θ are assumed to be known.

Figure 1.a: Estimated densities of the null distributions of SLR statistics,
 $\delta = -1, \lambda = 2.5, h = 0.06$

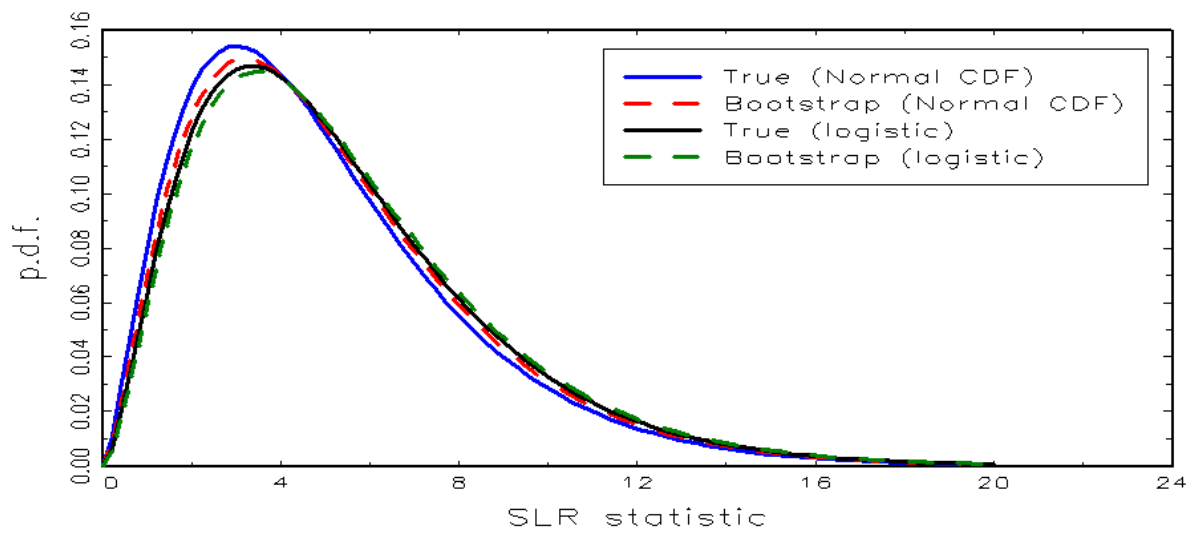


Figure 1.b: Estimated densities of the null distributions of SLR statistics,
 $\delta = 0, \lambda = 2.5, h = 0.06$

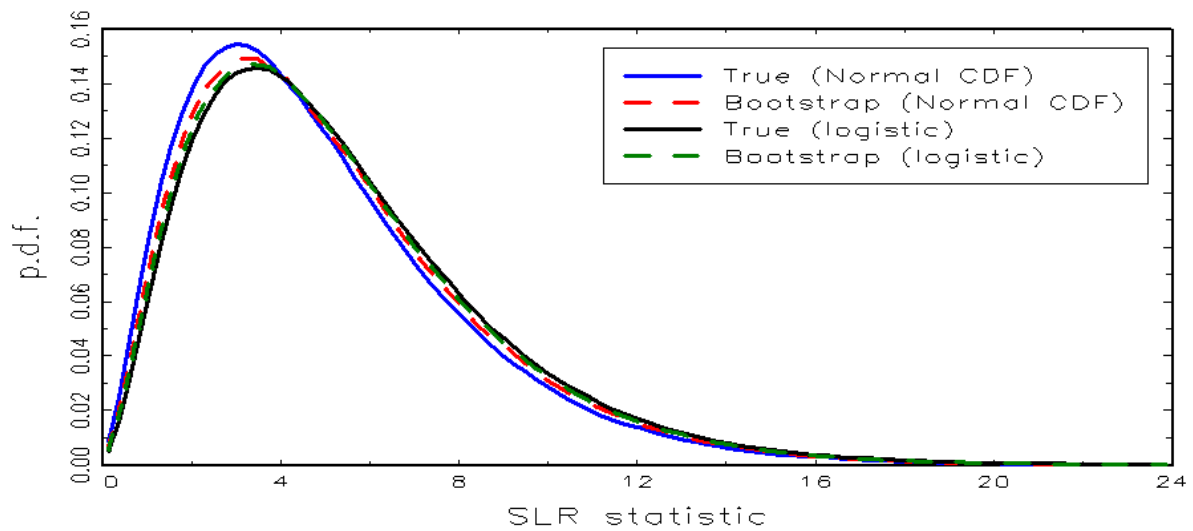


Figure 1.c: Estimated densities of the null distributions of SLR statistics,
 $\delta = 1, \lambda = 2.5, h = 0.06$

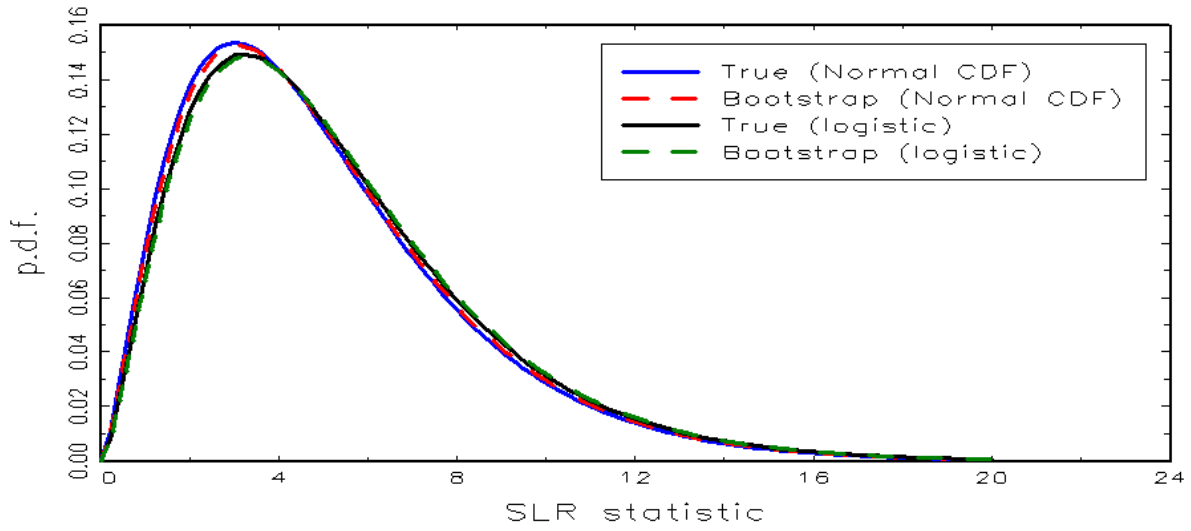


Figure 2: Estimated probability function $\pi(z)$

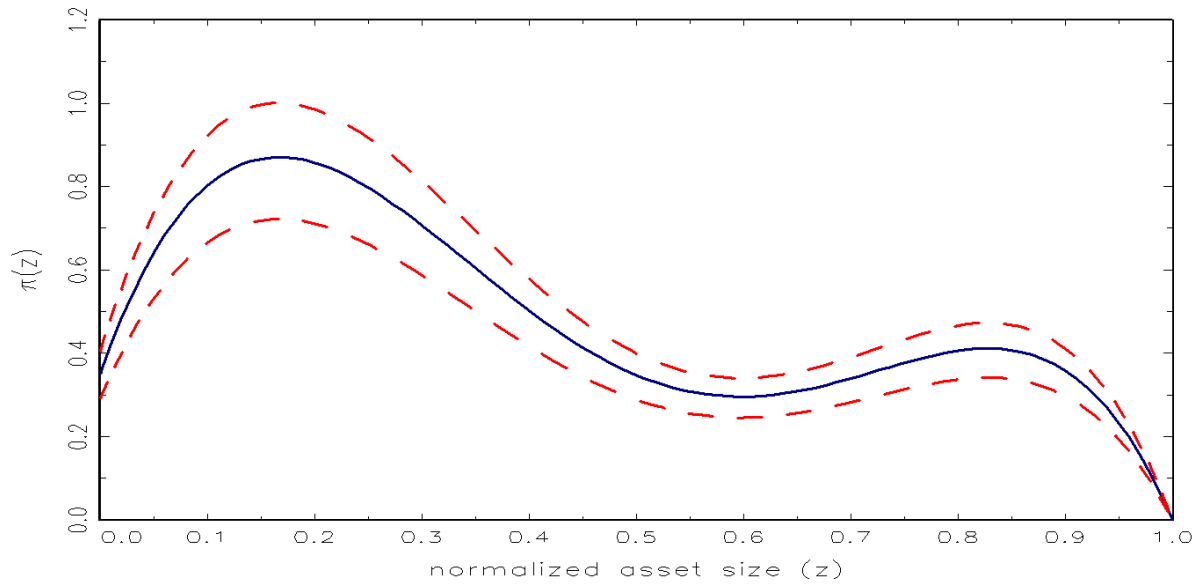


Figure 3: Technical inefficiency distributions

