

A ZERO INEFFICIENCY STOCHASTIC FRONTIER MODEL

SUBAL C. KUMBHAKAR[†], CHRISTOPHER F. PARMETER, AND EFTHYMIOS G. TSIONAS

ABSTRACT. Traditional stochastic frontier models impose inefficient behavior on all firms in the sample of interest. If the data under investigation represent a mixture of both fully efficient and inefficient firms then off-the-shelf frontier models are statistically inadequate. We introduce the zero inefficiency stochastic frontier model which can accommodate the presence of both efficient and inefficient firms in the sample. We derive the corresponding log-likelihood function, conditional mean of inefficiency to estimate observation-specific inefficiency and discuss testing for the presence of fully efficient firms. We provide both simulated evidence as well as an empirical example which demonstrates the applicability of the proposed method.

JEL Classification No.: C13, C23, C33

Keywords: Full Efficiency, Zero-Inflated, Mixture, Banking.

1. INTRODUCTION

The basic premise of neoclassical production theory is that every firm is fully efficient. This is, however, not the case in reality. The stochastic frontier (SF) methodology is designed to estimate the underlying production technology (represented by production, cost, revenue, profit, distance functions, among others) while accounting for random noise and incorporating inefficient behavior of firms simultaneously. Inefficiency is inherently unobservable, being convoluted with traditional stochastic shocks. By modeling inefficiency as a one-sided random variable, along with additional distributional assumptions on the noise component, one can cleave inefficiency from the composed error and generate estimates of firm level inefficiency.

Since its inception nearly three and a half decades ago (Aigner, Lovell and Schmidt 1977, Meeusen and van den Broeck 1977) specification of inefficiency in the SF model has been extended in many directions. However, one perspective on inefficient behavior that to date has largely been overlooked is the presence of both fully efficient and inefficient firms in the sample at hand. Put differently, standard SF models (and various extensions) currently in use cannot accommodate the case when some firms are fully efficient. Under the current SF modeling approach, which assumes that the inefficiency distribution governing behavior over all firms is continuous, even fully efficient firms will be deemed inefficient since the probability of having zero inefficiency is zero. A likely occurrence

Date: August 13, 2012.

[†]Subal C. Kumbhakar, Corresponding Author, Department of Economics, State University of New York at Binghamton, Binghamton, NY 13902-6000. E-mail: kkar@binghamton.edu, Phone: (607) 777-4762, Fax (607) 777-2681.

Christopher F. Parmeter, Department of Economics, University of Miami, Coral Gables, FL 33124-6520. Phone: (305) 284-4397, Fax: 305-284-2985, E-mail: cparmeter@bus.miami.edu.

Efthymios G. Tsionas, Department of Economics, Athens University of Economics and Business, 76 Patission Street, 10434 Athens, Greece. Phone: +30210-8203-338, Fax: +30210 8203-301, E-mail: tsionas@aueb.gr.

in applied work is the presence of fully efficient firms, which creates an abundance of zeros in the distribution of inefficiency.

The current menu of approaches within the SF paradigm do not offer econometrically feasible solutions for this issue. While one can test for the lack of inefficiency, this in essence treats all firms as fully efficient. Instead, a likely scenario is that there are both fully efficient and inefficient firms in the sample under scrutiny. Thus, the ideal situation is to have a model that allows *some* firms to be fully efficient, not necessarily *all* (as assumed in the neoclassical production theory) or *none* (as assumed under current SF models). To determine whether this is indeed the case in an empirical setup, it is necessary to consider an econometric model that allows firms to be fully efficient in a probabilistic framework. Our point of departure in this paper is that some firms can be fully efficient and so the distribution of inefficiency must be censored at zero.

To accommodate the fact that a subset of firms may be fully efficient, we develop a ‘zero inefficiency stochastic frontier’ (ZISF) model. This approach has connections with Lambert (1992) who developed the zero-inflated Poisson (ZIP) model to handle the over abundance of zeros arising in certain count data settings.¹ A key difference of the ZIP methodology with our approach is that the potential for an abundance of zeros is *unobserved*. In the ZIP model the outcome variable contains an abundance of zeros that are observed, whereas in our approach, given that inefficiency is convoluted with a random shock, we cannot directly observe firms that are fully efficient. The development and estimation of zero-inflated econometric models have become widespread. For example, Harris and Zhao (2007) develop a zero-inflated ordered probit model to estimate consumption levels of tobacco. Additionally these methods have recently appeared in recreation demand (Gurmu and Trivedi 1996), conflict analysis (Clark 2003), ecological modeling (Potts and Elith 2006), health economics (Bauer, Göhlmann and Sinning 2007) and traffic safety (Lord, Washington and Ivan 2005, 2007).

The ZISF model actually possesses an important link with the traditional SF model. If all firms were fully efficient (i.e., the probability of being fully efficient is 1) then a parametric restriction of zero variance for the inefficiency distribution makes the ZISF and SF models identical to the neoclassical production model. On the other hand, if none of the firms are fully efficient the ZISF model reduces to a SF model. Since the ZISF model allows each firm to be fully efficient with some probability, this setup is a generalization of the standard SF and the neoclassical production model which treats all firms as fully efficient. Moreover, the ZISF offers several different ways in which to categorize inefficiency that are unavailable in the traditional SF. Econometrically, both the ZIP and ZISF models can be viewed as models of regime membership, where membership in a particular regime is unobserved. In econometrics these regime membership issues are most commonly tackled using a finite mixture (latent class) modeling procedure.

Latent class SF models have recently appeared in the efficiency analysis literature (see Orea and Kumbhakar 2004, Greene 2005). These methods can provide a more parsimonious specification of the data at hand *a priori* by allowing for various degrees of heterogeneity. In the latent class

¹We thank an anonymous referee for drawing our attention to this connection and the surrounding literature.

SF model the focus is on technological and inefficiency induced heterogeneity. The finite mixture approach assumes that firms have access to a finite number of technologies and the inefficiency associated with each technology comes from a specific distribution (half-normal, truncated normal, exponential, etc.) in which the parameters differ. The idea is to identify (probabilistically) which firm is using what technology and then measure inefficiency as a probability weighted average computed using each of these technologies as the benchmark technology.

In this paper we abstract from technological heterogeneity and focus exclusively on the distribution of inefficiency. Furthermore, we assume that there are only two types of firms (efficient and inefficient). While perhaps providing misleading nomenclature, the level of inefficiency for fully efficient firms is zero; in essence the inefficiency distribution for these firms is a point mass at 0. On the other hand, the degree of inefficiency for inefficient firms is captured by any of the array of standard one-sided distributions, viz., half-normal, exponential, truncated normal, gamma, etc. When the number of observations is large and the market is competitive (which is the case in our banking application), it is likely that some firms will be fully efficient. Failure to recognize this feature leaves one with an incomplete analysis of inefficiency within the sample, irrespective of whether one uses a traditional mixture approach or not.

Recent empirical studies using latent class SF models unknowingly document full efficiency. Bos, Economidou and Koetter (2010) and Bos, Economidou, Koetter and Kolari (2010) use the finite mixture stochastic frontier model to determine the number of groups classified by technological heterogeneity and inefficiency. Both these papers find evidence of clustering and in each paper one of the clusters has near zero estimates of inefficiency. However, even with this finding of a subset of fully efficient firms, the use of a latent class SF methodology in no way guarantees that at least one of the estimated groups will be fully efficient. Moreover, it is not clear from the finite mixture approach whether identifying a group of efficient firms is actually predicated on overfitting from allowing technological heterogeneity across the regimes. Here we provide a framework to estimate firm level inefficiency when the underlying technology is homogeneous but not the distribution of inefficiency. We derive the likelihood function and discuss estimation of firm-specific inefficiency using a single equation. Additionally, we provide a test for the presence of fully efficient firms that is more appropriate than existing tests which only test for “full” efficiency (an all or nothing proposition).

The ZISF also has appeal outside of its econometric content. In many regulatory cases (electricity generation/distribution, water utilities, public transportation, post offices, etc.), regulators are interested in benchmark firms that are fully efficient either absolutely or relatively. The benchmark is typically preferred to be more than one firm. The benchmark firm(s) using the currently available SF models can be defined by the top $x\%$ of the firms which is subjective. This subjective element can be mitigated if we can statistically identify the group of firms that are fully efficient from the model. The ZISF offers several alternatives for classifying firms as fully efficient which the standard SF model does not provide.

A cognate strand of research can be found in Sickles and Qian (2009). They suggest modification of the standard stochastic frontier model by right truncating the distribution of inefficiency. That is, they argue that firms cannot be grossly *inefficient*, while we focus on the possibility of firms being fully efficient. To accommodate this feature they truncate the distribution of inefficiency by placing a threshold on the extent of inefficiency. By doing so they introduce an upper bound on inefficiency. Thus, they introduce truncation as a means to modify the SF model whereas we introduce censoring into the standard SF approach. The difference between these two modeling approaches underscores how one perceives the shape of the underlying distribution of inefficiency.

The rest of the paper proceeds as follows. Section 2 outlines the ZISF estimator, and discusses construction of efficiency scores and testing for the presence of fully efficient firms. Section 3 conducts several simulation exercises to determine the impact that fully efficient firms has within a stochastic frontier analysis. Section 4 provides a small empirical application to the U.S. banking industry to compare results across the SF and ZISF models. Section 5 presents some concluding remarks and extensions.

2. METHODOLOGY

2.1. The Zero Inefficiency Stochastic Frontier Model. The basic stochastic production frontier model is:

$$(1) \quad y_i = x_i' \beta + v_i - u_i = x_i' \beta + \varepsilon_i, \quad \text{for } i = 1, \dots, n,$$

where y_i is a scalar output, x_i is a $k \times 1$ vector of covariates, β is a $k \times 1$ vector of parameters, v_i is noise, and u_i represents technical inefficiency. In this literature it is standard to assume that $v_i \sim i.i.d. N(0, \sigma_v^2)$ and u_i is independently and identically distributed as half-normal (Aigner et al. 1977), exponential (Meeusen and van den Broeck 1977), truncated normal (Stevenson 1980) or gamma (Greene 1990).

As mentioned in the introduction, we assume that some firms are fully efficient ($u_i = 0$ for some i) while others are inefficient $u_i > 0$ and this information is not available to the econometrician. Thus, our problem is to assess which regime (efficient or inefficient) each firm belongs to. To do this we formulate our ZISF model as

$$ZISF \longrightarrow y_i = x_i' \beta + v_i \text{ with probability } p \text{ and } y_i = x_i' \beta + (v_i - u_i) \text{ with probability } (1 - p),$$

where p is the probability of a firm being fully efficient. Alternatively, p is the proportion of firms that are fully efficient and $1 - p$ is the proportion of firms that are technically inefficient.

Note that the technology in both regimes is the same. Since the regimes are unobserved, the ZISF falls into the category of a latent class model. Previous studies on latent class SF models either have different probabilities on the technology, $x_i' \beta$, or the composed error, $\varepsilon_i = v_i - u_i$ (or both). By acknowledging the fact that a subset of firms belong to a fully efficient regime, the ZISF model boils down to an analysis of regime membership based solely on inefficiency. The composed error term in the ZISF is $v_i - u_i(1 - 1\{u_i = 0\})$ where $1\{u_i = 0\} = p$. Thus, the ZISF error term

is not the same as the probability weighted composed error term $v_i - u_i$ which would be the case if one were to use a latent class SF model with identical technology.²

The density of our convoluted error term, assuming that $u_i \sim i.i.d. N_+(0, \sigma_u^2)$ and $v_i \sim i.i.d. N(0, \sigma_v^2)$, viz. is,

$$(2) \quad (p/\sigma_v)\phi(\varepsilon/\sigma_v) + (1-p) \left[\frac{2}{\sigma} \phi\left(\frac{\varepsilon}{\sigma}\right) \Phi\left(\frac{-\varepsilon}{\sigma_0}\right) \right],$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density and cumulative distribution functions of a standard normal random variate, respectively, and $\lambda = \sigma_u/\sigma_v$, $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $\sigma_0 = \lambda/\sigma$. The ZISF convoluted error distribution follows directly from the intuition afforded from our earlier discussion on regime membership.³ This density function is a mixture between a normally distributed random variable and the common, convoluted density from a normal/half-normal SF model. Thus, any of the array of standard SF densities can be used in deriving the likelihood function for the ZISF model. A natural extension of the model discussed here that does not require additional computational effort is to make the probability of full efficiency a parametric function of a set of covariates:

$$p_i = \frac{\exp(z'_i \gamma)}{1 + \exp(z'_i \gamma)}, \text{ or } p_i = \Phi(z'_i \gamma), \quad \text{for } i = 1, \dots, n,$$

where z_i is an $m \times 1$ vector of exogenous variables which influence whether a firm is inefficient or not and γ is an $m \times 1$ vector of parameters.

In any latent class model a key concern is identifying the number of classes. Here we do not face as difficult a problem because we are dealing *a priori* with two classes (fully efficient and inefficient firms). The only additional parameter in our ZISF model in comparison with the standard SF model is p . In a two class mixture model this is the mixing proportion parameter in the population. Thus, statistical identification of this parameter requires non-empty cells (non-zero observations in each group) in the sample (see McLachlan and Peel 2000). The variance of the inefficiency distribution, σ_u^2 , in the SF model is identified through moment restrictions on the composed density of the inefficiency and random components (skewness of the composed error distribution). When $\lambda \rightarrow 0$ identification of the model breaks down as the inefficient regime and the fully efficient regime become indistinguishable. However, this case does not pose a problem intuitively since $\lambda \rightarrow 0$ implies that firms are becoming close to fully efficient.

2.2. Estimating Firm-Specific Inefficiency. Here we discuss several approaches to estimate observation-specific inefficiency in the ZISF model. Natural estimators for inefficiency in the ZISF are the Jondrow, Lovell, Materov and Schmidt (1982) (ZI-JLMS hereafter) conditional mean or

²The case of a finite mixture of heterogeneous composed error distributions with homogeneous technology can be derived as a special case of Greene (2005).

³See Kumbhakar and Lovell (2000, ch. 3) for more on these derivations.

mode.⁴ The conditional density of u given ε is 0 with probability p and

$$\frac{\phi((u - \mu_*)/\sigma_*)}{\sigma_* \Phi(-\varepsilon/\sigma_0)},$$

with probability $(1 - p)$, where $\sigma_*^2 = \frac{\sigma_v^2 \sigma_u^2}{\sigma_v^2 + \sigma_u^2} = \frac{\sigma_v^2 \sigma_u^2}{\sigma^2}$ and $\mu_* = -\varepsilon \sigma_*^2 / \sigma_v^2 = -\varepsilon \sigma_u^2 / \sigma^2$. In other words, the conditional distribution of u given ε is normal with probability p and truncated normal, $N_+(\mu_*, \sigma_*^2)$, with probability $1 - p$. When $p = 0$ this is exactly the JLMS conditional density for the normal/half-normal setting (see Kumbhakar and Lovell 2000, pgs. 77-78). From this it follows that the conditional mean estimator for u in our ZISF model is

$$(3) \quad E[u|\varepsilon] = (1 - p) \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2} \left[\sigma_0 \frac{\phi(\varepsilon/\sigma_0)}{\Phi(-\varepsilon/\sigma_0)} - \varepsilon \right].$$

To implement this formula in practice one has to replace the unknown parameters by their ML estimates and ε by the residuals based on the ML estimates.

As opposed to using the conditional mean, the modal estimator might be preferred because it can be viewed as the ML estimator of $u|\varepsilon$ (see Materov 1981, Jondrow et al. 1982). The conditional mode estimator is found by taking the derivative of our conditional density:

$$(4) \quad M(u|\varepsilon) = \frac{df(u|\varepsilon)}{du} = 0.$$

This derivative is known to have a zero at $u = \mu_*$ when $\varepsilon < 0$ or 0 otherwise. Multiplying μ_* by $1 - p$ yields the fully efficient modal estimator. Intuitively, our conditional mode estimator is the same as the traditional JLMS estimator except that it is shrunk towards zero (as with the conditional mean estimator).

An interesting aspect of the ZI-JLMS estimator is that it is essentially a shrinkage estimator. However, all estimates of inefficiency are shrunk by the same proportion, p . This suggests that, on average, all firms are less inefficient. An alternative point of view is to consider that the truly inefficient firms in the ZISF are more inefficient than they are in the traditional SF. A flexible approach to measuring inefficiency with an eye towards this point is two fold. First, we can construct posterior estimates of inefficiency, defined as $\check{u}_i = (1 - \check{p}_i)\hat{u}_i$, where \check{p}_i is the posterior estimate of the probability of being fully efficient,

$$(5) \quad \check{p}_i = \frac{(\hat{p}/\hat{\sigma}_v)\phi(\hat{\varepsilon}_i/\hat{\sigma}_v)}{(\hat{p}/\hat{\sigma}_v)\phi(\hat{\varepsilon}_i/\hat{\sigma}_v) + (1 - \hat{p})\frac{2}{\sigma} \phi(\hat{\varepsilon}_i/\hat{\sigma}) \Phi(-\hat{\varepsilon}_i/\hat{\sigma}_0)}$$

and \hat{u}_i is the ZI-JLMS estimator of inefficiency in (3) with $p = 0$. The estimates in (5) represent ML parameter estimates and $\hat{\varepsilon}_i$ are the residuals based on these ML parameter estimates. These posterior estimates of inefficiency are firm specific and, while they also shrink the conditional mean towards zero, each inefficiency score is shrunk based on the posterior probability of being fully efficient. Second, we can censor the ZI-JLMS scores based on a pre-specified cut-off for the

⁴We thank Peter Schmidt for suggesting this to us.

estimated posterior probabilities. Once this cut-off has been established firms found to be inefficient can receive their ZI-JLMS scores as in (3) with $p = 0$.

Since \check{p}_i is firm-specific, researchers and regulators can either use these posterior estimates of inefficiency or the posterior odds ratio, $R_i = \check{p}_i/(1 - \check{p}_i)$ to define fully efficient firms given a pre-specified cutoff. These \check{p}_i will also give the regulators an idea of how likely (in a probabilistic sense) a particular firm (or group of firms) is to being fully efficient. This is particularly helpful in merger cases. For example, in recent merger hearing of water utilities in the UK the Competition Commission was critical about allowing an efficient utility to merge with an inefficient one on the ground that the number of efficient utilities would decline.

2.3. Testing for Full Efficiency. To test for full efficiency (or full inefficiency) one cannot resort to a simple t -test using the standard errors produced from maximization of the likelihood function. This is because the hypothesis of full efficiency, $H_0: p = 1$ (or inefficiency, $H_0: p = 0$) lies on the boundary of the parameter space (see Andrews 2001). It is well known that testing for fully efficient behavior in the traditional SF model results in a log-likelihood ratio statistic that is distributed as a mixture of χ^2 distributions (Lee 1993, Coelli 1995). An interesting extension of this insight is that of Chen and Liang (2010) who have shown that when a nuisance parameter such as p lies on the boundary, hypotheses testing regarding this parameter yields asymptotic mixture distributions. Furthermore, testing hypotheses on other parameters from the model will also result in a mixture of normal distributions. See their section 2.3 for additional details.

To test $p = 1$, one can use the pseudo-likelihood ratio (PLR) test. The PLR test is $PLR = -2(L_N - L_{ZI})$, where L_N is the log likelihood of the normal linear model (evaluated using the standard OLS estimates) and L_{ZI} is the log likelihood of the ZISF model. Using arguments similar to Chen and Liang (2010, p. 613) it can be shown that

$$(6) \quad \frac{\partial \ln f(\varepsilon)}{\partial p} = \frac{1}{f(\varepsilon)} \left[\sigma_v^{-1} \phi(\varepsilon/\sigma_v) - \frac{2}{\sigma} \phi(\varepsilon/\sigma_v) \Phi(-\varepsilon/\sigma_0) \right],$$

which under $p = 1$ has expectation zero. From the results in “Case 2” of Chen and Liang (2010, p. 608) the PLR test has an asymptotic distribution which constitutes a 50:50 mixture of χ_0^2 and χ_1^2 distributions (as in Coelli 1995).

A notable insight of our test for full efficiency here is that rejection of the null does not imply fully inefficient behavior as in Coelli (1995). In applied work where full efficiency is tested for the null hypothesis is $H_0: \sigma_u = 0$, so rejection of the null immediately implies inefficient behavior across all firms. Our hypothesis is more flexible towards investigating the presence of efficiency, since in our test we are focusing on the probability of being fully efficient (and not specifically on the variance of the inefficiency distribution). That is, a rejection of the null hypothesis is interpreted differently across the SF and ZISF settings.

3. SIMULATION RESULTS

The use of a ZISF model poses several interesting questions to practitioners. First, how misleading is the ZISF model when the true data generating process (DGP) is a standard SF? That is, are the posterior classification probabilities of full efficiency close to zero? Does one get the same inefficiency estimates from the two models in this setting? Second, what are the implications of using a standard SF model when, in fact, the true DGP is a ZISF? Third, when the true DGP is a ZISF how reliable is the application of the ZISF estimator? That is, how closely can we identify fully efficient firms and how accurately can we estimate the inefficiency levels of inefficient firms.

To address these questions we turn to Monte Carlo simulations. Given the myriad combinations of the set of parameters within the ZISF we limit our attention for the present purposes to one specific setup for our simulations. Results using alternative combinations are available from the authors upon request. For all of our simulations we set $\sigma_v = 0.1$. We use either $\lambda = 1$ or 5 and let the sample sizes vary over $n = 500$ or 1000. For each experiment we conduct 1,000 repetitions. Our basic frontier model is

$$(7) \quad y_i = 1 + x_i + \varepsilon_i,$$

where x_i is generated as a standard normal variable.

3.1. Finite Sample Results I: The true DGP is a normal/half-normal SF model but one estimates the ZISF model. Our first set of simulations, which assumes that u and v are generated from a standard SF model, i.e., $v \sim N(0, 0.1)$ and $u \sim N_+(0, \lambda \cdot 0.1)$, focuses on using the ZISF when it is clearly an over-specification of the problem in the sense that the true DGP is a SF (i.e., $p = 0$). The presence of p only serves to add an additional parameter to the problem. The ML estimator from the ZISF model is found to perform quite well. No numerical problems were encountered while using a standard conjugate gradients algorithm to maximize the likelihood function. Estimates of p were close to zero and statistically insignificant using the PLR test described in section 2.3.⁵

We construct posterior probabilities of full efficiency for each firm as well as estimates of inefficiency using the JLMS estimator in (3). We compare the density of these estimates of inefficiency with the density of estimated inefficiencies arising from the use of the proper, normal/half-normal SF model (that is, using (3) with $p = 0$). The densities of estimated inefficiencies are presented in Figure 1. The sampling distributions are obtained as averages (across firms) of the usual JLMS estimator of technical inefficiency.

[Figure 1 about here.]

It can be seen from Figure 1 that our ZISF estimates of inefficiency track the true inefficiency levels arising from the SF model (true DGP) quite well. The reason for this is that estimates of p are close to zero as expected since the ZISF generalizes the traditional SF estimator. We also find

⁵For further simulation results focusing on p see Rho and Schmidt (2012).

that the absolute efficiency rankings across firms are similar between the SF and ZISF models.⁶ One interesting insight on our results is the additional density that occurs in each of the four setups for very low levels of inefficiency. This is intuitive since the ZI-JLMS estimator ‘collapses’ the traditional conditional mean estimator towards zero and so we expect additional density here, though this weakens as the sample size increases.

3.2. Finite Sample results II: The true DGP is ZISF but one uses a normal/half-normal SF model. For this simulation setup we are concerned that the presence of fully efficient firms degrades our estimates of inefficiency if we ignore their presence. To assess this issue we allow the probability of full efficiency in our DGP to vary over $p = 0$ (no fully efficient firms), as well as $p = 0.1$, and 0.25 . We chose relatively low values of p to allay concerns that our results are driven by the inclusion of a large proportion of fully efficient firms. The sampling distribution is obtained as an average (across simulations) of the usual JLMS estimator of inefficiency. When $p = 0$ this estimator is the appropriate estimator. The sampling distributions of the sample mean JLMS inefficiency estimates are presented in Figure 2 for $p = 0$, $p = 0.1$ and $p = 0.25$.

Our main findings are: (i) The usual normal/half-normal SF model does produce sizeable differences in estimated inefficiency. Relative to the mean of the sampling distributions, the difference is close to one percentage point when $\lambda = 1$ but as large as 5 points when λ increases to 5. (ii) Ignoring the ZISF DGP leads to overestimation of technical inefficiency, even for small values of p .

The differences here also document an interesting feature. For $\lambda = 1$ it appears that the inefficiency scores are less inaccurate for $p = 0.1$ while for $\lambda = 5$ the inefficiency scores are roughly equivalent in terms of their inaccuracy. The intuition here is that for λ small, it is harder to distinguish between noise and inefficiency and so a larger p is required to see biases in inefficiency scores. Alternatively, when $\lambda = 5$ the density of estimated inefficiency for $p = 0.25$ always lies to the right of the equivalent density for $p = 0.1$ suggesting better identification of inefficiency and as such a narrowing of the gap between the two densities (this is expected since the inefficient firms in our simulated samples have the same level of inefficiency irrespective of the value of p).^{7,8}

[Figure 2 about here.]

3.3. Finite Sample Results III: Performance of the ZISF estimator when the DGP is a ZISF. Perhaps the most useful simulations for assessing the ZISF estimator are when it is appropriately applied to a ZISF DGP. We follow the same setup as our previous examples and focus attention on the results stemming from $p = 0.25$. Although we have conducted a large number of experiments for this simulation setup, for brevity we report here only the results for $p = 0.25$ (which

⁶Details on these results are not reported here to save space. However, these rankings can be obtained from the authors upon request.

⁷We mention, in passing, that the estimated differences between the underlying level of inefficiency and those estimated from the standard SF when the DGP is a ZISF model are even greater as σ increases from 0.1. In additional simulations we have seen differences as large as 15 to 20 percentage points depending on the relative sizes of λ and σ .

⁸To save space here we do not report results for other Monte Carlo trials using different values for n , λ and σ . These results for additional values of n , λ and σ are available from the authors upon request. The qualitative results remain the same as those described here.

is in the lower end of the estimated percentage of fully efficient banks in our sample). In addition to examining the sampling distribution of the ML estimators of p , λ , and σ , we examine the sampling distribution of the posterior odds ratio (the main classification tool) for the censored firms. This ratio should be greater than one for most of the censored (fully efficient) firms. Equivalent log posterior odds ratio in favor of non-censoring should be greater than one for inefficient firms.

Figure 3 presents the representative sampling distributions over 1,000 Monte Carlo repetitions. The sampling distributions of the ML estimators for p , λ and σ are centered around the true values and they appear reasonably symmetric though the distribution for λ_{ML} does have a long right tail. The sampling distribution of the log posterior odds ratio for the censored firms is constructed as follows. For each replication we compute \check{p}_i as in (5). The \check{p}_i s for the censored firms are then singled out (since we have an artificial experiment we know exactly which firms are censored). We then take the mean value of $\check{p}_i/(1 - \check{p}_i)$, the odds ratio, which is used to formally classify firms into censored and non-censored groups. The upper left panel in Figure 3 presents the density of the log odds ratios (we take logs for clarity in presentation). For $p = 0.25$, $\lambda = 5$ and $n = 1,000$ the ZISF model does a good job in discriminating the efficient firms since most of the mass of the sampling distribution of the log posterior odds ratio for the censored firms is to the right of zero.

[Figure 3 about here.]

3.4. Finite Sample Results IV: Performance of the PLR Statistic. While the asymptotic distribution of the PLR statistic is known to be $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$, it is important to discern the finite sample performance of this statistic to gauge its usefulness in applied settings. Here we use the DGP $y = \beta + v$, where $\beta = 0$, $v \sim N(0, 1)$. For each one of 10,000 replications we fit the fully efficient stochastic frontier to obtain L_{ZI} and we compute L_N using the OLS estimates. We use sample sizes of $n = 50, 100, 200$ and 500 . The actual critical values (90%, 95% and 99%) from $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ were obtained by taking 10,000,000 draws from this distribution and calculating them numerically.

[Table 1 about here.]

Table 1 presents the finite sample critical values from our 10,000 replications for each sample size considered. As we see for n as small as 200 the performance of the PLR statistic appears to follow closely the asymptotic distribution. It appears that the finite distribution of the PLR statistic has a somewhat thinner tail as the 99% critical values are smaller than the asymptotic critical value for n equal to 500. Overall, these results suggest that the PLR can be relied upon in empirical work to adequately statistically discriminate full efficiency.

4. APPLICATION

In this section we provide an empirical example to showcase the merit of the ZISF model. For our data we use detailed cost information from the commercial bank and bank holding company database managed by the Federal Reserve Bank of Chicago. This data is based on the Report of Condition and Income for all U.S. commercial banks that report to the Federal Reserve banks and

the FDIC. We use a balanced panel of 4,985 banks over the period 1989-2000; this provides us with a sample of 59,820 observations. Kumbhakar and Tsionas (2005) used a random sample of 500 of these banks for the period 1996-2000 for their study on measuring technical and allocative inefficiency in a cost system.

Given the recent downfall of the U.S. financial system it is evident that the commercial banking industry is one of the most important drivers of the U.S. economy. The 1990s saw widespread changes in the structure of the banking industry given consolidation and mergers. These acts of shedding can be justified based on economies of scale and efficiency. Since we are using panel data (whereas our model in Section 2 is designed for cross-sectional data) we control for time in the SF function to allow for technical change (shift in the frontier). We also need to make assumptions on the temporal behavior of inefficiency and noise. Here, for simplicity, we are assuming both u and v to be independently and identically distributed.⁹ By assuming no specific temporal behavior on inefficiency, our model allows a bank to be fully efficient in one year but not in others. Also, we do not restrict the firm level of inefficiency to change over time in any specific manner which is commonly assumed to be the same for all firms in applied settings (see for example the models in Kumbhakar and Lovell 2000, chapter 3). If we follow the literature and impose such behavior by assuming $u_{it} = u_i\gamma(t)$ where $\gamma(t)$ is a parametric function of time, then fully efficient banks will be efficient across the entire time period regardless of $\gamma(t)$. Clearly this is a restrictive assumption. Thus, the i.i.d. assumption on u fits quite well in our ZISF model.

The choice of input and output variables in the banking literature on efficiency studies is subject to controversy. We follow the widely used intermediation approach (Kaparakis, Miller and Noulas 1994) which views banks as financial firms whose job is to transform various financial and physical resources into loans and investments. Our outputs are installment loans (y_1), real estate loans (y_2), business loans (y_3), federal funds sold and securities purchased (y_4) and other assets¹⁰ (y_5). Our inputs are labor (X_1), capital (X_2), purchased funds (X_3), and interest-bearing deposits in total transaction and non-transaction accounts (X_4 and X_5 , respectively). To calculate input prices (w_1 through w_5) we divided the total expense for each input by the corresponding input quantities.

Since we have multiple outputs and multiple inputs we cannot use a production frontier approach. Furthermore, it is customary to view banks as cost minimizers because outputs are exogenously given (outputs are services which cannot be stored like manufacturing outputs). Thus, we use a cost function approach and change the previous notation of the stochastic production frontier to that of a stochastic cost frontier for both the standard SF and ZISF models. This is done by a simple change in the sign of the inefficiency term and of course a change in the interpretation of inefficiency. The arguments in the cost function are input prices, outputs and a time trend. Our discussion of the empirical results are based on a multiple output cost frontier for both the standard and ZISF models.

⁹The model can easily be extended to accommodate heteroskedasticity in both u and v .

¹⁰These are assets that cannot be properly included in any other asset items balance sheet.

4.1. Empirical Results. To begin we focus on estimation of a cost function. We use the translog functional form, written out fully as

$$\begin{aligned}
 \ln C_{it} = & \alpha_0 + \sum_{j=1}^4 \alpha_j \ln w_{j,it} + \sum_{k=1}^5 \beta_k \ln y_{k,it} + 0.5 \sum_{m=1}^5 \sum_{k=1}^5 \beta_{mk} \ln y_{k,it} \ln y_{m,it} \\
 & + 0.5 \sum_{j=1}^4 \sum_{q=1}^4 \alpha_{jq} \ln w_{j,it} \ln w_{q,it} + \sum_{k=1}^5 \sum_{j=1}^4 \gamma_{kj} \ln y_{k,it} \ln w_{j,it} + \alpha_t t \\
 (8) \quad & + 0.5 \alpha_{tt} t^2 + \sum_{j=1}^4 \delta_{jt} \ln w_{j,it} t + \sum_{k=1}^5 \theta_{kt} \ln y_{k,it} t + v_{it} + u_{it}.
 \end{aligned}$$

We assume the usual symmetry restrictions ($\beta_{mk} = \beta_{km}$ and $\gamma_{kj} = \gamma_{jk}$) hold. Since the cost function is homogeneous of degree one in input prices, we need to impose linear homogeneity restrictions. These restrictions are automatically imposed if cost and input prices are normalized by one input price (we use w_5). Furthermore, we assume that v_{it} is normally distributed and model u_{it} as a half-normal. Both are assumed to be independently and identically distributed over i and t .

For brevity we do not report the parameter estimates (more than 50) stemming from our translog cost function.¹¹ Instead, we report the parameters associated with the variances from each of the noise terms as well as the probability of full efficiency. Table 2 presents these estimates across the standard SF and ZISF estimators.¹²

[Table 2 about here.]

Our estimate of the probability of being fully efficient is 36% and this is statistically different from zero at the 1% level using the PLR test discussed in section 2.3. Inspection of the estimates of the variance parameters within the composed error reveals that while the ZISF and standard half-normal SF provide almost equivalent estimates for the noise variance, the estimates for the variance of inefficiencies are, however, markedly different. The ZISF has a variance that is almost 50% larger. This result is intuitive. In the standard SF model the variance of u is lower because the presence of fully efficient banks (for which the variance is zero) pulls down the variance. When the ZISF model allows for the presence of fully efficient banks, the variance of u for the inefficient banks is increased.

To further investigate this issue we re-estimated both stochastic frontier models by excluding the top 1%, 2%, ..., 5% of the most efficient banks (based on efficiency scores). These are presumably the banks that are most likely to be fully efficient. The variation in $\hat{\sigma}_v$ remained almost the same but $\hat{\sigma}_u$ increases steadily to about 0.300 when we exclude the top 5% in the SF model. Eliminating the top 1%, 2%, ..., 5% gives estimates for σ_u of 0.219, 0.243, 0.255, 0.276, 0.289, respectively. The probability of full efficiency from the ZISF, p is reduced considerably to about 0.03 when the top 5% is omitted, which is to be expected. In essence, removing the bank-year observations which

¹¹These estimates are available upon request.

¹²All code is available from the authors upon request. We also note that the ZISF can be implemented using LIMDEP version 9.0.

have the lowest, *ex poste* estimates of technical inefficiency results in the ZISF and SF producing identical results. Additionally, we used the posterior probabilities in (5) to distinguish between fully efficient banks ($\check{p}_i > 0.5$). Using this set of banks we estimated a stochastic frontier model and obtained an estimate of $\sigma_u \approx 0$, suggesting that our classification did produce a set of banks for which a traditional stochastic frontier model could not detect inefficiency.

[Figure 4 about here.]

Figure 4 presents several measures of assessing inefficiency in the ZISF model along with the traditional measurement of inefficiency within the SF model. The upper left block is the histogram of the standard distribution of conditional inefficiency using the JLMS estimator with the SF ML estimates. The upper right block is the ZI-JLMS estimates as calculated in (3). The figures in the lower blocks represent alternative characterizations of bank specific inefficiency in the ZISF framework described in section 2.2. First, the figure in the lower left block uses a generic cutoff rule to determine inefficiency. For banks' whose posterior probability of full efficiency is greater than 0.95 we assign an inefficiency level of zero, whereas for those banks who have a posterior probability of full efficiency less than 0.95 we assign to them their JLMS inefficiency score constructed using the ZISF ML estimates. The histogram in the lower right block represents the posterior estimates of inefficiency, using (5) to calculate the posterior probability of full efficiency. As is quite noticeable, how one chooses to think of measuring inefficiency has an impact on the shape of the distribution of conditional inefficiency. Moreover, as discussed earlier the ZISF approach offers an array of methods to estimate and display inefficiency beyond the traditional SF.

Table 3 presents summary statistics for our four measures of inefficiency from Figure 4. We present the upper and lower deciles (D_{90} , D_{10}), the quartiles (Q_{25} , Q_{50} , Q_{75}) as well as the mean and standard deviation. The first row (Standard JLMS) focuses on the exact JLMS estimates obtained from the standard SF model. The second row (Censored JLMS) presents inefficiency scores based on censoring given the level of the posterior probabilities from equation (5); only inefficiency levels for banks not deemed fully efficient are presented. We can see that, consistent with our estimates from Table 2, our censored JLMS conditional mean estimates of inefficiency are higher than the standard JLMS conditional mean estimates. The third row (ZI-JLMS) lists the full-fledged ZI-JLMS using our maximum likelihood estimates in equation (3). The JMLS inefficiency estimates from the ZISF are substantially lower than either of the other inefficiency estimates. This follows since, per (3), accounting for the uncertainty of a bank being fully efficient or not results in a shrinking of expected inefficiency. The fourth row (Posterior JLMS) presents the inefficiency scores based on ZI-JLMS but multiplied by the estimated posterior probabilities in (5) as opposed to \hat{p} in Table 2. Given that our estimated posterior probabilities are quite large, the shrunken individual inefficiency scores is not surprising. Regardless of which method is used to discern inefficient behavior, the message is clear. For the set of banks considered here, the extent of inefficiency is limited with respect to a more traditional SF analysis.

[Table 3 about here.]

A clear signal from application of the ZISF is the large proportion of banks (across different time periods) that are deemed to be fully efficient, using the posterior probability cutoff described above. In this setting, given that the banking market can be viewed as reasonably competitive we believe the presence of fully efficient banks is indicative of the underlying market structure more so than a statistical artifact. This also links nicely with Sickles and Qian (2009) who find that there are no grossly inefficient banks in the U.S.

Along with the estimation of inefficiency the translog cost frontier can be used to estimate additional measures of interest such as technical change (TC) and returns to scale (RTS), which are functions of the estimated parameters and data.¹³ TC for a cost function is defined as cost diminution (rate of change in cost over time, holding everything else constant), i.e., $\partial \ln C / \partial t$. A negative value suggests technical progress (reduction in cost over time, *ceteris paribus*). RTS from a cost function measures the proportional increase in costs given an increase in all outputs, which is defined as the reciprocal of $\sum_m \partial \ln C / \partial \ln y_m$. If RTS exceeds unity a proportional increase in all outputs will lead to a less than proportional increase in cost. In such a case scale economies are said to exist, i.e., the scale of operation is below optimum, and therefore there are benefits from expansion. The opposite is true if RTS is less than unity. Note that both RTS and TC are observation-specific (we dropped the i and t subscripts for simplicity).

We provide density plots of both RTS and TC for our cost function estimates in Figure 5.¹⁴ We can see that both RTS (panel (a)) and TC estimates (panel (b)) are nearly identical across the SF and ZISF parameter estimates. The Li (1996) test returns bootstrap p -values of 0.279 and 0.488 (using 999 replications) respectively, suggesting that statistically these densities are indistinguishable at conventional significance levels. Most of the banks experienced technical progress (indicated by the negative value of $\partial \ln C / \partial t$).

[Figure 5 about here.]

Our TC and RTS results are robust across the two estimation methodologies. This is expected because in our theoretical model inefficiency does not affect technical change and returns to scale. Note that the $x'\beta$ part determines both TC and RTS. In fact, it would be difficult to reconcile our results if they were statistically different because it would contradict the modeling setup which assumes that all producers (efficient and inefficient alike) have access to the same technology.

In summary, the lesson from the simulations and the banking application is that the ZISF model gives lower estimates of inefficiency scores compared to those from the standard SF model. This is especially true when the proportion of fully efficient observations are not negligible. In other words, large differences in the estimated inefficiency between the SF and ZISF models would indicate that the proportion of fully efficient observations cannot be ignored, especially if one were to draw policy implications from it. For example, if the estimated efficiency scores are used by regulators (as is the case with many industries such as electric, gas, and water utilities, railroads, airlines, postal services, etc.), it is important to know the group of efficient banks to construct the benchmark.

¹³The parameters of the translog function do not have a direct economic interpretation.

¹⁴We use Silverman rule-of-thumb bandwidths to construct all densities in Figure 5.

Under incentive regulation, incentives are provided to the inefficient entities (laggards) to improve their efficiency levels (to catch-up to the efficiency levels of the benchmark entities). The observed efficiency scores of laggards can then be compared with the benchmark banks to determine the potential for efficiency improvement which, in turn, might determine the types of incentives to be given for efficiency improvement. Given that inefficiency estimates from the SF models are higher, its use will give the regulators a false signal regarding the the extent of potential efficiency improvement. Consequently both targets and incentives might be wrong (set too high).

5. CONCLUSIONS

Zero-inflated methodologies represent a useful modeling strategy when observed outcomes display an over abundance of zeros. In applied production economics this over abundance of zeros can be viewed as the inclusion of both efficient and inefficient firms in the sample under econometric scrutiny. However, a key difference with this approach and existing work on zero-inflated methods is that the main outcome of interest in this setting (inefficiency) is unobserved. To accommodate this scenario in applied work this paper develops a stochastic frontier model that allows the probability of any firm to be fully efficient to be non-zero. The ZISF model marks a realistic and useful generalization of traditional stochastic frontier techniques, filling the previously existing lacuna in this literature. By focusing attention on the question of regime membership, the ZISF completely characterizes the inefficiency landscape and contains the traditional stochastic frontier as a parametric restriction. This model encapsulation of the benchmark stochastic frontier model offers the attraction of testing for the presence of fully efficient firms without connection to an everything or else alternative hypothesis.

Further, the ZISF helps one to define benchmark firms in a statistically meaningful manner as opposed to the current practice of using the firm with the lowest level of inefficiency in the sample or some given percentage of firms. Additionally, several methods were presented for measuring inefficiency that should prove useful for practitioners. Moreover, this new approach is conducive to application of the multiple comparisons with the best methodology introduced to the SF literature by Horrace and Schmidt (2000). Also stemming from estimation of a ZISF is a more natural test of the hypothesis of fully efficient behavior in one's sample of firms. By not restricting the alternative hypothesis to be inefficient behavior for all firms, our pseudo-likelihood ratio test is a more parsimonious characterization of the efficiency landscape. Lastly, this method is an extension of the existing studies on mixture models that allow for different technologies/efficiency distributions.

Useful extensions of the methods developed here include (i) a system approach using profit/cost functions, (ii) allowing both error components to be heteroskedastic and (iii) developing alternative zero inefficiency approaches exclusively for panel data. The most important aspect of a panel extension will be the temporal modeling of inefficiency. Here we placed no restrictions on the efficiency level of a given firm across time periods, but one could think of imposing the requirement that the probability of firm efficiency is increasing over time. We leave these extensions for future research.

Overall our simulated and applied results show that the ZISF methodology successfully allows fully efficient firms to be accounted for from the onset of a stochastic frontier study and that the method can uncover anomalies that traditional methods wash away with the rigid assumption of full inefficiency. We feel that these methods will be useful for applied researchers conducting efficiency studies and for regulators interested in pursuing alternative measures of benchmarking firms across a spectrum of industries.

REFERENCES

- Aigner, D. J., Lovell, C. A. K. and Schmidt, P. (1977), ‘Formulation and estimation of stochastic frontier production models’, *Journal of Econometrics* **6**(1), 21–37.
- Andrews, D. W. K. (2001), ‘Testing when a parameter is on the boundary of the maintained hypothesis’, *Econometrica* **69**, 683–734.
- Bauer, T., Göhlmann, S. and Sinning, M. (2007), ‘Gender differences in smoking behavior’, *Health Economics* **16**, 895–909.
- Bos, J. W. B., Economidou, C. and Koetter, M. (2010), ‘Technology clubs, R&D and growth patterns: Evidence from EU manufacturing’, *European Economic Review* **54**(1), 60–79.
- Bos, J. W. B., Economidou, C., Koetter, M. and Kolari, J. W. (2010), ‘Do all countries grow alike?’, *Journal of Development Economics* **91**(1), 113–127.
- Chen, Y. and Liang, K.-Y. (2010), ‘On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems’, *Biometrika* **97**(3), 603–620.
- Clark, D. H. (2003), ‘Can strategic interaction divert diversionary behavior? A model of U.S. conflict propensity’, *The Journal of Politics* **65**(4), 1013–1039.
- Coelli, T. J. (1995), ‘Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis’, *Journal of Productivity Analysis* **6**(3), 247–268.
- Greene, W. H. (1990), ‘A gamma-distributed stochastic frontier model’, *Journal of Econometrics* **46**(1/2), 141–164.
- Greene, W. H. (2005), ‘Reconsidering heterogeneity in panel data estimators of the stochastic frontier model’, *Journal of Econometrics* **126**(2), 269–303.
- Gurmu, S. and Trivedi, P. K. (1996), ‘Excess zeros in count models for recreational trips’, *Journal of Business and Economic Statistics* **14**(4), 469–477.
- Harris, M. N. and Zhao, X. (2007), ‘A zero-inflated ordered probit model, with an application to modelling tobacco consumption’, *Journal of Econometrics* **141**, 1073–1099.
- Horrace, W. and Schmidt, P. (2000), ‘Multiple comparisons with the best, with economic applications’, *Journal of Applied Econometrics* **15**(1), 1–26.
- Jondrow, J., Lovell, C. A. K., Materov, I. S. and Schmidt, P. (1982), ‘On the estimation of technical efficiency in the stochastic frontier production function model’, *Journal of Econometrics* **19**(2/3), 233–238.
- Kaparakis, E. I., Miller, S. M. and Noulas, A. (1994), ‘Short-run cost-inefficiency of commercial banks: A flexible stochastic frontier approach’, *Journal of Money, Credit and Banking* **26**, 875–893.
- Kumbhakar, S. C. and Lovell, C. A. K. (2000), *Stochastic Frontier Analysis*, Cambridge University Press, New York, NY.
- Kumbhakar, S. C. and Tsionas, E. (2005), ‘Measuring technical and allocative inefficiency in the translog cost system: A Bayesian approach’, *Journal of Econometrics* **126**, 355–384.
- Lambert, D. (1992), ‘Zero-inflated Poisson regression, with an application to defects in manufacturing’, *Technometrics* **34**(1), 1–14.
- Lee, L.-F. (1993), ‘Asymptotic distribution of the maximum likelihood estimator for a stochastic frontier function model with singular information matrix’, *Econometric Theory* **9**, 413–430.

- Li, Q. (1996), 'Nonparametric testing of closeness between two unknown distribution functions', *Econometric Reviews* **15**, 261–274.
- Lord, D., Washington, S. and Ivan, J. N. (2005), 'Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory', *Accident Analysis & Prevention* **37**, 35–46.
- Lord, D., Washington, S. and Ivan, J. N. (2007), 'Further notes on the application of zero-inflated models in highway safety', *Accident Analysis & Prevention* **39**, 53–57.
- Materov, I. S. (1981), 'On full identification of the stochastic production frontier model', *Ekonomika i Matematicheskie Metody* **17**, 784–788 (in Russian).
- McLachlan, G. J. and Peel, D. (2000), *Finite Mixture Models*, Wiley, New York, NY.
- Meeusen, W. and van den Broeck, J. (1977), 'Efficiency estimation from Cobb-Douglas production functions with composed error', *International Economic Review* **18**(2), 435–444.
- Orea, L. and Kumbhakar, S. C. (2004), 'Efficiency measurement using a latent class stochastic frontier model', *Empirical Economics* **29**(1), 169–183.
- Potts, J. M. and Elith, J. (2006), 'Comparing species abundance models', *Ecological Modelling* **199**, 153–163.
- Rho, S. and Schmidt, P. (2012), Are all firms inefficient? Paper presented at North American Productivity Workshop VII, Rice University, Houston, TX, June 6-9.
- Sickles, R. and Qian, J. (2009), Stochastic frontiers with bounded inefficiency. Rice University Working paper.
- Stevenson, R. E. (1980), 'Likelihood functions for generalized stochastic frontier estimation', *Journal of Econometrics* **13**(1), 57–66.

FIGURE 1. Comparison of average inefficiency (from each simulation) densities for the standard and ZISF when the true DGP is a standard SF.

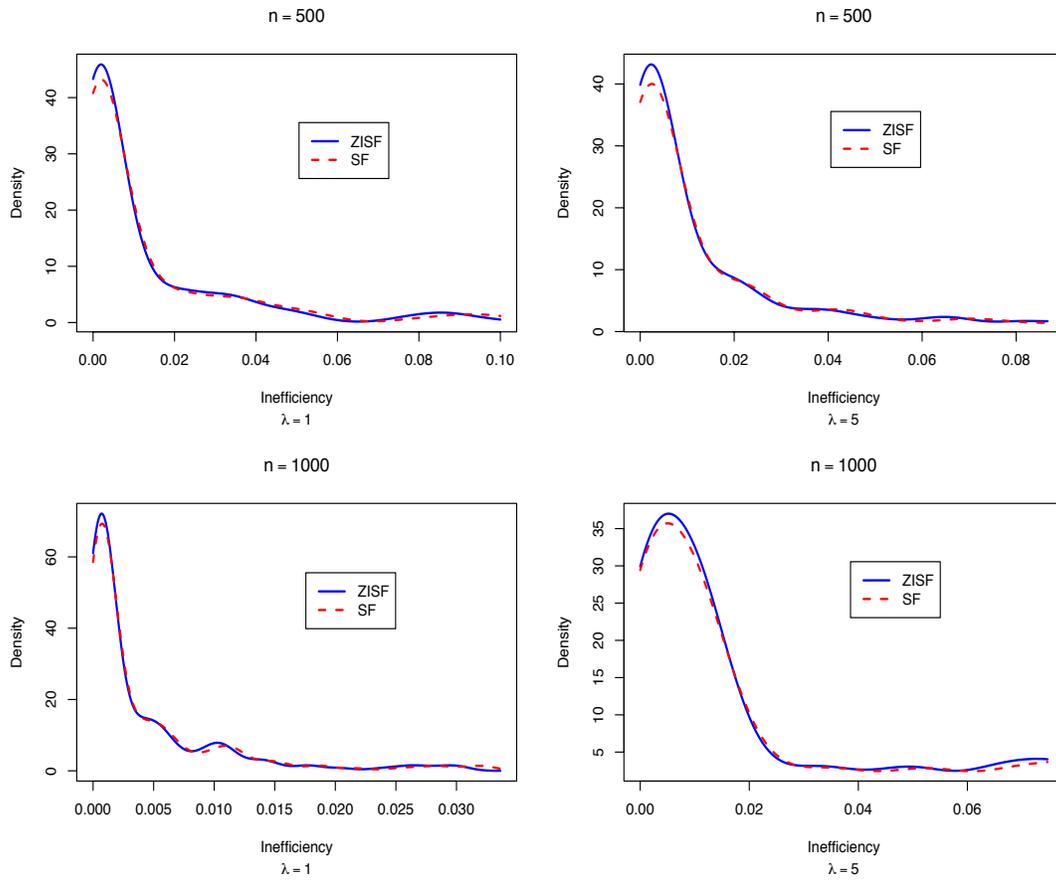


FIGURE 2. Sampling distributions of estimated technical inefficiency. The true DGP is a ZISF and a usual SF is used to estimate the model. — $\sim p = 0$, - - - - $\sim p = 0.10$, $\cdots \cdots p = 0.25$. Sampling distributions are estimated as simulation averages of the sample mean JLMS technical efficiency estimator for 1,000 simulations. The true DGP is ZISF with $p = 0, p = 0.1$ or $p = 0.25$ and maximum likelihood is used to produce estimates for the adopted normal/half-normal model for $\lambda = 1$ (left panel) and $\lambda = 5$ (right panel). The upper panel corresponds to $n = 500$, and the lower panel to $n = 1,000$.

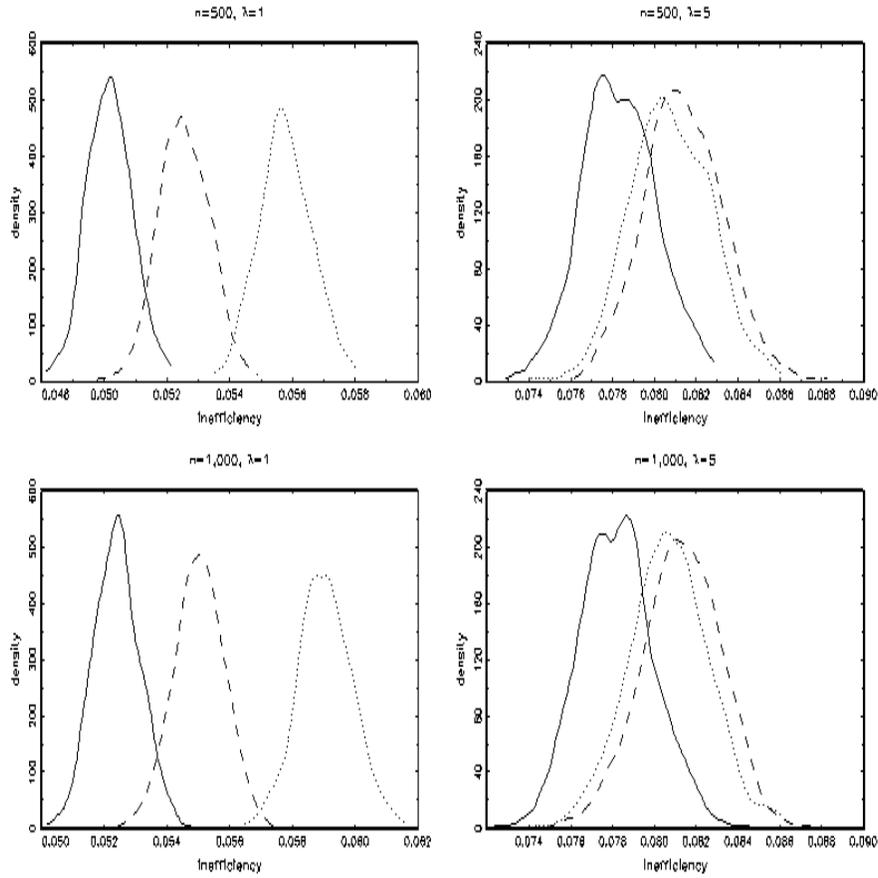


FIGURE 3. Sampling distributions of the estimated parameters from the ZISF model. The true DGP process is a ZISF model, the sample size $n = 1,000$, $\sigma = 0.1$, $\lambda = 5$ and $p = 0.25$. ML estimation is applied over 1,000 Monte Carlo replications. Vertical lines denote the true value of the parameter (or zero in the case of the upper left figure). Here p_{ML} , σ_{ML} and λ_{ML} denote ML estimates of p , σ and λ , respectively.

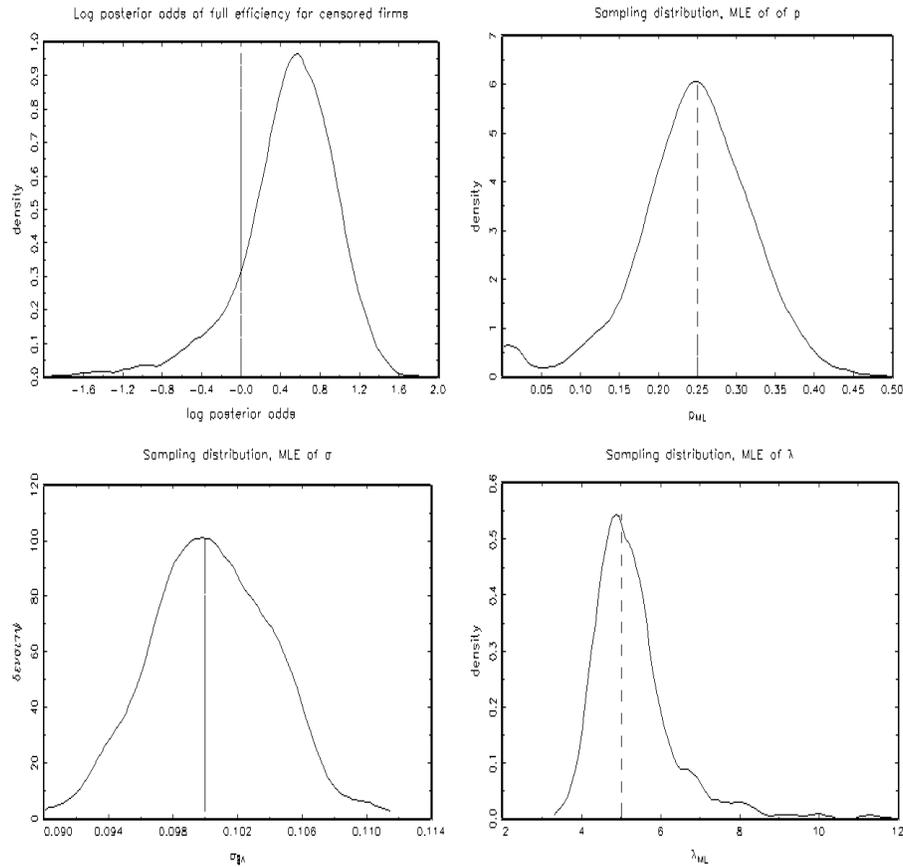


FIGURE 4. Histograms of Inefficiency and Posterior Probabilities for both the SF and ZISF maximum likelihood estimators.

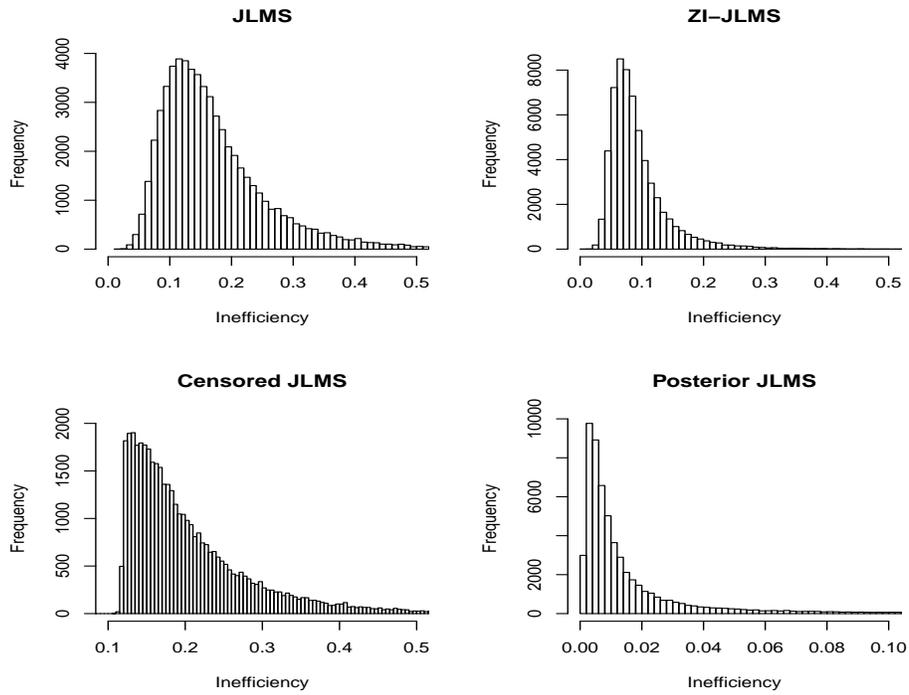


FIGURE 5. Densities of returns to scale and technical change for the standard SF and ZISF models.

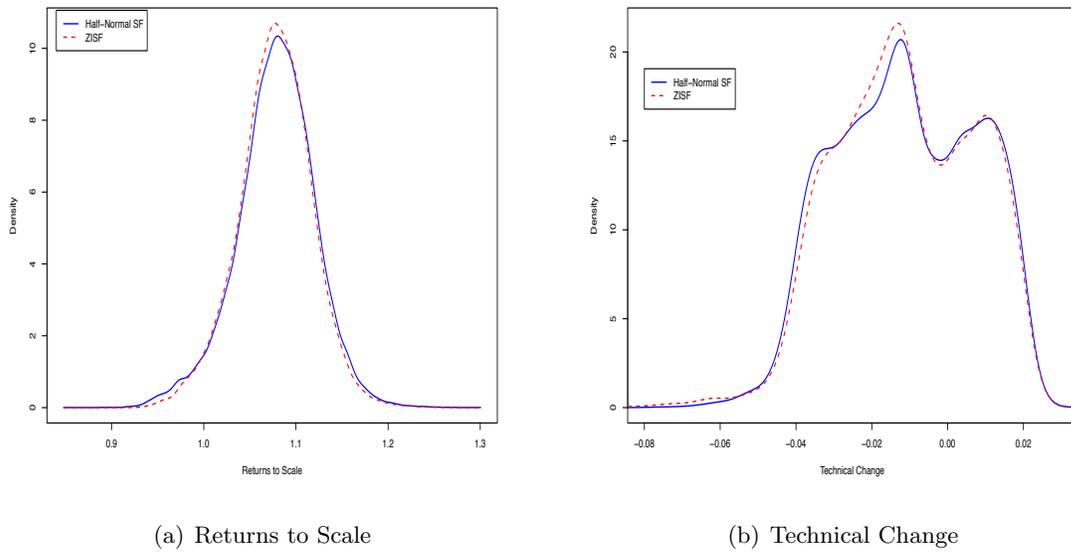


TABLE 1. Monte Carlo simulated critical values for the PLR statistic testing full efficiency in the ZISF.

n	90%	95%	99%
50	1.9261	3.6418	6.0827
100	1.6296	2.8421	6.0516
200	1.6624	2.7063	6.1835
500	1.6085	2.7042	5.0382
Asymp. Crit	1.6420	2.7043	5.4133

TABLE 2. Parameter estimates for the noise components of the SF and ZISF models. Asymptotic p -values are in parentheses beneath each estimate. The asymptotic p -value for the estimate of p is based on the PLR statistic described in section 2.3.

	Standard SF	ZISF
σ_v	0.151 (0.000)	0.174 (0.000)
σ_u	0.217 (0.000)	0.300 (0.000)
p	—	0.358 (0.000)

TABLE 3. Summary statistics for various measures of inefficiency. We report the upper and lower deciles (D_{90} , D_{10}), the quartiles (Q_{25} , Q_{50} , Q_{75}) as well as the mean and standard deviation (SD).

	D_{10}	Q_{25}	Q_{50}	Mean	Q_{75}	D_{90}	SD
Standard JLMS	0.085	0.111	0.151	0.172	0.208	0.286	0.091
Censored JLMS	0.130	0.147	0.181	0.207	0.238	0.319	0.087
ZI-JLMS	0.050	0.062	0.080	0.093	0.108	0.149	0.050
Posterior JLMS	0.003	0.004	0.009	0.029	0.020	0.056	0.071