Regularized conditional estimators of unit inefficiency in stochastic frontier analysis, with application to electricity distribution market

Zangin Zeebari [1,2] · Kristofer Månsson[1] · Pär Sjölander[1] · Magnus Söderberg[3,4,5]

Abstract

In stochastic frontier analysis, the conventional estimation of unit inefficiency is based on the mean/mode of the inefficiency, conditioned on the composite error. It is known that the conditional mean of inefficiency shrinks towards the mean rather than towards the unit inefficiency. In this paper, we analytically prove that the conditional mode cannot accurately estimate unit inefficiency, either. We propose regularized estimators of unit inefficiency that restrict the unit inefficiency estimators to satisfy some a priori assumptions, and derive the closed form regularized conditional mode estimators for the three most commonly used inefficiency densities. Extensive simulations show that, under common empirical situations, e.g., regarding sample size and signal-to-noise ratio, the regularized estimators outperform the conventional (unregularized) estimators when the inefficiency is greater than its mean/mode. Based on real data from the electricity distribution sector in Sweden, we demonstrate that the conventional conditional estimators and our regularized conditional estimators provide substantially different results for highly inefficient companies.

Keywords Uncertainty modelling · Productivity · Regularized Estimators · Constrained Estimators · Conditional Estimators

JEL C01 · C12 · C51 · H21

1 Introduction

Since the publication of the papers by Aigner et al. (1977) and Meeusen and van den Broeck (1977), stochastic frontier analysis (SFA) has been a common approach to gain deeper insights into the potential for productivity improvement

✉ Zangin Zeebari
zangin.zeebari@ju.se

1 Department of Economics, Finance and Statistics, Jönköping International Business School (JIBS), Jönköping University, P.O. Box 1026, SE-551 11 Jönköping, Sweden

2 Department of Global Public Health, Karolinska Institutet, Stockholm, Sweden

3 Halmstad University, Kristian IV:s väg 3, 301 18 Halmstad, Sweden

4 Ratio Institute, Box 3203, S-103 64 Stockholm, Sweden

5 Centre for Applied Energy Economics and Policy Research, Griffith University, 226 Grey St, South Bank, Brisbane, QLD 4101, Australia

(Kumbhakar et al. 2020) and cost reduction in monopolized markets (Bogetoft and Otto 2011). For unit inefficiency, the standard estimation approach was developed by Jondrow et al. (1982), acronymed "JLMS" in the SFA literature. The JLMS estimator is based on the mean (and the mode) of the inefficiency conditioned on the composite error (for later studies of JLMS estimators, see Kumbhakar and Lovell 2000, and Battese and Coelli 1988).

Despite its widespread use, the JLMS estimator has been criticized. Wang and Schmidt (2009) show that it shrinks the inefficiency towards its mean, leading to a distribution that is different from the distribution of the inefficiency. The mean and mode are not fully representative characteristics of the conditional distribution of the inefficiency, especially if each unit is observed once. Thus, in cross-sectional context, conditional estimator produces an inconsistent estimator of the inefficiency. Moreover, a conditional estimator is conditioned on an estimated composite error rather than on the composite error itself (Horrace 2005 and Kumbhakar et al. 2015). Therefore, the sampling distribution of the conditional estimator is different from the theoretically assumed conditional distribution of the

inefficiency. Consequently, the inefficiencies are inaccurately estimated, something regulatory agencies have stated as an impediment for the practical use of SFA (Badunenko et al. 2012; Stone 2002 and Tsionas 2017). This is also illustrated in a simulation study by Andor et al. (2019), where they show that both the SFA and Data envelopment analysis (DEA) methods used by regulators underestimate the true efficiency values. One way to reduce this problem is to combine the SFA and the DEA (Andor et al. 2019 and Tsionas 2021), but such combinatory approaches are not able to eliminate the underestimation problem.

Our approach is similar to the combinatory approach in that it can be viewed as a weighted average of unit inefficiency estimators, but in contrast, it is a weighted average of the sample (industry), solely based on the SFA approach. The proposed regularized estimators can be used as standalone estimators along with any other estimators in a combinatory approach. In addition, the regularized estimators described here can be used in a variety of situations but in this paper, we limit ourselves to studying unit inefficiency estimation in a cross-sectional context, using the classical stochastic frontier model suggested by Aigner et al. (1977).

We propose a regularized (constrained) estimator based on Bayesian risk (expected loss) that restricts the inefficiencies to satisfy some underlying theoretical (and/or intuitive) conditions. Restrictions on the moments are common options for the imposed constraints upon the likelihood functions (e.g., Hall and Presnell 1999). Our regularized estimators are easily calculated, e.g., they can be the JLMS (unregularized) estimators, with imposed constraints on the first and the second moments of the conditional distribution of the inefficiency.

The proposed methodology is different from other recent contributions in the field. For example, Kumbhakar et al. (1991) suggest a single step procedure for the estimation of unit inefficiency when they deploy firm-specific determinants of the inefficiency in the maximum likelihood estimation of the SFA model. They show that ignoring the determinants would lead to biased and inconsistent estimators. However, firm-specific determinants are often unobserved, and even unknown. Another recent contribution is the use of non-parametric and semi-parametric estimation methods, in which the JLMS estimator is used for estimating unit inefficiency (Kumbhakar et al. 2007). Another avenue of research is the use of quantile regression into the estimation of the production function (Bernini et al. 2004; Wang et al. 2014 and Behr 2010). However, this approach introduces a new challenge, specifically that one needs to pay more attention to the selection of appropriate quantiles which can be different for distinct densities of the composite error (Jradi et al. 2019). In addition, no sound and clear post-estimation method for estimating unit

inefficiency exists when using quantile regression (Kumbhakar et al. 2020).

Under mild assumptions, e.g., log-concavity of the distribution, which covers many of the distributions used in the SFA literature, we analytically investigate some properties of the conditional mode (maximum a posteriori probability estimator) and give a general formula for the conditional mode (and its functions) that can be used with any inefficiency density. Next, we derive a regularized conditional mode estimator with the three most commonly used inefficiency densities, i.e., the half-normal, truncated normal and exponential distributions. The proposed unit inefficiency estimation is considered a restricted or penalized estimation method that improves the estimation of unit inefficiency based on the conditional mean/mode. Therefore, the regularized estimators can be easily derived wherever the JLMS (unregularized) estimators are used. Therefore, the idea of our manuscript can be easily used in any context where the unit inefficiency estimation is calculated based on the optimization of an objective function, by imposing some restrictions to improve the accuracy of the unit inefficiency estimator.

An extensive simulation study is conducted, with varying factors, such as the sample size, inefficiency density and signal-to-noise ratio (relative variation of the inefficiency to the variation of random shocks). The simulation results show that the regularized estimators outperform the conventional (unregularized) estimators when the inefficiencies are greater than their mean/mode, especially with a larger signal-to-noise ratio. As the unregularized conditional mean/mode shrinks towards the mean/mode, the regularized conditional mean/mode shrinks less towards the mean/mode, especially for larger inefficiencies.

We apply both unregularized and regularized estimators to data from the Swedish electricity distribution sector. The results show that the estimated inefficiencies from the two regularized and unregularized estimators are substantially different, particularly for firms that are in the right tail of the inefficiency distribution. Considering the results from the simulations which are supported analytically (Theorem 3), we recommend that regulators use the results from the regularized estimators for the firms that are ranked highly inefficient, based on their estimated inefficiency scores using the conditional mean/mode.

The remainder of this paper is structured as follows. In Section 2, we derive a general formula for the conditional mode of the inefficiency and analytically investigate some of its properties under mild distributional assumptions of unconditional inefficiency. Next, the regularized estimator is discussed and formally derived for both production and cost functions. It is analytically shown that the regularized conditional mode estimators serve to reduce the shrinkage towards the mode. Regularized conditional mode estimators

are presented under three different distributional assumptions. In Section 4, both regularized and unregularized estimators are evaluated using extensive Monte Carlo simulations. In Section 5, we present an application based on real data. The data represent electricity distribution firms in Sweden, and we estimate the cost inefficiency, which is used by the Energy Markets Inspectorate as an input in their revenue cap regulation. Section 6 concludes the paper and discusses avenues for future research.

2 Theory

A stochastic frontier, cross-sectional, production model can be formulated as

$$y_i = \mathbf{x}'_i\beta + v_i - u_i \tag{1}$$

where i indicates the unit, y_i is the observed output, x_i is the given $k \times 1$ vector of inputs, u_i is the unobserved inefficiency, v_i is the unobserved noise and β is an unknown $k \times 1$ vector of functional parameters.

The conventions of a simple parametric cross-section SFA assume i.i.d. random noise terms with a density function $g_v(v)$ that is symmetric around zero and i.i.d. nonnegative inefficiencies with a density function $f_u(u)$. For example, the most common (semistandard) $g_v(v)$ is assumed to be the density of a zero-mean normal distribution $N(0, \sigma_v^2)$[1], and the equivalent candidates for $f_u(u)$ are assumed to be the densities of a half-normal distribution $N^+(0, \sigma_u^2)$, an exponential distribution $\text{Exp}(\sigma_u)$ with scale parameter σ_u, and a truncated normal distribution $N^+(\mu, \sigma_u^2)$ with a general μ that can take any real number.

The maximum likelihood estimation of an SFA model is based on maximizing the likelihood of the i.i.d. composite errors $\varepsilon_i = v_i - u_i$ with the density function

$$h_\varepsilon(\varepsilon) = \int_0^{+\infty} f_u(u)\, g_v(u + \varepsilon)\, du$$

where the composite error ε_i is $\varepsilon_i = y_i - \mathbf{x}'_i\beta$.

It has been argued (e.g., Greene 1990 and Ruggiero 1999), that the selection of different inefficiency density functions should not result in noticeable differences between the fit of the SFA models, or the ranks of the estimated conditional unit inefficiencies. However, they may differ in the magnitude of the estimated inefficiencies, especially for highly inefficient units.

As mentioned in the Introduction, the most common way of scoring the unit inefficiency is through the JLMS

estimators. For the ith unit, the inefficiency is estimated as $\widehat{u}_i = E(u|\varepsilon_i)$ or $\widehat{u}_i = Mode(u|\varepsilon_i)$ using the following conditional density function of inefficiency u given a composite error ε.

$$f_{u|\varepsilon}(u) = \frac{f_u(u)\, g_v(u + \varepsilon)}{h_\varepsilon(\varepsilon)}$$

However, as stated by Kumbhakar et al. (2020), the conditional estimator of the inefficiency is an estimator of a characteristic (mean or mode) of the conditional inefficiency rather than of the inefficiency itself. Such a distinction between the two remains unchanged regardless of the sample size. In fact, it depends on the size of the noise variance rather than on the sample size. This fact is proven by Wang and Schmidt (2009) for the conditional mean when the inefficiency follows a half-normal distribution, and they argue that it also holds when the inefficiencies are drawn from exponential and general truncated normal distributions. However, such argument has not been proven for the conditional mode, although there is a general belief in the SFA literature that the JLMS estimators, whether mean or mode, are shrinkage estimators.

In Theorem 3, we provide a proof that, under mild distributional assumptions, the conditional mode of the inefficiency analogously shrinks towards the mode of the inefficiency. This means that the conditional mode estimator underestimates large inefficiencies. It also overestimates the inefficiencies of almost fully efficient firms when the inefficiency mode is a positive number (as it is the case for a truncated normal distribution with location parameter $\mu > 0$).

The conditional mode estimator is the maximum a posteriori probability estimator, which is the mode of the a posteriori distribution, i.e., $Mode(u|\varepsilon_i) = \text{Argmax}_{u \in \mathbb{R}^+} f_{u|\varepsilon}(u) = \text{Argmax}_{u \in \mathbb{R}^+} (f_u(u)g_v(u + \varepsilon))$. In Theorem 1, we give a general formula to calculate the conditional mode of the inefficiency for any inefficiency density function that fulfills the mild assumptions stated below. We take smooth density functions $f_u(u)$ whose supports are the real numbers in the range $u \in [0, \infty)$. For such densities, we define the "*form*" (kernel) of the density, denoted as $K(w)$, for $w \in \mathbb{R}$, as follows,

$$f(u) = \begin{cases} cK(u), & \text{if } u \geq 0 \\ 0, & \text{if } u < 0 \end{cases}$$

where, c is a constant and possibly an expression containing the parameters of the density function, but not a function of the variable u itself. Note that the form of the density is in fact the kernel of the density defined for both the support of the density and its complement in the set of the real numbers. First, we present the following corollary.

---

[1] Other zero-mean symmetric distributions have been suggested, such as Laplace (Horrace and Parmeter, 2018; Nguyen, 2010) and t-distribution by Tancredi (2002) and Wheat et al. (2019).

Corollary: Suppose the function $K(w) > 0$ is defined for all $w \in \mathbb{R}$, with the following properties:

1. $K(w)$ is continuously differentiable for all $w \in \mathbb{R}$, and log-concave.
2. $\frac{K'(w)}{K(w)}$ is bounded above, such that $\lim_{w \to +\infty} \frac{K'(w)}{K(w)} < |M| < +\infty$.
3. $\frac{K'(w)}{K(w)}$ is bounded below, such that $\lim_{w \to -\infty} \frac{K'(w)}{K(w)} > -|m| > -\infty$.

For the convergence of the integral $\int_{-\infty}^{+\infty} e^{-s\left[w - \frac{K'(w)}{K(w)}\right]} dw$, with $l < s < r$, we must have $\lim_{w \to \pm\infty} \left| \frac{\frac{K'(w)}{K(w)}}{w} \right| < 1$.

We give the detailed proof in the Appendix. The results of the following theory are based on the results of the Corollary.

Theorem 1: Suppose the noise of the production function in (1) is $v \sim N(0, \sigma_v^2)$, and the inefficiency has the density $f_u(u)$, for $u \geq 0$. If $f_u(u)$ has the form $K(w)$ which is continuously differentiable, with its first derivative $K'(w)$, such that $\lim_{w \to \pm\infty} \frac{K'(w)}{K(w)}$ is bounded, then the inefficiency mode conditioned on the composite error ε, denoted as $\hat{u} = Mode(u|\varepsilon)$, is unique. In addition, if

$$\lim_{\tilde{u} \to \pm\infty} \left| \frac{\frac{K'(\tilde{u})}{K(\tilde{u})}}{\tilde{u}} \right| < \frac{1}{\sigma_v^2}$$

then the conditional mode is $\hat{u} = Mode(u|\varepsilon) = \max\{0, \tilde{u}\}$, where

$$\tilde{u} = -\varepsilon - \sum_{k=1}^{+\infty} \frac{(\sigma_v^2)^k}{k!} \left[\frac{\partial}{\partial \varepsilon} \right]^{k-1} \left\{ -\frac{K'(-\varepsilon)}{K(-\varepsilon)} \right\}^k$$

Proof:
Obviously, $\hat{u} = Mode(u|\varepsilon) = \text{Argmax}_{u \in \mathbb{R}^+} (f_u(u)g_v(u + \varepsilon)) = \max\{0, \tilde{u}\}$, where $\tilde{u}$ is defined as

$$\tilde{u} = \underset{u \in \mathbb{R}}{\text{Argmax}} (K(u)g_v(u + \varepsilon))$$

By solving the above maximization problem, we have

$$K'(\tilde{u})\, g_v(\tilde{u} + \varepsilon) + g_v'(\tilde{u} + \varepsilon)\, K(\tilde{u}) = K'(\tilde{u})\, g_v(\tilde{u} + \varepsilon)$$
$$- \frac{(\tilde{u} + \varepsilon)}{\sigma_v^2} g_v(\tilde{u} + \varepsilon)\, K(\tilde{u}) = 0$$

$$\tilde{u} = -\varepsilon + \sigma_v^2 \frac{K'(\tilde{u})}{K(\tilde{u})} \qquad (2)$$

With $f_u(\tilde{u})$ a log-concave density, $\frac{d \ln[f_u(\tilde{u})]}{du} = \frac{f_u'(\tilde{u})}{f_u(\tilde{u})} = \frac{K'(u)}{K(u)}$ is a decreasing function of $u \geq 0$. Thus, there is a unique solution of $\tilde{u}$ in terms of bounded ε in (2), since $\sigma_v^2 \frac{K'(\tilde{u})}{K(\tilde{u})} - \tilde{u} = \varepsilon$ also becomes a monotonically decreasing function of $\tilde{u} \geq 0$.

For the Lagrange reversion theorem (see Whittaker and Watson (1927), pp. 132–133, and Grossman (2005)[2]) of Eq. (2), we start with the bilateral Laplace transform of the function $\tilde{u}'(-\varepsilon) = \frac{d\tilde{u}}{d(-\varepsilon)} = -\frac{d\tilde{u}}{d\varepsilon}$. Suppose for a complex number s whose real part is in the interval $k < Re(s) < r$, the bilateral Laplace transform exists, as follows.

$$\mathcal{L}\left\{ \frac{d\tilde{u}}{d(-\varepsilon)} \right\}(s) = \int_{-\infty}^{+\infty} e^{-s(-\varepsilon)} \frac{d\tilde{u}}{d(-\varepsilon)} d(-\varepsilon)$$
$$= \int_{-\infty}^{+\infty} e^{-s\left[\tilde{u} - \sigma_v^2 \frac{K'(\tilde{u})}{K(\tilde{u})}\right]} d\tilde{u} = \int_{-\infty}^{+\infty} e^{-s\tilde{u}} e^{s\sigma_v^2 \frac{K'(\tilde{u})}{K(\tilde{u})}} d\tilde{u}$$

For the convergence of the above integral, we must have the following condition (see Corollary above):

$$\lim_{\tilde{u} \to \pm\infty} \left| \frac{\frac{K'(\tilde{u})}{K(\tilde{u})}}{\tilde{u}} \right| < \frac{1}{\sigma_v^2}$$

Returning to the above integral, and using Taylor expansion of $s\sigma_v^2 \frac{K'(\tilde{u})}{K(\tilde{u})}$, we have

$$\mathcal{L}\left\{ \frac{d\tilde{u}}{d(-\varepsilon)} \right\}(s) = \int_{-\infty}^{+\infty} e^{-s\tilde{u}} \left[\sum_{k=0}^{+\infty} \frac{s^k}{k!} \left(\sigma_v^2 \frac{K'(\tilde{u})}{K(\tilde{u})} \right)^k \right] d\tilde{u}$$
$$= \sum_{k=0}^{+\infty} \left[\int_{-\infty}^{+\infty} e^{-s\tilde{u}} \frac{s^k}{k!} \left(\sigma_v^2 \frac{K'(\tilde{u})}{K(\tilde{u})} \right)^k d\tilde{u} \right]$$

Applying the properties of bilateral Laplace transform $\left(\mathcal{L}\{f^{(n)}\}(s) = s^n \mathcal{L}\{f\}(s) \right)$ to the integral inside the summation, the summation is equal to

$$= \sum_{k=0}^{+\infty} \left[\int_{-\infty}^{+\infty} \frac{1}{k!} e^{-s\tilde{u}} \frac{d^k}{du^k} \left(\sigma_v^2 \frac{K'(\tilde{u})}{K(\tilde{u})} \right)^k d\tilde{u} \right]$$
$$= \int_{-\infty}^{+\infty} e^{-s\tilde{u}} \left[\sum_{k=0}^{+\infty} \frac{1}{k!} \cdot \frac{d^k}{du^k} \left(\sigma_v^2 \frac{K'(\tilde{u})}{K(\tilde{u})} \right)^k \right] d\tilde{u}$$

By changing the notation of the last integral and due to the uniqueness of bilateral Laplace transform,

$$\frac{d\tilde{u}}{d(-\varepsilon)} = \sum_{k=0}^{+\infty} \frac{(\sigma_v^2)^k}{k!} \frac{d^k}{d(-\varepsilon)^k} \left(\frac{K'(-\varepsilon)}{K(-\varepsilon)} \right)^k$$

$$\frac{d\tilde{u}}{d(-\varepsilon)} = 1 + \sum_{k=1}^{+\infty} \frac{(\sigma_v^2)^k}{k!} \frac{d^k}{d\varepsilon^k} \left(-\frac{K'(-\varepsilon)}{K(-\varepsilon)} \right)^k$$

$$\frac{d\tilde{u}}{d\varepsilon} = -1 - \sum_{k=1}^{+\infty} \frac{(\sigma_v^2)^k}{k!} \frac{d^k}{d\varepsilon^k} \left(-\frac{K'(-\varepsilon)}{K(-\varepsilon)} \right)^k$$

---

[2] Contrast to Grossman (2005) in which the proof is based on unilateral Laplace and the necessary condition that $f_u'(0) = 0$, in our proof such condition $f_u'(0) = 0$ is not necessary since we use bilateral Laplace transform.

Solving the above differential equation gives us the solution

$$\tilde{u} = -\varepsilon - \sum_{k=1}^{+\infty} \frac{(\sigma_v^2)^k}{k!} \frac{d^{k-1}}{d\varepsilon^{k-1}} \left(-\frac{K'(-\varepsilon)}{K(-\varepsilon)} \right)^k$$

This completes the proof. ∎

Note that the commonly used inefficiency densities of half normal, exponential, general truncated normal and gamma (with shape parameter ≥1) are log-concave distributions. When the noise v and the inefficiency u are distributed as assumed in Theorem 1, for each of the density and distribution functions $q \in \left\{ \tilde{f}_{\widehat{u}|u}, \tilde{F}_{\widehat{u}|u}, \tilde{f}_{\widehat{u}|\varepsilon}, \tilde{F}_{\widehat{u}|\varepsilon}, \tilde{f}_{u|\widehat{u}}, \tilde{F}_{u|\widehat{u}}, f_u, F_u, \tilde{f}_{\widehat{u}}, \tilde{F}_{\widehat{u}} \right\}$, and any other differentiable function of $\widehat{u} \geq 0$, we have,

$$q(\widehat{u}) = q(-\varepsilon) - \sum_{k=1}^{+\infty} \frac{(\sigma_v^2)^k}{k!} \left[\frac{\partial}{\partial\varepsilon} \right]^{k-1} \left\{ \left(-\frac{f_u'(-\varepsilon)}{f_u(-\varepsilon)} \right)^k q'(-\varepsilon) \right\}$$

When the $[\ln(f_u(u))]' = \frac{d\ln[f_u(u)]}{du}$ is linear, as it is the case for the half-normal, exponential and general truncated normal densities, the conditional mode, in terms of the mode of inefficiency m, is simplified as

$$\widehat{u} = Mode(u|\varepsilon) = \frac{-\varepsilon + \sigma_v^2[\ln(f_u(m))]' - \sigma_v^2 m[\ln(f_u(m))]''}{1 - \sigma_v^2[\ln(f_u(m))]''}$$

where, $[\ln(f_u(m))]'$ and $[\ln(f_u(m))]''$ are the first and second derivatives evaluated at the mode of the inefficiency (m). In general, the conditional mode, theoretically and empirically, is less covered in the SFA literature when JLMS estimators are used, in favor of the conditional mean. To the best of authors' knowledge, the article by Papadopoulos (2021) is an exception, in that the author elaborates on the conditional model and proves its monotonicity in terms of the composite error when the inefficiency follows a generalized exponential distribution. Monotonicity of the conditional mode in terms of the composite error is important in that both (mean/mode) JLMS estimators must rank the unit inefficiencies identically. If so, using the conditional mode, the inefficiencies can be ranked based on their corresponding composite errors, i.e., the coefficient of ranked correlation between the conditional mode scores and the composite errors becomes almost negative one. A similar argument holds for conditional mean scores, as shown by Bera and Sharma (1999) and Ondrich and Ruggiero (2001). In Theorem 2, we show that under mild distributional assumptions, the monotonicity of the conditional mode in terms of the composite error is generalizable to any other inefficiency distribution.

Theorem 2: Suppose $v \sim N(0, \sigma_v^2)$. The inefficiency density $f_u(u)$ is nonzero, twice differentiable and log-concave at $u \geq 0$. The conditional mode estimator of the inefficiency conditioned on the composite error ε is a monotonically decreasing function of the composite error.

Proof:

Since $f_u(u)$ is log-concave, we have $\frac{d^2 \ln[f_u(u)]}{(du)^2} \leq 0$ for all $u \geq 0$. As shown in Theorem 1, we can write Eq. (2) as $\tilde{u} = -\varepsilon + \sigma_v^2 \frac{d\ln[f_u(\tilde{u})]}{d\tilde{u}}$. Then, by the chain rule of derivatives, we have

$$\frac{\partial \tilde{u}}{\partial \varepsilon} = -1 + \sigma_v^2 \frac{d^2 \ln[f_u(\tilde{u})]}{(d\tilde{u})^2} \frac{\partial \tilde{u}}{\partial \varepsilon} = \frac{-1}{1 - \sigma_v^2 \frac{d^2 \ln[f_u(\tilde{u})]}{(d\tilde{u})^2}} < 0$$

The above negative derivative would imply strict monotonicity if negative scores ($\tilde{u}$) were acceptable. Since they are restricted to $\widehat{u} = \max\{0, \tilde{u}\}$, monotonicity is not strict, in general. Thus, the proof is complete. ∎

Wang and Schmidt (2009) show that the conditional mean shrinks towards the mean of inefficiency rather than towards inefficiency itself. This property is disadvantageous to the unit inefficiencies that depart from the mean since it underestimates highly inefficient firms and overestimates the inefficiencies lower than the mean. It is also a disadvantage of the conditional mean for regulators to accurately estimate the inefficiency in the lower and, especially, in the upper tail of the inefficiency distribution. Although being able to rank the units based on their inefficiencies is of regulators' interest, in some cases the magnitude of the inefficiency is of crucial importance, for instance, EU countries' (in)efficiencies in their climate plans to cut emissions of greenhouse gases.

In Theorem 3, we prove that the conditional mode has a similar property, in that it is a shrinkage estimator towards the mode of inefficiency rather than towards inefficiency itself. With such property, although the conditional mode would outperform the conditional mean in estimating the lower tail of an inefficiency distribution with its mode in a narrow positive neighborhood of zero, it is still a poor estimator for highly inefficient firms, i.e., the right tail of the distribution.

Theorem 3: Suppose $v \sim N(0, \sigma_v^2)$ and, the inefficiency density $f_u(u)$ is nonzero, twice differentiable, log-concave for $u > 0$ and with $m = Mode(u) = \text{Argmax}_{u \in \mathbb{R}^+} f_u(u)$. Let the conditional mode estimator of the inefficiency be $\widehat{u} = Mode(u|\varepsilon)$. Then,

a) as $\sigma_v^2 \to 0$, $\widehat{u} \to_p u$,
b) as $\sigma_v^2 \to 0$, $\widehat{u} \to_d u$,
c) as $\sigma_v^2 \to 0$, $\frac{\widehat{u}-u}{\sigma_v} \to_d N(0,1)$,
d) as $\sigma_v^2 \to \infty$, $\widehat{u} \to_p m = Mode(u)$.
e) as $\sigma_v^2 \to \infty$, $\sigma_v^2[\ln(f_u(m))]' + (\sigma_v^2[\ln(f_u(m))]'' - 1)$ $(\widehat{u} - m) \to_d (\varepsilon + m)$.

Proof:

By assumption, $f_u(u)$ is differentiable and nonzero, for $u > 0$. Then, $\frac{d\ln[f_u(u)]}{du}$ is bounded (differentiable functions

have their derivatives bounded). Then, we can write
$$\tilde{u} = \underset{u \in \mathbb{R}^+}{\text{Argmax}}(f_u(u) \cdot g_v(u + \varepsilon)),$$

$$f_u'(\tilde{u}) \, g_v(\tilde{u} + \varepsilon) + f_u(\tilde{u}) \, g_v'(\tilde{u} + \varepsilon) = 0$$

$$\frac{1}{\sigma_v^2} = \frac{1}{\tilde{u} + \varepsilon} \frac{d \ln[f_u(\tilde{u})]}{d\tilde{u}}$$

In addition, as $\sigma_v^2 \to 0$, the normal density tends to Dirac's delta function,[3] with its mass concentrated around the mean, i.e., $g_v(v) \to_d \delta(E(v))$. Then, $v \to_p E(v) = 0$. (Note that it might be that $\tilde{u} < 0$, then the conditional mode is $\hat{u} = \max\{0, \tilde{u}\}$).

a) As $\sigma_v^2 \to 0$, $v \to_p E(v) = 0$ (Dirac's delta function). Then $\varepsilon \to_p -u$ (since $\varepsilon = v - u$), or equivalently $-\varepsilon \to_p u$. Additionally, as $\sigma_v^2 \to 0$, $\frac{1}{u+\varepsilon} \frac{d \ln[f_u(\tilde{u})]}{du} \to \infty$ (due to the fact that $\frac{1}{\sigma_v^2} = \frac{1}{u+\varepsilon} \frac{d \ln[f_u(\tilde{u})]}{du}$). Since $\frac{d \ln[f_u(u)]}{du}$ is bounded, then, for a bounded composite error, it must be that $\tilde{u} \to -\varepsilon$. It means $\tilde{u} \to_p u$. Then, $\hat{u} = \max\{0, \tilde{u}\} \to_p \max\{0, u\} = u$.

b) Although the convergence in probability, as shown in point (a), automatically implies the convergence in distribution, another direct proof, independent from the result of point (a) above, can be as follows.

$$\varepsilon = \sigma_v^2 \frac{f_u'(\tilde{u})}{f_u(\tilde{u})} - \tilde{u}$$

Then,

$$\left| \frac{d\varepsilon}{d\tilde{u}} \right| = \left| \sigma_v^2 \frac{d^2 \ln[f_u(\tilde{u})]}{(d\tilde{u})^2} - 1 \right|$$

$$\tilde{f}_{\tilde{u}}(\tilde{u}) = h_\varepsilon\left(\sigma_v^2 \frac{f_u'(\tilde{u})}{f_u(\tilde{u})} - \tilde{u} \right) \left| \frac{d\varepsilon}{d\tilde{u}} \right| = \left| \sigma_v^2 \frac{d^2 \ln[f_u(\tilde{u})]}{(d\tilde{u})^2} - 1 \right|$$
$$\times \int_0^{+\infty} f_u(u) \, g_v\left(u + \sigma_v^2 \frac{f_u'(\tilde{u})}{f_u(\tilde{u})} - \tilde{u} \right) du$$

As $\sigma_v^2 \to 0$,

$$\tilde{f}_{\tilde{u}}(\tilde{u}) \to_d \lim_{\sigma_v^2} h_\varepsilon\left(\sigma_v^2 \frac{f_u'(\tilde{u})}{f_u(\tilde{u})} - \tilde{u} \right) \left| \frac{d\varepsilon}{d\tilde{u}} \right|$$
$$= \lim_{\sigma_v^2} \left| \sigma_v^2 \frac{d^2 \ln[f_u(\tilde{u})]}{(d\tilde{u})^2} - 1 \right| \int_0^{+\infty} f_u(u) \, g_v\left(u + \sigma_v^2 \frac{f_u'(\tilde{u})}{f_u(\tilde{u})} - \tilde{u} \right) du$$

$$= \int_0^{+\infty} f_u(u) \, \delta(u - \tilde{u}) du = f_u(\tilde{u})$$

As $\sigma_v^2 \to 0$, $\Pr(\tilde{u} < 0) \to 0$, and $\tilde{f}_{\tilde{u}}^*(u) = \frac{\tilde{f}_{\tilde{u}}(u)}{1 - \Pr(\tilde{u} < 0)} \to \tilde{f}_{\tilde{u}}(u)$, meaning that $\tilde{f}_{\tilde{u}}^*(u) \to f_u(u)$, for $u \geq 0$.

c) Since $v \sim N\left(0, \sigma_v^2\right)$, then $-\frac{v}{\sigma_v} \sim N(0, 1)$. It means

$$-\frac{v}{\sigma_v} = -\frac{u + \varepsilon}{\sigma_v} = \frac{\tilde{u} - u - \sigma_v^2 \frac{f_u'(\tilde{u})}{f_u(\tilde{u})}}{\sigma_v} \sim N(0, 1)$$

As $\sigma_v^2 \to 0$, $\frac{\tilde{u} - u - \sigma_v^2 \frac{f_u'(\tilde{u})}{f_u(\tilde{u})}}{\sigma_v} \to_p \frac{\tilde{u} - u}{\sigma_v}$, then $\frac{\tilde{u} - u}{\sigma_v} \to_d N(0, 1)$. From point b, as $\sigma_v^2 \to 0$, $\Pr(\tilde{u} < 0) \to 0$, and we have $\hat{u} = \max\{0, \tilde{u}\}$, i.e., as $\sigma_v^2 \to 0$, $\hat{u} \to_p \tilde{u}$. Then, $\frac{\hat{u} - u}{\sigma_v} \to_p \frac{\tilde{u} - u}{\sigma_v}$, meaning that $\frac{\hat{u} - u}{\sigma_v} \to_d N(0, 1)$.

d) Since we have $\frac{1}{\sigma_v^2} = \frac{1}{\tilde{u} + \varepsilon} \frac{d \ln[f_u(\tilde{u})]}{d\tilde{u}}$, for a bounded value of ε, as $\sigma_v^2 \to \infty$, it implies two possibilities. First, if $\frac{d \ln[f_u(\tilde{u})]}{d\tilde{u}} \to 0$, then $\tilde{u} \to Mode(u)$ since $f_u(u)$ is unimodal (log-concave). Whether the $Mode(u)$ is zero or a positive number, we have $\hat{u} = max\{0, \tilde{u}\} \to Mode(u)$. Second, if $\frac{d \ln[f_u(\tilde{u})]}{d\tilde{u}} \not\to 0$, then we must have $\frac{d \ln[f_u(\tilde{u})]}{d\tilde{u}} < 0$ and the density must be a monotonically decreasing function of $u \geq 0$, since for $\tilde{u} \to +\infty$, $\frac{d \ln[f_u(\tilde{u})]}{d\tilde{u}} \not> 0$. Thus, it must be that $\tilde{u} \to -\infty$ (with a bounded ε), which is restricted to $\hat{u} = max\{0, \tilde{u}\} = 0$. In such case, $\hat{u} = max\{0, \tilde{u}\} = 0$ is again the mode of u since $f_u(u)$ (that cannot have $\frac{d \ln[f_u(u)]}{du} = 0$, for any $u > 0$, according to the second possibility) must be strictly monotonically decreasing with its $Mode(u) = 0$.

e) As $\sigma_v^2 \to \infty$, we can use the fact in point d and the mean value theorem around the mode m to write $\lim_{\sigma_v^2 \to \infty} \frac{[\ln(f_u(\tilde{u}))]' - [\ln(f_u(m))]'}{\tilde{u} - m} = [\ln(f_u(m))]''$, since $\lim_{\sigma_v^2 \to \infty} (\tilde{u} - m) = 0$. Therefore, using this fact and the Eq. (2), $\tilde{u} = -\varepsilon + \sigma_v^2 \frac{d \ln[f_u(\tilde{u})]}{d\tilde{u}}$, we have $\lim_{\sigma_v^2 \to \infty} \tilde{u} = \lim_{\sigma_v^2 \to \infty} \frac{-\varepsilon + \sigma_v^2 [\ln(f_u(m))]' - \sigma_v^2 m [\ln(f_u(m))]''}{1 - \sigma_v^2 [\ln(f_u(m))]''}$. This means, as $\sigma_v^2 \to \infty$, $\sigma_v^2 [\ln(f_u(m))]' + \left(\sigma_v^2 [\ln(f_u(m))]'' - 1 \right)(\hat{u} - m) \to_d (\varepsilon + m)$. ∎

Note that in point e of Theorem 3, for half normal and general truncated normal densities, the first derivate of the log density evaluated at the mode is $[\ln(f_u(m))]' = 0$, while for the exponential density, only the second derivate at the mode is $[\ln(f_u(m))]'' = 0$.

A question that might arise is, does a larger sample prevent the shrinkage of the JLMS estimators? In cross-sectional context, the simple answer to this question is 'no' in the literature, for several reasons. First, since the inefficiencies are unobservable, the conditional estimators cannot be improved by learning from more data (in contrast to regression models). Second, the productivity of each unit is observed only once, therefore, due to the assumption of independence between the units, conditional estimator of each unit inefficiency is conditioned on a single composite error which corresponds to the unit itself. Third, due to lack

---

[3] see Horrace and Wright (2020), for a similar argument, when the half normal inefficiency density tends to Dirac's delta function as $\sigma_u^2 \to 0$.

of replication, the JLMS estimator is based on a guess (a typical value, like the mean or the mode) from the conditional distribution of the inefficiency, conditioned on a single composite error. Therefore, inconsistency and high uncertainty of the JLMS estimators are expected in the cross-sectional context. In econometrics literature, it is known that regularization increases the accuracy of an estimator by reducing its variance. The accuracy of a regularized estimator is due to a trade-off between decreased variance and increased bias.

3 Regularization

It has been shown in the literature that a maximum likelihood estimator is improved by maximizing an *a posteriori* or regularized (penalized) likelihood function; see for example, Cox and O'Sullivan (1990) and Flynn et al. (2013). One can consider the conditional mode and the conditional mean from Bayes expected loss and the Bayes risk minimization perspective. For example, for the conditional mean, the loss function is $(u - \widehat{u})^2$, whose risk minimization yields $\text{Argmin}_{\widehat{u} \in \mathbb{R}^+} E\{(u - \widehat{u})^2 | \varepsilon\} = E(u | \varepsilon)$. For the conditional mode, the loss function is a zero-one indicator function $(I(u \neq \widehat{u}) - 1)$, whose risk minimization yields $\text{Argmin}_{\widehat{u} \in \mathbb{R}^+} E\{I(u \neq \widehat{u}) - 1 | \varepsilon\} = \text{Argmin}_{\widehat{u} \in \mathbb{R}^+} [-f_{u|\varepsilon}(\widehat{u})] = \text{Argmax}_{\widehat{u} \in \mathbb{R}^+} f_{u|\varepsilon}(\widehat{u}) = Mode(u | \varepsilon)$.

A regularization of the risk minimization is achieved by adding extra information to, or imposing more constraints on, the risk function (expected loss). Suppose the constraints are a set of m zero-equality equations of twice differentiable functions $R(u)$, i.e., $R(u) = 0_{m \times 1}$. The regularized conditional mean of the inefficiency is the solution to the following constrained objective function.

$$\min_{\widehat{u} \in \mathbb{R}^+} E\{(u - \widehat{u})^2 | \varepsilon\}$$

Subject to : $\boldsymbol{R}(\widehat{u}) = 0$

The solution is $\text{Argmin}_{\widehat{u} \in \mathbb{R}^+} E\{(u - \widehat{u})^2 | \varepsilon\} + \lambda' \boldsymbol{R}(\widehat{u})$, where λ is the vector of Lagrange multipliers. The regularized conditional mean is the solution to the following system of equations.

$$\begin{cases} \widehat{u} - E(u | \varepsilon) + 0.5 \lambda' \nabla R(\widehat{u}) = 0 \\ R(\widehat{u}) = 0 \end{cases}$$

For the conditional mode of the inefficiency, the objective function and the constraints are as follows.

$$\max_{\widehat{u} \in \mathbb{R}^+} f_{u|\varepsilon}(\widehat{u})$$

Subject to : $\boldsymbol{R}(\widehat{u}) = 0$

The regularized conditional mode is the solution to the following system of equations.

$$\begin{cases} f'_{u|\varepsilon}(\widehat{u}) + \lambda' \nabla R(\widehat{u}) = 0 \\ R(\widehat{u}) = 0 \end{cases}$$

The regularized JLMS estimators can be developed for both the mean and the mode. However, in the next section, we develop only the regularized conditional mode estimators for the three most commonly used inefficiency densities, which are the half normal, exponential and general truncated normal. In fact, the idea of a restricted estimator of the unit inefficiency can be easily generalized to any method that estimates the unit inefficiency based on the optimization of an objective functions (e.g., methods based on the determinants of inefficiency, such as Wang and Schmidt; 2002, Tran and Tsionas; 2009, Simar et al. 2017, and Parmeter et al. 2017). This can be done by imposing proper restrictions on the optimization of the objective function that is used for estimating the unit inefficiency.

Restricted moments are common imposed constraints on the likelihood functions (e.g., Hall and Presnell 1999). The constraints can, for instance, be on the sum of the inefficiencies or the sum of squared inefficiencies. These are constraints on the first and second moments of the inefficiencies, respectively. Based on the result of point e from Theorem 3, we can write

$$\lim_{\sigma_v^2 \to \infty} (\tilde{u} - m) = \lim_{\sigma_v^2 \to \infty} \frac{-(\varepsilon + m) + \sigma_v^2 [\ln(f_u(m))]'}{1 - \sigma_v^2 [\ln(f_u(m))]''}$$

$$= -\frac{[\ln(f_u(m))]'}{[\ln(f_u(m))]''}$$

which shows the shrinkage of the conditional mode towards the mode in response to the noise variance inflation. For half normal and general truncated normal densities, the limit is zero. With an exponential density, the $\tilde{u}$ is restricted to be non-negative, i.e., $\widehat{u} = \max\{0, \tilde{u}\}$ and $m = 0$. Then, $\lim_{\sigma_v^2 \to \infty} (\widehat{u} - m) = 0$. With a vector of constraints $\boldsymbol{R}(\widehat{u}) = 0$, the mode-difference limit becomes

$$\lim_{\sigma_v^2 \to \infty} (\tilde{u} - m) = \lim_{\sigma_v^2 \to \infty} \frac{-(\varepsilon + m) + \sigma_v^2 [\ln(f_u(m))]' + \sigma_v^2 \lambda' \nabla \boldsymbol{R}(\widehat{u})}{1 - \sigma_v^2 [\ln(f_u(m))]''}$$

$$= -\frac{[\ln(f_u(m))]' + \lambda' \nabla \boldsymbol{R}(\widehat{u})}{[\ln(f_u(m))]''}$$

Therefore, the above limit can be different from zero with properly selected constraints. This means that regularization can serve to reduce the shrinkage of the conditional mode estimator towards the mode. For the developed regularized conditional modes in Table 1, this fact is

Table 1 The unit inefficiency estimator $\hat{u} = \max\{0, -\tilde{u}\}$, with $\tilde{u}$ given in the cells of the following table

Measure	Half Normal, $u_i \sim N^+(0, \sigma_u^2)$	Truncated Normal, $u_i \sim N^+(\mu, \sigma_u^2)$	Exponential, $u_i \sim Exp\left(\sqrt{\sigma_u^2}\right)$
Unregularized Mode $(u|\varepsilon_i)$	$\frac{\sigma_u^2 \varepsilon}{\sigma_v^2 + \sigma_u^2}$	$\frac{\sigma_u^2 \varepsilon - \mu \sigma_v^2}{\sigma_v^2 + \sigma_u^2}$	$\varepsilon + \frac{\sigma_v^2}{\sqrt{\sigma_u^2}}$
Mode $(u|\varepsilon_i)$ with 1st Moment Constraint	$\frac{\sigma_u^2 \varepsilon + \sigma_v^2 \bar{\varepsilon}}{\sigma_v^2 + \sigma_u^2}$	$\frac{\sigma_u^2 \varepsilon + \sigma_v^2 \bar{\varepsilon}}{\sigma_v^2 + \sigma_u^2}$	ε
Mode $(u|\varepsilon_i)$ with 1st & 2nd Moment Constraints	$\dfrac{\varepsilon + \bar{\varepsilon}\left(\sqrt{\frac{\sigma_\varepsilon^2}{\sigma_u^2 - \bar{\varepsilon}^2}} - 1\right)}{\sqrt{\frac{\sigma_\varepsilon^2}{\sigma_u^2 - \bar{\varepsilon}^2}}}$	$\dfrac{\varepsilon_i + \bar{\varepsilon}\left(\sqrt{\frac{\sigma_\varepsilon^2}{E(u^2) - \bar{\varepsilon}^2}} - 1\right)}{\sqrt{\frac{\sigma_\varepsilon^2}{E(u^2) - \bar{\varepsilon}^2}}}$	$\dfrac{\varepsilon + \bar{\varepsilon}\left(\sqrt{\frac{\sigma_\varepsilon^2}{2\sigma_u^2 - \bar{\varepsilon}^2}} - 1\right)}{\sqrt{\frac{\sigma_\varepsilon^2}{2\sigma_u^2 - \bar{\varepsilon}^2}}}$

obvious. For example, for the general truncated normal, the conditional mode is $\frac{-\sigma_u^2 \varepsilon + \mu \sigma_v^2}{\sigma_v^2 + \sigma_u^2}$, which shrinks toward the mode μ with noise variance inflation, i.e., $\lim_{\sigma_v^2 \to \infty} \frac{-\sigma_u^2 \varepsilon + \mu \sigma_v^2}{\sigma_v^2 + \sigma_u^2} = \mu$. But, its regularized version based on the first moment restriction is $-\frac{\sigma_u^2 \varepsilon + \sigma_v^2 \bar{\varepsilon}}{\sigma_v^2 + \sigma_u^2}$, and its limit when $\sigma_v^2 \to \infty$ is $\lim_{\sigma_v^2 \to \infty} -\frac{\sigma_u^2 \varepsilon + \sigma_v^2 \bar{\varepsilon}}{\sigma_v^2 + \sigma_u^2} = -\bar{\varepsilon} \neq \mu$. The same argument holds for the other two cases (half normal and exponential densities and their conditional modes and regularized conditional modes). Therefore, a weaker shrinkage property of the regularized estimators, especially for highly inefficient units, is expected. This fact is also supported by the simulation results in Section 4, as expected.

4 First- and second-moment constraints

Inconsistency and high uncertainty of the JLMS estimators are expected in the cross-sectional context since any JLMS estimator of a unit inefficiency is conditioned on a single composite error corresponding to the unit itself. For each unit inefficiency estimation, we can also exploit extra information from other composite errors. For example, one can impose a restriction on all estimated inefficiencies such that their sample mean equals the sample mean of the composite errors. Such a restriction is equivalent to a sample zero-mean constraint on the estimated random shocks.

In terms of economic theory, the zero-mean random shock constraint is interpreted as a condition where the unit's productivity is invariant to the random shocks in the market. Let us take a production frontier model such as the Cobb-Douglas or a translog model with the inefficiency as the single source of shortfall.

$$y_i = f(\boldsymbol{x}_i; \beta) e^{-u_i}$$

If firm i experiences a random shock (v_i), its production can expand or shrink, depending on the sign of v_i.

$$y_i = f(\boldsymbol{x}_i; \beta) e^{-u_i} e^{v_i}$$

Random shocks for some units can cover part of their inefficiencies, while for others, they might worsen their productivities, depending on whether the random shocks and the firm specific inefficiencies are in the same or opposite directions. An assumption of the SFA model in (1) is the zero-mean random shocks with a normal distribution. A similar restriction can also be imposed on the estimated random shocks, such that their sample mean is equal to zero. This is equivalent to the counterfactual assumption that if a firm were consecutively exposed to the shocks from the whole market, its productivity would eventually return to the same level, since it is assumed that the sum of the random socks in the market is equal to zero.

$$\begin{aligned} y_i &= f(\boldsymbol{x}_i; \beta) \cdot e^{-u_i} = f(\boldsymbol{x}_i; \beta) \cdot e^{-u_i} \cdot e^{v_1} \dots e^{v_i} \dots e^{v_n} \\ &= f(\boldsymbol{x}_i; \beta) \cdot e^{-u_i} e^{\sum_{i=1}^n v_i} = f(\boldsymbol{x}_i; \beta) \cdot e^{-u_i} \end{aligned}$$

Imposing the above-mentioned market (industry) shock-invariance assumption on the conditional mode of the inefficiency and the constraint that $\varepsilon_i = v_i - u_i$, for $i = 1, \dots, n$, with a sample of n units, is translated into the inefficiency sum (mean) restriction. The regularized conditional mode is the solution of the following constrained objective function.

$$\max_{u_1, \dots, u_n} \left\{ \sum_{i=1}^n \ln[g_v(u_i + \varepsilon_i)] + \ln[f_u(u_i)] \right\}$$

subject to:

$$\sum_{i=1}^n (u_i + \varepsilon_i) = 0$$

Using the Lagrange multiplier method, the above constrained objective function is written as

$$\begin{aligned} &m_0(u_1, \dots, u_n | \varepsilon_1, \dots, \varepsilon_n) \\ &= \operatorname*{Argmax}_{u_1, \dots, u_n} \left\{ \sum_{i=1}^n \ln[g_v(u_i + \varepsilon_i)] + \ln[f_u(u_i)] + \lambda \sum_{i=1}^n (u_i + \varepsilon_i) \right\} \end{aligned}$$

Then, the estimated inefficiencies are forced to fulfill the constraint $\sum_{i=1}^n v_i = 0$. We can extend the number of restrictions, for example, by adding a restriction on the

variance or the sum of squares of the estimated conditional modes, as follows.

$$\max_{u_1,\dots,u_n} \left\{ \sum_{i=1}^n \ln[g_v(u_i + \varepsilon_i)] + \ln[f_u(u_i)] \right\}$$

Subject to:

$$\sum_{i=1}^n (u_i + \varepsilon_i) = 0$$

$$\sum_{i=1}^n u_i^2 = c$$

The c on the right side of the second constraint can be, for example, $c = nE(u^2)$. With the Lagrange multiplier method, the above constrained objective function is written as

$$m_0(u_1,\dots,u_n|\varepsilon_1,\dots,\varepsilon_n) = \operatorname*{Argmax}_{u_1,\dots,u_n}$$

$$\left\{ \sum_{i=1}^n \ln[g_v(u_i + \varepsilon_i)] + \ln[f_u(u_i)] + \lambda \sum_{i=1}^n (u_i + \varepsilon_i) + \theta \left[\sum_{i=1}^n u_i^2 - c \right] \right\}$$

Table 1 shows the regularized conditional mode estimators for a production model. Note that; $\bar\varepsilon = \frac{\sum_{i=1}^n \varepsilon_i}{n}$ and $\hat\sigma_\varepsilon^2 = \frac{\sum_{i=0}^n (\varepsilon_i - \bar\varepsilon)^2}{n}$, and for $u_i \sim N^+(\mu, \sigma_u^2)$,

$$E(u^2) = \sigma_u^2 \left(1 + \frac{\mu}{\sigma_u} \frac{\phi(\mu/\sigma_u)}{\Phi(\mu/\sigma_u)} + \frac{\mu^2}{\sigma_u^2} \right)$$

For half-normal and exponential distributions, $E(u^2)$ is σ_u^2 and $2\sigma_u^2$, respectively. Thus, with the first- and second-moment constraints, $\tilde u$ has the same closed-form solution in terms of $E(u^2)$. The conditional mean $E(u|\varepsilon_i)$ with each of the densities shown in Table 1, has the following general form:

$$E(u|\varepsilon_i) = \tilde\sigma \frac{\phi(\tilde\mu/\tilde\sigma)}{\Phi(\tilde\mu/\tilde\sigma)} - \tilde\mu$$

where $\tilde\mu$ is the negative of the cells of the 1st row in Table 1 (corresponding to $-\tilde u$ in Theorem 1), and $\tilde\sigma = \sigma_v$ for the exponential density and $\tilde\sigma = \frac{\sigma_v \sigma_u}{\sigma}$ for each of the half-normal and truncated normal densities, where $\sigma^2 = \sigma_v^2 + \sigma_u^2$.

As stated in Theorem 3, the conditional mode of inefficiency shrinks towards the mode of inefficiency in response to any noise variance inflation. From Table 1, we realize that the regularized estimators (explicitly with the first-moment restriction) serve to hold the unit inefficiency estimators away from the inefficiency mode by adding fractions of the noise variance to the conditional mode estimators, i.e., they serve to reduce the shrinkage of the

conditional mode estimator towards the mode of inefficiency (0 or $\mu > 0$).

For each of the above inefficiency densities and the set of the constraints, the same estimators are developed for a cost function. To save space, they are not presented here, but they are obtained straightforwardly by altering the signs of ε and $\bar\varepsilon$ inside the above closed-form formulae in Table 1. The purpose of presenting regularized conditional mode estimators is to introduce the methodology with closed-form mathematical expressions. Analogous to the conditional mode, the methodology can also be applied to regularized conditional mean estimators, with properly selected constraints[4].

5 Simulations

An extensive simulation study is conducted to assess the performance of the proposed methodology relative to the JLMS estimators. The varying factors of the simulation study are (i) the sample size, (ii) the inefficiency distribution, (iii) the noise variance, (iv) the inefficiency variance and (iv) the location parameter when the inefficiency follows a truncated normal distribution.

In addition, the performance of the regularized estimators is assessed when the distribution is incorrectly specified. For this purpose, we see Gompertz distribution a suitable option, due to its desirable characteristics for being a probability distribution of inefficiency. More precisely, Gompertz distribution is a non-negative and log-concave distribution with non-zero density and with flexibility to be skewed to the left and to the right, and to have its mode equal to zero or to a positive number. For more on Gompertz distribution, see e.g., Lenart 2012.

Samples of size 20, 30, 50, 100 and 250 were simulated. A Cobb-Douglas production model was assumed, and each sample consisted of three simulated variables: production output and labor and capital inputs. The values for the intercept, elasticities, and means and variances of labor and capital were selected to imitate a production model originally used in Cobb-Douglas (1928). Specifically, the regression coefficients were selected as $\beta = \{-0.25, 0.25, 075\}$ and labor and capital were drawn from the bivariate normal distribution $\boldsymbol{x} \sim N_2([5.5, 5], \begin{bmatrix} 0.25 & 0 \\ 0 & 0.04 \end{bmatrix})$.

The noise from $N(0, \sigma_v^2)$ and inefficiencies from $N^+(0, \sigma_u^2)$, $N^+(\mu, \sigma_u^2)$, $Exp(\sigma_u)$, and $Gompertz(\eta, b)$ are

---

[4] For instance, a quadratic interpolated polynomial through the points 0, maximum deterministic inefficiency (the negative of the minimum composite error in a production function) and the estimated mean of inefficiency.

Table 2 Distributional Parameters for $Var(u) = 1 - \sigma_v^2$

Distribution	$Var(u) = 0.1$	$Var(u) = 0.5$	$Var(u) = 0.9$
$N^+(\mu, \sigma_u^2)$, with $\mu = -0.1$	$\sigma_u^2 = 0.3068604$	$\sigma_u^2 = 1.4465384$	$\sigma_u^2 = 2.5713615$
$N^+(\mu, \sigma_u^2)$, with $\mu = 0.1$	$\sigma_u^2 = 0.2440010$	$\sigma_u^2 = 1.3058240$	$\sigma_u^2 = 2.3825505$
$Gompertz(\eta, b)$, with $b = 1$	$\eta = 1.6707464$	$\eta = 0.2768683$	$\eta = 0.0805014$

drawn and used to simulate the production model. The noise variance was given the values of $\sigma_v^2 = \{0.1, 0.5, 0.9\}$ and the inefficiency variance $Var(u)$ was selected such that $Var(u) = 1 - \sigma_v^2$, i.e., the variance of the composite error was kept at unity with each simulated sample ($\sigma_\varepsilon^2 = 1$). For the truncated normal inefficiency, the variance is also affected by μ. In the simulations, $\mu = 0$ (for half normal) and $\mu = 0.1$, and (for truncated normal). We select the basic Gompertz distributions with their scale parameters equal to one ($b = 1$). Then, their shape parameters η are selected such that $Var(u) = 1 - \sigma_v^2$.

We randomly draw inefficiencies from basic *Gompertz(η, b)* distributions. However, to assess the performance of the regularized estimators of unit inefficiency with an incorrectly specified distribution, we use $N^+(0, \sigma_u^2)$, $N^+(\mu, \sigma_u^2)$ and $Exp(\sigma_u)$ when estimating the SFA model and the unit inefficiency scores. The results are shown in Fig. 4 for samples of size 100. The results are consistent with other sample sizes, too.

Distributional parameters must be properly selected for having $Var(u) = 1 - \sigma_v^2$. For $Exp(\sigma_u)$, we select $\sigma_u = \sqrt{1 - \sigma_v^2}$, and for $N^+(0, \sigma_u^2)$, we select $\sigma_u^2 = (1 - \sigma_v^2)/(1 - \frac{2}{\pi})$. However, for each of $N^+(\mu, \sigma_u^2)$ and $Gompertz(\eta, b)$, there is no closed form solution to select the proper distributional parameters. We found the parameters numerically, as shown in Table 2.

The simulations were implemented as follows. For each sample size, the simulated design matrix was fixed across all simulations. To assess the performance of each estimator across different ranks of inefficiency, we considered two different scenarios. The first scenario is to rank the firms constantly based on their inefficiencies such that the first simulated firm always receives the lowest simulated inefficiency, and the last simulated firm always receives the largest inefficiency. The second scenario is to randomly rank the firms based on their inefficiencies. The results of the two scenarios were consistent; hence, the second scenario was followed to avoid any potential effect due to differences in production input across the firms. This process was repeated 100 times, i.e., 100 samples of ranked inefficiencies were simulated from the above-mentioned inefficiency probability distributions. For each of the 100 samples of inefficiencies, 100 samples of noise terms were randomly generated from the normal distributions. This resulted in 10,000 replications for each of the 60

combinations of the above factors (sample size, probability distribution, σ_v^2, σ_u^2 and μ).

With each replication of the simulation process, the four measures of unit-level inefficiencies were calculated, which were the conditional mean, conditional mode, conditional mode with first-moment constraint and conditional mode with first- and second-moment constraint. The Mean Squared Error (MSE) for the ith firm's inefficiency was calculated as follows.

$$MSE(\widehat{u}_i) = \frac{\sum_{k=1}^{100} \sum_{j=1}^{100} (\widehat{u}_{ji} - u_{ki})^2}{10000}$$

The squared bias for the ith firm's inefficiency measure was calculated as follows.

$$Bias^2(\widehat{u}_i) = \frac{\sum_{k=1}^{100} \left(\frac{\sum_{j=1}^{100} \widehat{u}_{ji}}{100} - u_{ki} \right)^2}{100}$$

Each measure's relative efficiency to the conditional mean efficiency was calculated as

$$Relative\,MSE(\widehat{u}_i) = \frac{MSE(E(u|\varepsilon_i))}{MSE(\widehat{u}_i)}$$

In the above formulae, i represents the unit, j is the noise replication and k represents inefficiency replications. The results of the relative MSE are shown in Fig. 1, Fig. 2, Fig. 3 and Fig. 4. In the graphs, the x-axis represents the rank of the inefficiency[5] and the y-axis represents the relative MSE. All the simulations and calculations were run in STATA/SE 16 for Windows 64 bit using the *sfcross* command by Belotti et al. (2013).

The results of the simulations in Figs. 1–4 show that when estimating large inefficiencies, the regularized conditional mode estimator, especially with the first-moment constraint, outperforms the JLMS estimators as the signal-to-noise ratio (σ_u/σ_v) increases. While the signal-to-noise ratio seems to be more decisive for the relative performance of the regularized estimator than the sample size and distributional assumption, its performance improves further when inefficiencies are exponentially distributed and when the sample size is not very large. Also, the larger the signal-

---

[5] An alternative is to use the inefficiencies or technical efficiencies. The conclusion is the same but on different scales.
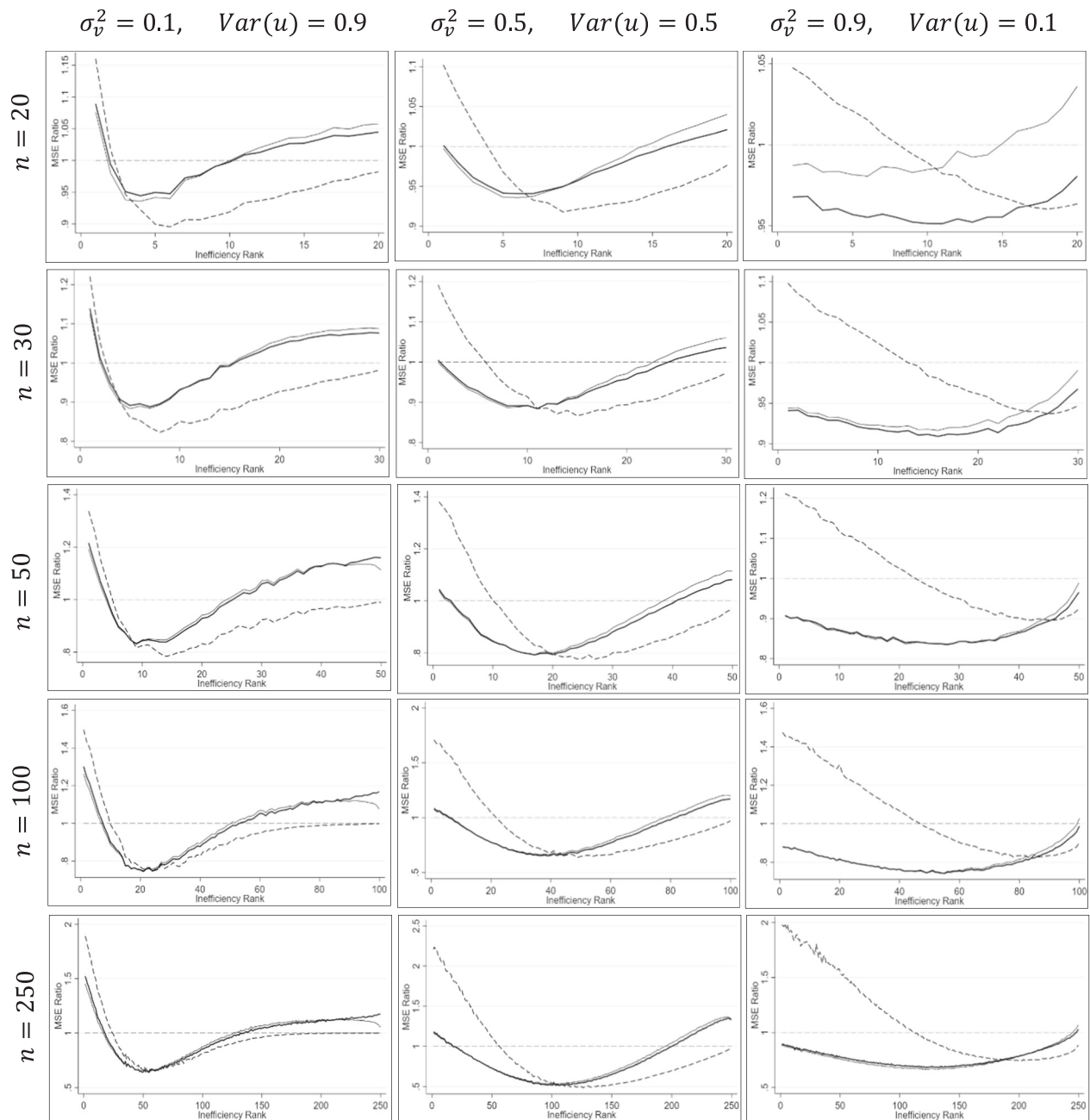
Fig. 1 Normal—Half-Normal Model: Relative Inefficiency (MSE Ratio) compared to $E(u|\varepsilon)$. Note: The larger the noise variance, the choppier the curves of the regularized conditional mode estimators are
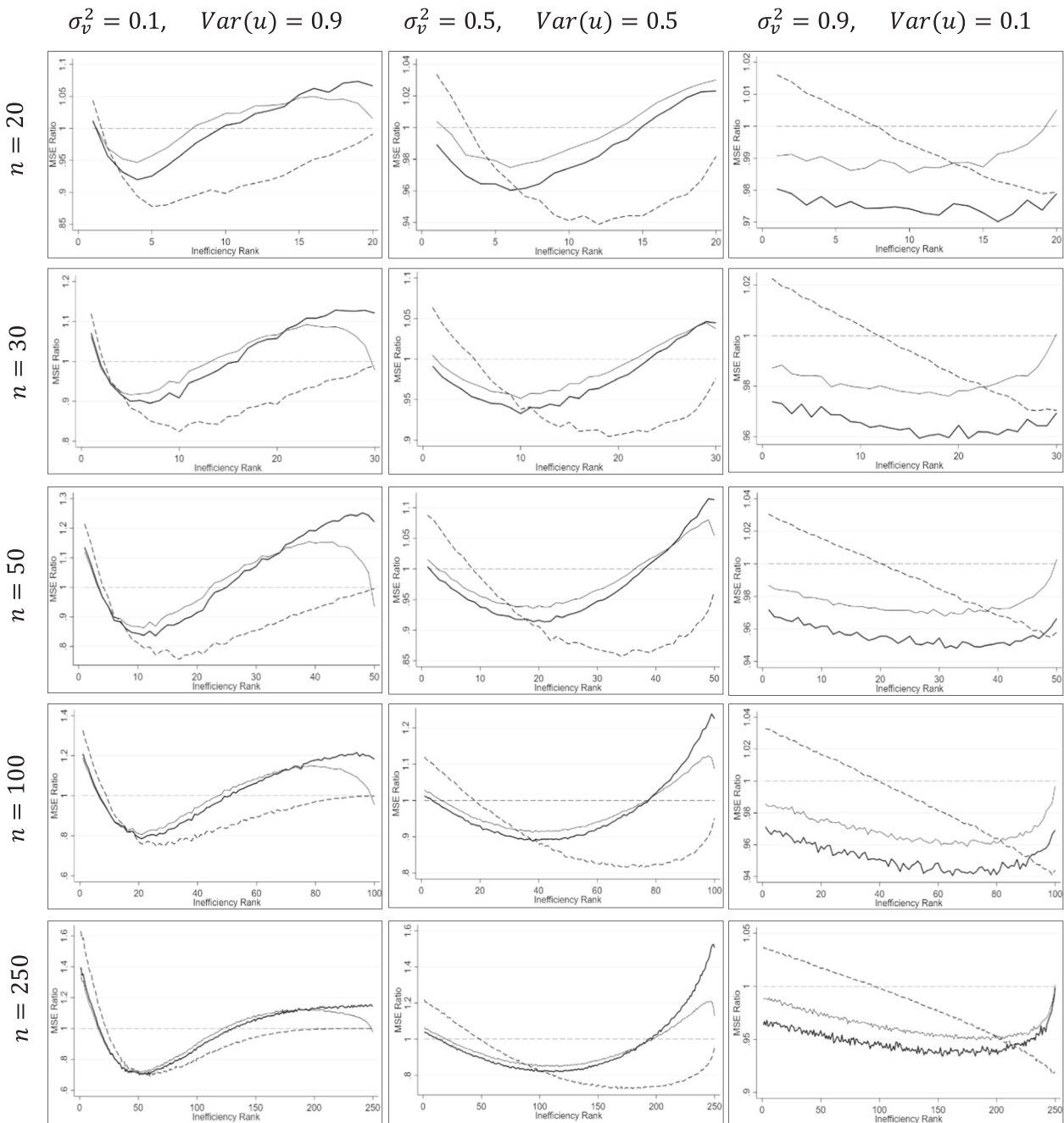
to-noise ratio, the more robust the regularized estimators are, if the distribution is incorrectly specified, especially for larger inefficiencies. Some points are listed as follows:

- The unregularized conditional mode is almost always the most accurate estimator for units with no or small

inefficiency— a result that is expected due to its shrinkage-towards-mode property. [6]

---

[6] The modes are zero (for half-normal and exponential distributions) and 0.1 for truncated normal distributions.

- - - - Conditional mean $E(u|\varepsilon)$, reference   - - - - Conditional $Mode(u|\varepsilon)$

——— $Mode(u|\varepsilon)$, restricted with $\sum u$ ——— $Mode(u|\varepsilon)$, restricted with $\sum u$ & $\sum u^2$
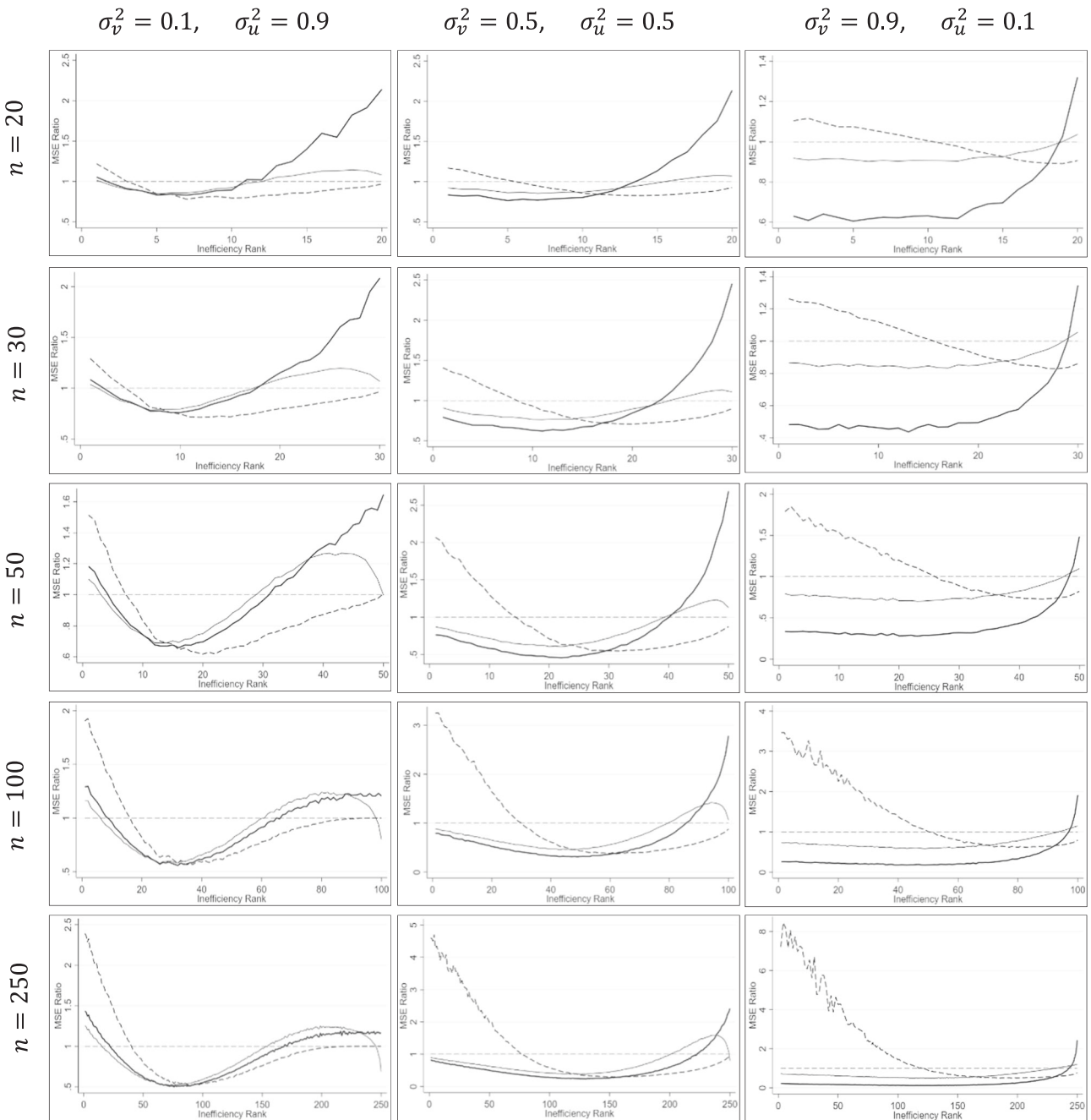
Fig. 2 Normal—Exponential Model: Relative Inefficiency (MSE Ratio) compared to $E(u|\varepsilon)$. Note: The larger the noise variance, the choppier the curve of the unregularized conditional mode is

- The unregularized conditional mean is the most accurate estimator of the unit inefficiency for middle ranks since it is a shrinkage estimator towards the mean.
- Regularized conditional mode estimators, especially the one with the first-moment constraint, are the most accurate estimators of unit inefficiencies that are more to

the right tail of the distribution (highly ranked), unless it is a case with low signal-to-noise ratio, in which the unconditional mode (for lower ranks) and unconditional mean (for higher ranks) outperform the regularized estimators. The same argument holds if the inefficiency distribution is incorrectly specified, especially when an
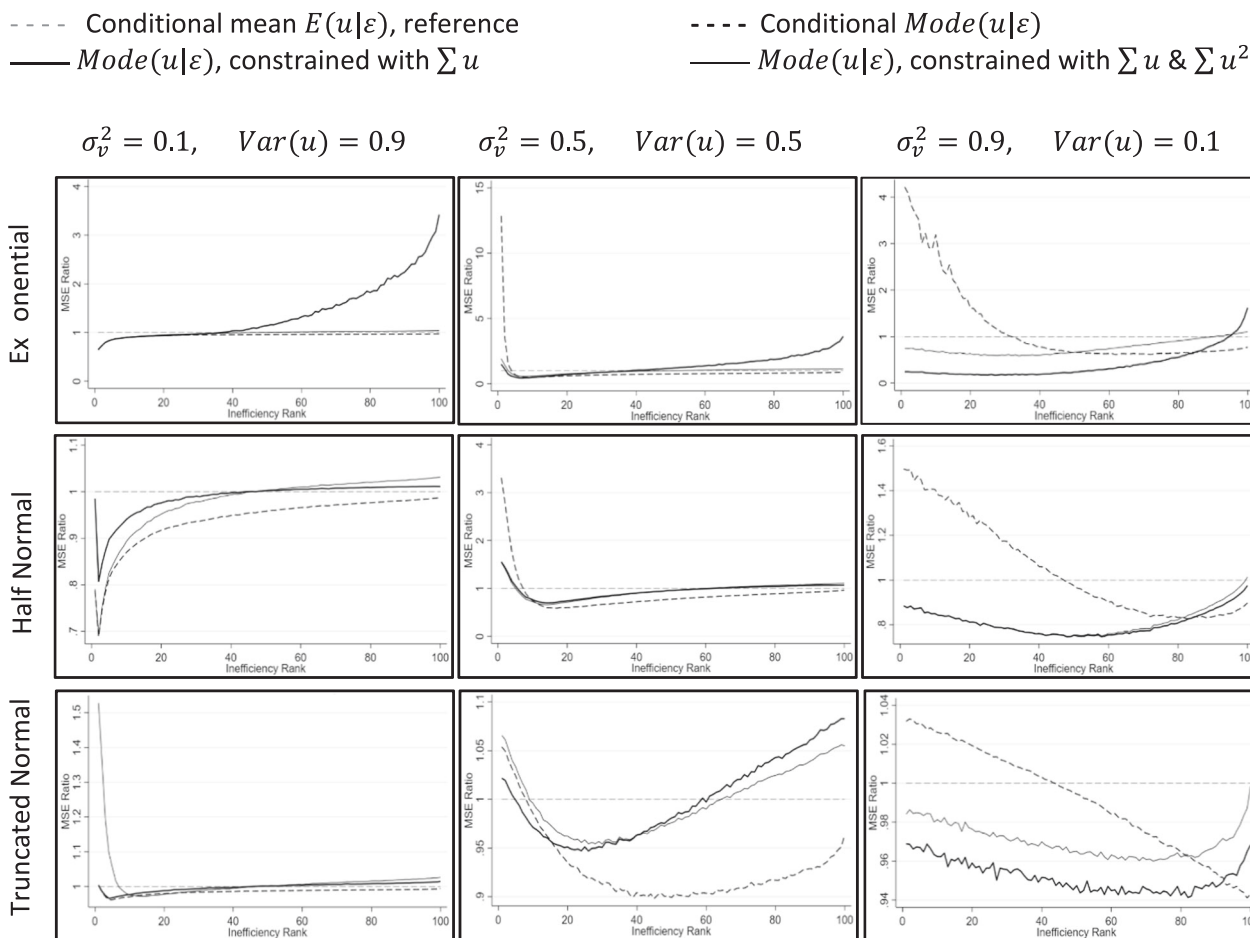
Fig. 3 Normal—Truncated Normal Model (with $\mu = 0.1$): Relative Inefficiency (MSE Ratio) compared to $E(u|\varepsilon)$. Note: The larger the noise variance, the choppier the curves of the regularized conditional mode estimators are

exponential distribution is incorrectly chosen as the inefficiency distribution.

- A summary of the above 3 points is that the analysts should make an effort to learn the characterizing conditions of the application at hand since the preferred estimation approach is determined by the signal-to-noise

ratio and the location of the inefficiencies (see Badunenko et al. 2012). The averaging of different estimators has been proposed by Sickles (2005) with simple (naïve) averaging, Huang and Lai (2012), with different functional models, Parmeter et al. (2019), Andor et al. (2019) and Tsionas (2021). A different weighted average

- - - Conditional mean $E(u|\varepsilon)$, reference   - - - - Conditional $Mode(u|\varepsilon)$
—— $Mode(u|\varepsilon)$, constrained with $\sum u$ —— $Mode(u|\varepsilon)$, constrained with $\sum u$ & $\sum u^2$

Fig. 4 Inefficiencies are drawn from basic Gompertz distributions (scale parameter $b = 1$), but the inefficiency distributions are incorrectly specified, using each of exponential, half normal and general truncated normal distributions. Relative Inefficiency (MSE Ratio) compared to $E(u|\varepsilon)$, with sample size $n = 100$. Note: The larger the noise variance, the choppier the curves of the regularized conditional mode estimators are

estimator can be developed using the regularized estimators, depending on the signal-to-noise ratio and the rank of the inefficiency. The preferred estimator can be a mixture (or weighted sum) of the conditional mode (for lower inefficiency ranks), conditional mean (for middle inefficiency ranks) and regularized conditional mode, especially the one subject to the first-moment constraint (for high inefficiency ranks). Also, a smoother estimator rather than the averaging can be an interpolated one that passes through some points (nodes and their values), like the unconditional mode, the unconditional mean and the most extreme conditional inefficiencies from both tails (skipping the lowest inefficiency if the mode inefficiency is zero).

6 Application

We consider the Swedish electricity distribution market that consisted of 154 local monopolies with complete data in

2013. The regulator wants to know the extent to which each firm can improve relative to the efficient frontier. For that purpose, we specify and estimate a variable cost (c) function where the number of customers/connection points (s) is the relevant output variable and the price of labor (l) and electricity (e) are the corresponding input prices. This production process is similar to what has been used in the past in this field; see e.g., Söderberg ([2008](#)), p. 65-66, for an extensive literature review. The price of electricity is included because firms purchase electricity to cover network losses and pay for transit on the high voltage network. The electricity price is calculated as the total costs of transit and the losses divided by the sum of the losses and high voltage deliveries.

Since the estimation of the unit inefficiency is a post-estimation procedure in SFA, entering the discussion of the selection between different productivity models, for instance between Cobb-Douglas and translog, might divert our attention away from the purpose of our proposed regularized estimators. Therefore, to save space, we only
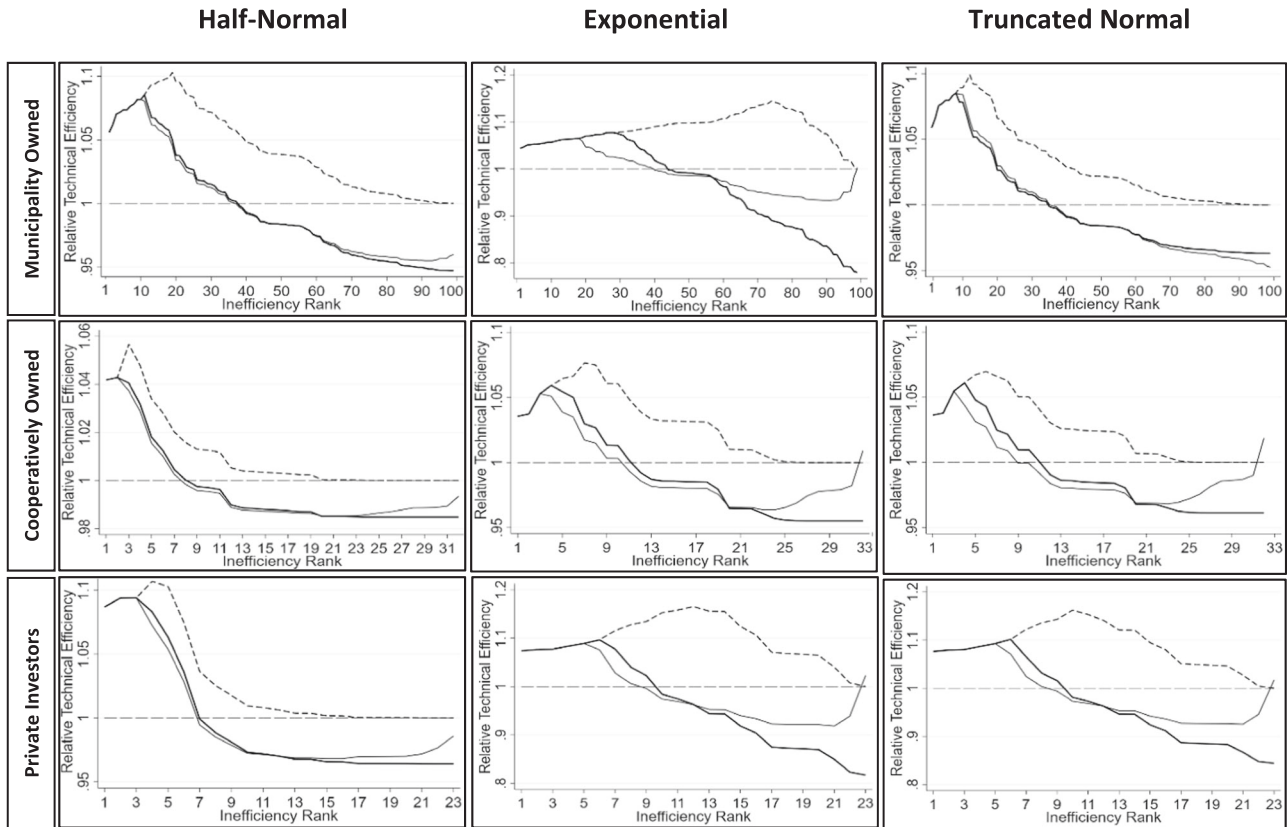
Fig. 5 Relative technical efficiency based on restricted and unrestricted conditional mode estimates, compared to technical efficiency based on unrestricted $E(u|\varepsilon)$ estimtes

assume a Cobb-Douglas production model, and specify the variable cost function as $c_i = \alpha s_i^\beta e_i^\gamma l_i^\delta$, where i denotes the firm. The homogeneity restriction can be imposed by normalizing c_i and l_i by e_i, which after natural logarithm transform allows us to write the model as:

$$ln\left(\frac{c_i}{e_i}\right) = \beta_0 + \beta_1 \ln(s_i) + \beta_2 \ln\left(\frac{l_i}{e_i}\right)$$

This expression has normal Cobb-Douglas properties, e.g., β_1 reveals the nature of the scale of production. Specifically, if $\beta_1 < 1$, then there are economies of scale; if $\beta_1 = 1$, then there is constant returns to scale; and if $\beta_1 > 1$, there are diseconomies of scale. It is straightforward to extend this Cobb-Douglas model to a stochastic frontier setting with inefficiency (u) and idiosyncratic error (v) terms (Coelli et al. 2005):

$$ln\left(\frac{c_i}{e_i}\right) = \beta_0 + \beta_1 \ln(s_i) + \beta_2 \ln\left(\frac{l_i}{e_i}\right) + \nu_i + u_i$$

Data on variable costs (Opex), the number of customers, and the price of electricity are collected from the Swedish energy regulator (the Energy Markets Inspectorate). The price of labor, which measures the average regional salary for employees in the public sector, is collected from Statistics Sweden. Data are cross-sectional from the year 2013. Because the objective function, or type of customers, can be different for different ownership forms, as shown by Meade and Söderberg (2020), we argue that the regulator has to restrict the benchmark to the firms that have the same type of owners. For Swedish electricity distribution, therefore, we need three different benchmark samples: (i) municipality owned firms ($n = 99$), (ii) cooperatively owned firms ($n = 32$), and (iii) firms owned by private investors ($n = 23$). Some descriptive statistics of the data are presented in Table 4 in the Appendix.

In Fig. 5, we see that the regularized estimators suggest that the highly inefficient firms have less technical efficiency (or equivalently larger inefficiency scores)

Table 3 Model estimates

Estimate (S.E.)		Half Normal	Exponential	Truncated Normal
Municipality Owned, $n = 99$	β_0	−8.510 (0.3309)	−8.4294 (0.329)	−8.5354 (0.4585)
	β_1	0.8087 (0.0255)	0.8075 (0.0259)	0.8090 (0.0257)
	β_2	1.1079 (0.0325)	1.1078 (0.0328)	1.1079 (0.0325)
	σ_u^2	0.0529 (0.0301)	0.013 (0.0104)	0.0459 (0.0669)
	σ_v^2	0.0222 (0.0097)	0.0287 (0.0090)	0.0204 (0.0243)
	μ	–	–	0.0827 (0.8431)
	H_0: Noinefficiency	LR test, $\overline{\chi}^2$ (p-value): 1.58 (0.104)	LR test, $\overline{\chi}^2$ (p-value): 1.24 (0.133)	Z test, z (p-value): 1.24 (0.108)
Cooperatively Owned, $n = 32$	β_0	−9.122 (0.3741)	−9.0215 (0.3026)	−9.0329 (0.3117)
	β_1	0.8854 (0.0279)	0.8912 (0.0275)	0.8903 (0.0278)
	β_2	1.1198 (0.0336)	1.1101 (0.0276)	1.1113 (0.0286)
	σ_u^2	0.0925 (0.0393)	0.0358 (0.018)	0.5632 (2.3897)
	σ_v^2	0.0064 (0.0090)	0.0087 (0.005)	0.0083 (0.0051)
	μ	–	–	−2.5088 (12.5981)
	H_0: Noinefficiency	LR test, $\overline{\chi}^2$ (p-value): 5.67 (0.009)	LR test, $\overline{\chi}^2$ (p-value): 6.47 (0.005)	Z test, z (p-value): 2.321 (0.010)
Private Investors, $n = 23$	β_0	−8.0991 (1.2217)	−7.9061 (0.6246)	−7.919 (0.6574)
	β_1	0.9424 (0.0435)	0.9441 (0.028)	0.9442 (0.0283)
	β_2	0.9589 (0.1033)	0.9513 (0.0679)	0.9514 (0.0685)
	σ_u^2	0.0201 (0.1272)	0.0367 (0.0357)	0.8233 (11.2301)
	σ_v^2	0.1549 (0.4719)	0.0388 (0.023)	0.0375 (0.029)
	μ	–	–	−3.6906 (58.5673)
	H_0: Noinefficiency	LR test, $\overline{\chi}^2$ (p-value): 1.22 (0.134)	LR test, $\overline{\chi}^2$ (p-value): 1.24 (0.132)	Z test, z (p-value): 1.163 (0.122)

compared to what the unregularized estimators estimate. Any inference regarding unit inefficiency can be poor when only a single sample is available, as it is in a cross-sectional context. However, we know that the conditional mean and the conditional mode (Theorem 3) are shrinkage estimators, i.e., they underestimate larger inefficiencies.

Therefore, our regularized estimators behave better in that sense, i.e., they estimated larger inefficiencies further from the mean/mode compared to the unregularized estimators. In addition, they have desired properties in the sense that they follow the theoretical first and second moments of inefficiency, i.e., their sample mean and variance are close to the estimated industry mean and variance.

The inference with a single sample is challenging. We checked the relative performance of each estimator by running a simulation with the same sampled data (number of customers and prices) but with the costs generated from the estimated parameters ($\widehat{\sigma}_v^2$, $\widehat{\sigma}_u^2$, $\widehat{\mu}$ and $\widehat{\beta}$) in Table 3. The simulation procedure was the same as that explained in the simulation section (Fig. 6 in the Appendix).

7 Conclusions

The conditional mean/mode estimator of unit inefficiency is a shrinkage estimator towards the inefficiency mean/mode, depending on the noise variance (or signal-to-noise ratio). It is mostly different from the firm's inefficiency itself unless there is no noise in the productivity model. The proposed regularized conditional mode estimators outperform the classical conditional mode/mean estimators, especially for highly inefficient units, and when the inefficiency distribution is incorrectly specified.

The constraints used in this paper were imposed on the first and the second moments of the inefficiencies when estimating the conditional mode of inefficiency. The idea can be further generalized to other types of constraints, distributions, or constraints on the conditional mean. In this article, the methodology is discussed in a cross-sectional context. However, it can be directly applied to a panel data context wherever the conditional mode/mean of the unit inefficiency is estimated. According to Tsionas (2017), one issue that continues to plague SFA is the endogeneity of the inputs. Our methodology is also
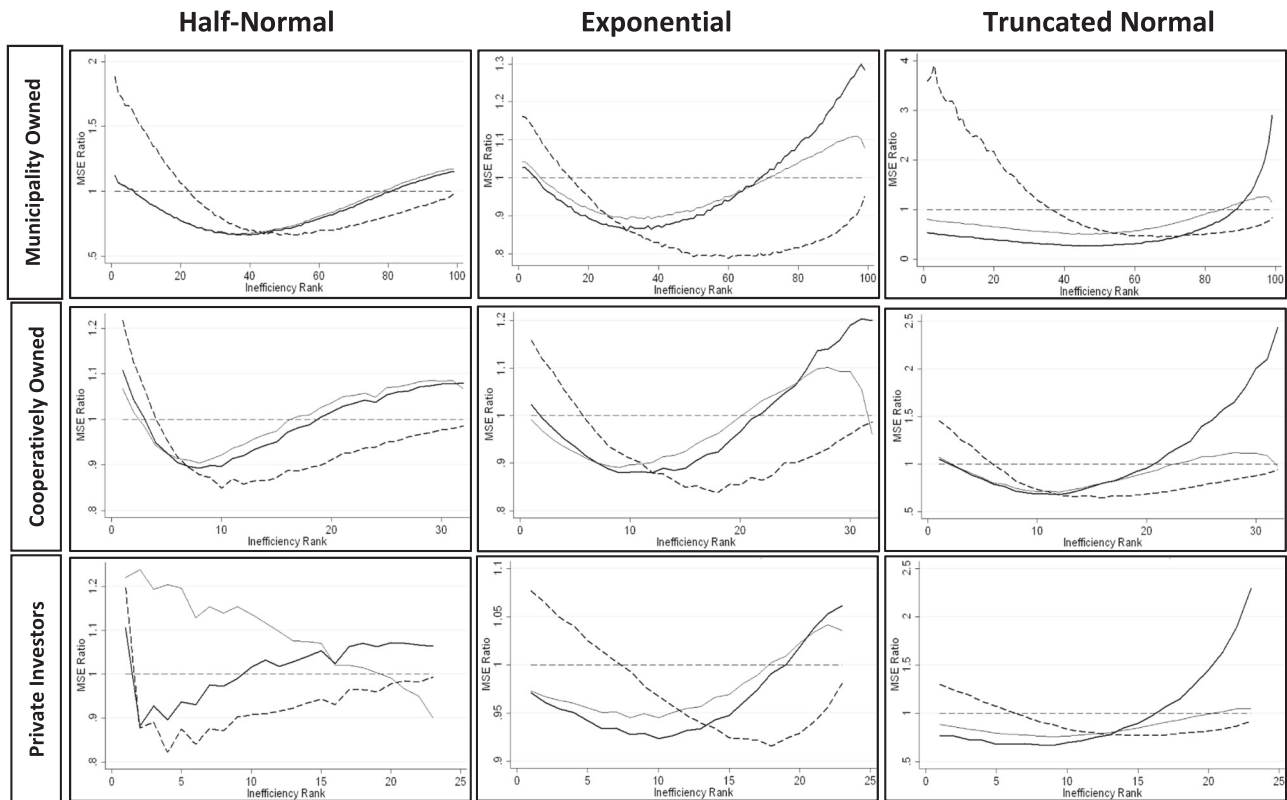
Fig. 6 The MSE ratio of unrestricted and restricted conditional mode estimators, compared to the unrestricted $E(u|\varepsilon)$ with the simulations based on the data in the Application section

Table 4 Descriptive statistics of the data used in the application section

Variable	Mean	S.D.	Min	Max
Panel A: municipality owned (n = 99)				
Variable cost (*c*)	57 593	75 404	7 207	612 999
Number of customers (*s*)	22 160	31 480	2 303	256 549
Price of electricity (*e*)	0.3673	0.3379	0.0390	1.9990
Price of labor (*l*)	21 746	356	21 250	23 160
Panel B: cooperatively owned (n = 32)				
Variable cost (*c*)	18 602	14 536	2 851	57 421
Number of customers (*s*)	4 916	4 599	808	19 120
Price of electricity (*e*)	1.3177	1.0874	0.0660	4.4560
Price of labor (*l*)	21 860	390	21 250	23 160
Panel C: owned by private investors (n = 23)				
Variable cost (*c*)	242 954	506 236	509	2 304 885
Number of customers (*s*)	89 709	193 788	158	802 484
Price of electricity (*e*)	0.5683	0.5803	0.0380	2.5280
Price of labor (*l*)	21 777	586	21 250	23 160

directly applicable to SFA methods dealing with endogeneity. And most importantly, the proposed regularized estimators are beneficial to regulators for accurately estimating high unit inefficiencies since the benchmark methods systematically underestimate the inefficiency of less efficient units.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

8 Appendix

Theorem: Suppose the function $K(w) > 0$ is defined for all $w \in \mathbb{R}$, with the properties:

1. $K(w)$ is continuously differentiable for all $w \in \mathbb{R}$, and log-concave, with its first derivative $K'(W)$.
2. $\frac{K'(w)}{K(w)}$ is bounded above, such that $\lim_{w \to +\infty} \frac{K'(w)}{K(w)} < |M| < +\infty$.
3. $\frac{K'(w)}{K(w)}$ is bounded below, such that $\lim_{w \to -\infty} \frac{K'(w)}{K(w)} > -|m| > -\infty$.

For the convergence of the integral $\int_{-\infty}^{+\infty} e^{-s\left[w - \frac{K'(w)}{K(w)}\right]} dw$, with $l < s < r$, we must have $\lim_{w \to \pm\infty} \left|\frac{\frac{K'(w)}{K(w)}}{w}\right| < 1$.

Proof:

We simplify the notation and use $\frac{K'(w)}{K(w)} = H(w)$. Then, the integral can be written as follows.

$$\int_{-\infty}^{+\infty} e^{-s[w - H(w)]} dw = \int_{0}^{+\infty} e^{-s[w - H(w)]} dw + \int_{0}^{+\infty} e^{s[w + H(-w)]} dw$$

Taking the first integral on the right side of the equation, $[w - H(w)]$ is always increasing. Then, $\lim_{w \to +\infty}[w - H(w)] = +\infty$. In its functional form, it is similar to Laplace transform of $e^{sH(w)}$. Then, it must be the case that (see Widder 1946, Theorem 2.4a)

$$\lim_{w \to +\infty} \frac{sH(w)}{w} = l \Rightarrow sl = \lim_{w \to +\infty} \frac{H(w)}{w} < \lim_{w \to +\infty} \frac{M}{w} = 0$$

which implies that $l < 0$. For all s such that $l < s$, the integral converges if the rate of change is $H(w)$ is slower than the rate of change in w, i.e., $\lim_{w \to +\infty} \frac{H(w)}{w} < 1$.

Taking the second integral on the right side of the equation, $\lim_{w \to +\infty}[w + H(-w)] = +\infty$. In its functional form, it is similar to Laplace transform of $e^{sH(w)}$ (with $-s$ from s-domain). Then, it must be the case that

$$\lim_{w \to +\infty} \frac{sH(-w)}{w} = r \Rightarrow sr = \lim_{w \to +\infty} \frac{H(-w)}{w} > \lim_{w \to +\infty} \frac{-|m|}{w} = 0$$

which implies that $r > 0$. For all s such that $s < r$, the integral converges if the rate of change in $H(-w)$ is faster than the rate of change in w, i.e.,

$1 < \lim_{w \to +\infty} \frac{H(-w)}{w} = \lim_{w \to -\infty} \frac{H(w)}{-w}$, which means $\lim_{w \to -\infty} \frac{H(w)}{w} > -1$.

Taking the conditions of both integrals together for the convergence, it means $\lim_{w \to \pm\infty} \left|\frac{H(w)}{-w}\right| < 1$. This completes the proof.

Figure 6 and Table 4

References

Aigner D, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. J Econ 6:21–37

Andor MA, Parmeter C, Sommer S (2019) Combining uncertainty with uncertainty to get certainty? Efficiency analysis for regulation purposes. Eur J Oper Res 274(1):240–252

Badunenko O, Henderson DJ, Kumbhakar CS (2012) When, where and how to perform efficiency estimation. J R Stat Soc Series A 175:863–892

Battese GE, Coelli TJ (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. J Econ 38:387–399

Behr A (2010) Quantile regression for robust bank efficiency score estimation. Eur J Oper Res 200:568–58

Belotti F, Daidone S, Ilardi G, Atelaa V (2013) Stochastic frontier analysis using Stata. Stata J 13(4):719–758

Bera A, Sharma S (1999) Estimating production uncertainty in stochastic frontier production function models. J Prod Anal 12:187–210

Bernini C, Freo M, Gardini A (2004) Quantile estimation of frontier production function. Empir Econ 29:373–381

Bogetoft P, Otto L, (2011) *Benchmarking with DEA, SFA, and R, International Series in Operations Research and Management Science*, 157. Springer https://doi.org/10.1007/978-1-4419-7961-2_10.

Cobb CW, Douglas PH (1928) A theory of production, american economic review. Supplementary, Papers and Proceedings of the Fortieth Annual Meeting of the American Economic Association 18(1):139–165

Coelli TJ, Rao DSP, O'Donnell CJ, Battese GE (2005), *An introduction to efficiency and productivity analysis*. Springer Science and Business Media.

Cox D, O'Sullivan F (1990) Asymptotic analysis of penalized likelihood and related estimators. Ann Stat 18:1676–1695

Flynn CJ, Hurvich CM, Simonoff JS (2013) Efficiency for regularization parameter selection in penalized likelihood estimation of misspecified models. J Am Stat Assoc 108(503):1031–1043. https://doi.org/10.1080/01621459.2013.801775.

Greene WH (1990) A gamma-distributed stochastic frontier model. J Econ 46(1–2):141–164

Grossman N (2005) A C^∞ Lagrange Inversion Theorem. Am Math Mon 112(6):512–514

Hall P, Presnell B (1999) Density estimation under constraints. J Comput Gr Stat 8(2):259–27

Horrace WC (2005) On ranking and selection from independent truncated normal distributions. J Econ 126:335–354

Horrace WC, Parmeter CF (2018) A Laplace stochastic frontier model. Econ Rev 37(3):260–280

Horrace WC, Wright IA (2020) Stationary points for parametric stochastic frontier models. J Bus Econ Stat 38(3):516–526

Huang CJ, Lai H (2012) Estimation of stochastic frontier models based on multimodel inference. J Prod Anal 38(3):273–284. http://www.jstor.org/stable/43549952

Jondrow J, Lovell CAK, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. J Econ 19:233–8

Jradi S, Parmeter CF, Ruggiero J (2019) Quantile estimation of the stochastic frontier model. Econ Lett. 182:15–18

Kumbhakar SC, Ghosh S, McGuckin JT (1991) A generalized production frontier approach for estimating determinants of inefficiency in U.S. dairy Farms. J Bus Econ Stat 9(3):279–286

Kumbhakar SC, Park BU, Simar L, Tsionas EG (2007) Nonparametric stochastic frontiers: A local maximum likelihood approach. J Econ 137(1):1–27

Kumbhakar S, Wang HJ, Horncastle A (2015) A practitioner guide to stochastic frontier analysis using Stata. Cambridge University Press, New York, NY

Kumbhakar SC, Lovell CAK (2000) Stochastic frontier analysis. Cambridge University Press, Cambridge

Kumbhakar SC, Parmeter CF, Zelenyuk V (2020) Stochastic Frontier Analysis: Foundations and Advances II. In: Ray S, Chambers R, Kumbhakar S (eds) Handbook of Production Economics. Springer, Singapore, p 1–40

Lenart A (2012) The moments of the Gompertz distribution and maximum likelihood estimation of its parameters. Scand Act J 2014(3):255–277

Meade R, Söderberg M (2020) Is welfare higher when utilities are owned by customers instead of investors? Evidence from electricity distribution in New Zealand. Energy Econ 86:104700

Meeusen W, van den Broeck J (1977) Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. Int Econ Rev 18:435–44

Nguyen N.B (2010), *Estimation of Technical Efficiency in Stochastic Frontier Analysis, Doctoral Dissertation*, Bowling Green State University.

Ondrich J, Ruggiero J (2001) "Efficiency measurement in the stochastic frontier model". Eur J Oper Res 129(2):434–442

Papadopoulos A (2021) Stochastic frontier models using the generalized exponential distribution. J Prod Anal 55:15–29. https://doi.org/10.1007/s11123-020-00591-9

Parmeter CF, Wang HJ, Kumbhakar SC (2017) Nonparametric estimation of the determinants of inefficiency. J Prod Anal 47:205–221

Parmeter CF, Wan ATK, Zhang X (2019) Model averaging estimators for the stochastic frontier model. J Prod Anal 51:91–103. https://doi.org/10.1007/s11123-019-00547-8

Ruggiero J (1999) Efficiency estimation and error decomposition in the stochastic frontier model: A Monte Carlo analysis. Eur J Oper Res 115(6):555–563

Sickles R (2005) Panel estimators and the identification of firm-specific efficiency levels in parametric, semiparametric and nonparametric settings. J Econ 126(2):305–334

Simar L, Van Keilegom I, Zelenyuk V (2017) Nonparametric least squares methods for stochastic frontier models. J Prod Anal 47:189–204

Stone M (2002) How not to measure the efficiency of public services (and how one might). J R Stat Soc Series A (Statistics in Society) 165:405–434

Söderberg M (2008) *Four essays on efficiency in Swedish electricity distribution*. PhD thesis University of Gothenburg.

Tancredi A (2002) *Accounting for heavy tails in stochastic frontier models, Working Paper Series 16/2002*, Department of Statistical Sciences, University of Padua. http://hdl.handle.net/11577/3442318

Tran KC, Tsionas EG (2009) Estimation of nonparametric inefficiency effects stochastic frontier models with an application to British manufacturing. Econ Model 26(5):904–909

Tsionas MG (2017) When, where, and how of efficiency estimation: improved procedures for stochastic frontier modeling. J Am Stat Assoc 112(519):948–965

Tsionas MG (2021) Optimal combinations of stochastic frontier and data envelopment analysis models. Eur J Oper Res 294(2):790–800

Wang HJ, Schmidt P (2002) One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. J Prod Anal 18:129–144

Wang WS, Schmidt P (2009) On the distribution of estimated technical efficiency in stochastic frontier models. J Econ 148:35–45

Wang Y, Wang S, Dang C, Ge W (2014) Nonparametric quantile frontier estimation under shape restriction. Eur J Oper Res 232:671–6

Wheat P, Stead AD, Greene WH (2019) Robust stochastic frontier analysis: a Student's t-half normal model with application to highway maintenance costs in England. J Prod Anal 51:21–38. https://doi.org/10.1007/s11123-018-0541-y

Whittaker ET, Watson GN (1927) A Course of Modern Analysis: An Introduction to the General Theory of Infinite Processes and of Analytic Functions with an Account of the Principal Transcendental Functions, 4th Edition. Cambridge University Press, New York

Widder DV (1946) The Laplace transform. Princeton Math Series 6:42