8-2022

The Conditional Mode in Parametric Frontier Models

William C. Horrace
Syracuse University

Hyunseok Jung
University of Arkansas, Fayetteville

Yi Yang
Amazon

CENTER FOR POLICY RESEARCH
THE MAXWELL SCHOOL

WORKING PAPER SERIES

The Conditional Mode in Parametric Frontier Models

William C. Horrace, Hyunseok Jung, and Yi Yang

Paper No. 249
August 2022

ISSN: 1525-3066

426 Eggers Hall
Syracuse University
Syracuse, NY 13244-1020
T 315.443.3114 **E** ctrpol@syr.edu
https://surface.syr.edu/cpr_workingpapers/

Syracuse University
Maxwell School of Citizenship & Public Affairs
Center for Policy Research

CENTER FOR POLICY RESEARCH - Fall 2022

Shannon Monnat, Director
Professor of Sociology, Lerner Chair in Public Health Promotion & Population Health

Associate Directors
Margaret Austin
John Yinger

SENIOR RESEARCH ASSOCIATES

Badi Baltagi, ECON
Robert Bifulco, PAIA
Carmen Carrión-Flores, ECON
Sean Drake, SOC
Alfonso Flores-Lagunes, ECON
Joss Greene, SOC
Sarah Hamersma, PAIA
Madonna Harrington Meyer, SOC
Colleen Heflin, PAIA
Yilin Hou, PAIA
Hugo Jales, ECON
Gabriela Kirk, SOC
Jeffrey Kubik, ECON
Yoonseok Lee, ECON
Leonard M. Lopoo, PAIA
Amy Lutz, SOC
Yingyi Ma, SOC
Jerry Miner, ECON
Shannon Monnat, SOC
Jan Ondrich, ECON
David Popp, PAIA
Stuart Rosenthal, ECON
Michah Rothbart, PAIA
Alexander Rothenberg, ECON
Rebecca Schewe, SOC
Ying Shi, PAIA
Saba Siddiki, PAIA
Perry Singleton, ECON
Michiko Ueda, PAIA
Yulong Wang, ECON
John Yinger, PAIA
Maria Zhu, ECON

RESEARCH AFFILIATES

William Horrace, ECON
Jennifer Karas Montez, SOC
Wei Li, MATH
Jianxuan Liu, MATH
Andrew London, SOC
Merril Silverstein, SOC
Emily Wiemers, PAIA
Peter Wilcoxen, PAIA
Janet Wilmoth, SOC
Douglas Wolf, PAIA

GRADUATE ASSOCIATES

Rhea Acuña, PAIA
Graham Ambrose, PAIA
Mayra Cervantes, PAIA
Brandon Charles, PAIA
Adam Cucchiara, PAIA
William Clay Fannin, PAIA
Myriam Gregoire-Zawilski, PAIA
Joshua Grove, SOC
Zhe He, ECON
Harneet Kaur, SOC
Hyojeong Kim, PAIA
Jooyoung Kim, ECON
Shuyuan Li, ECON
Maeve Maloney, ECON
Mitchell McFarlane, PAIA
Austin McNeill Brown, SOC. SCI.
Qasim Mehdi, PAIA
Nicholas Oesterling, PAIA
Michael Quinn, ECON
Radine Rafols, ECON
Sarah Souders, PAIA
Yue Sun, SOC
Nastassia Vaitsiakhovich, SOC
Rui Xu, ECON
Zhanhan Yu, ECON
Bo Zheng, PAIA
Dongmei Zuo, SOC. SCI.

POSTDOCTORAL SCHOLARS

Michael Dunaway, Postdoctoral Research Scholar
Xiaohan Sun, Postdoctoral Research Scholar
Xue Zhang, Postdoctoral Research Scholar

RESEARCH STAFF

Emily Graham, Assistant Director, Lerner Center
Michelle Kincaid, Senior Associate, Maxwell X Lab
Hannah Patnaik, Managing Director, Maxwell X Lab
Stephanie Potts, Research Associate, CPR
Alex Punch, Director, Lerner Center

CPR STAFF

Katrina Fiacchi, Assistant Director
Abby Frey, Administrative Assistant
Alyssa Kirk, Office Coordinator
Davor Mondom, CPDG Center Coordinator
Candi Patterson, Computer Consultant

Abstract

We survey formulations of the conditional mode estimator for technical inefficiency in parametric stochastic frontier models with normal errors and introduce new formulations for models with Laplace errors. We prove the conditional mode estimator converges pointwise to the true inefficiency value as the noise variance goes to zero. We also prove that the conditional mode estimator in the normal-exponential model achieves near-minimax optimality. Our minimax theorem implies that the worst-case risk occurs when many firms are nearly efficient, and the conditional mode estimator minimizes estimation risk in this case by estimating these small inefficiency firms as efficient. Unlike the conditional expectation estimator, the conditional mode estimator produces multiple firms with inefficiency estimates exactly equal to zero, suggesting a rule for selecting a subset of maximally efficient firms. Our simulation results show that this "zero-mode subset" has reasonably high probability of containing the most efficient firm, particularly when inefficiency is exponentially distributed. The rule is easy to apply and interpret for practitioners. We include an empirical example demonstrating the merits of the conditional mode estimator.

JEL No.: C14, C23 and D24.

Keywords: Stochastic Frontier Model, Efficiency Estimation, Laplace Distribution, Minimax Optimality, Ranking and Selection.

Authors: William C. Horrace, Department of Economics and Center for Policy Research, Syracuse University, 426 Eggers Hall, Syracuse, NY 13244, whorrace@syr.edu; Hyunseok Jung, Department of Economics, University of Arkansas, Fayetteville, AR 72701, hj020@uark.edu; Yi Yang, Amazon, yangnyi@amazon.com.

The Conditional Mode in Parametric Frontier Models

William C. Horrace *   Hyunseok Jung†   Yi Yang ‡

August 2022

Abstract

We survey formulations of the conditional mode estimator for technical inefficiency in parametric stochastic frontier models with normal errors and introduce new formulations for models with Laplace errors. We prove the conditional mode estimator converges pointwise to the true inefficiency value as the noise variance goes to zero. We also prove that the conditional mode estimator in the normal-exponential model achieves near-minimax optimality. Our minimax theorem implies that the worst-case risk occurs when many firms are nearly efficient, and the conditional mode estimator minimizes estimation risk in this case by estimating these small inefficiency firms as efficient. Unlike the conditional expectation estimator, the conditional mode estimator produces multiple firms with inefficiency estimates exactly equal to zero, suggesting a rule for selecting a subset of maximally efficient firms. Our simulation results show that this "zero-mode subset" has reasonably high probability of containing the most efficient firm, particularly when inefficiency is exponentially distributed. The rule is easy to apply and interpret for practitioners. We include an empirical example demonstrating the merits of the conditional mode estimator.

Keywords: Stochastic Frontier Model, Efficiency Estimation, Laplace Distribution, Minimax Optimality, Ranking and Selection.

*Economics Department, Syracuse University: whorrace@maxwell.syr.edu
†Economics Department, University of Arkansas: hj020@uark.edu
‡Amazon: yyang64@syr.edu

1 Introduction

This paper studies estimation of technical inefficiency in the canonical cross-sectional stochastic frontier model (SFM) introduced by Aigner et al. (1977) and Meeusen and van den Broeck (1977):

$$y_i = X_i' \quad + \varepsilon_i, \qquad \varepsilon_i = v_i - u_i \qquad \text{for } i = 1, ..., n, \tag{1}$$

where y_i is productive output and X_i is a vector of production inputs, with error components $-\infty < v_i < \infty$ (statistical noise) and $u_i \geq 0$. The nonnegative error u_i represents technical inefficiency and captures the deviation of firm's production from the efficient frontier.[1]

An important feature of the parametric SFM is that it yields estimates of firm-level technical inefficiency. Under the assumptions that u_i and v_i are *i.i.d.* over i, and u_i, v_i and X_i are mutually independent, consistent estimation of the vector follows by maximum likelihood estimation (MLE) or corrected ordinary least squares estimation (COLS, Olson et al., 1980), producing a consistent estimate of the composed error, $\hat{\varepsilon}_i$. Then, estimation of inefficiency reduces to the problem of recovering it from $\hat{\varepsilon}_i$.[2]

Jondrow et al. (1982) recommend the mean or mode of the conditional distribution $f(u_i|\varepsilon_i)$ as an estimator of u_i. We denote them as the "conditional expectation" estimator, $\hat{u}_i^e \equiv E(u_i|\varepsilon_i)$, and the "conditional mode" estimator, $\hat{u}_i^m \equiv M(u_i|\varepsilon_i)$. They derive closed-form formulae for the conditional expectation and mode under a normal-half normal assumption, i.e., $v_i \sim i.i.d.N(0, \sigma_v^2)$ and $u_i \sim i.i.d.|N(0, \sigma_u^2)|$, and a normal-exponential assumption, i.e., $v_i \sim i.i.d.N(0, \sigma_v^2)$ and $u_i \sim i.i.d.\text{Exp}(1/\sigma_u)$.[3] For notational simplicity, we will omit the subscript "i" unless necessary for exposition.

[1] Here, we present the simplest model, but there are also panel data versions of the model and versions that include a third error component that captures firm heterogeneity (e.g., Greene, 2005). While we do not consider them directly, our results hold for these models provided that they rely on parametric assumptions for identification of model parameters.

[2] To simplify discussion, we adopt the usual practice of ignoring estimation error in $\hat{\varepsilon}_i$, treating it as ε_i, which is justified for $\sqrt{n}$-consistent estimators of .

[3] $\text{Exp}(1/\sigma_u)$ denotes the exponential density $f(u) = \exp(-u/\sigma_u)/\sigma_u$.

In the SF literature, the conditional expectation estimator, $\hat{u}^e$, has become a common way to estimate firm-level inefficiency, while the conditional mode estimator $\hat{u}^m$ has received less attention. In terms of its closed-form expression, $\hat{u}^e$ has been derived under a wide range of distributional assumptions: the normal-truncated normal case by Kumbhakar and Lovell (2003); the normal-gamma case by Greene (1990); and the normal-uniform, Laplace-exponential, Cauchy-half Cauchy, etc., cases by Nguyen (2010). Its statistical properties are also well known. Waldman (1984) compares $\hat{u}^e$ with two other estimators (e.g., $-\varepsilon$) and shows that $\hat{u}^e$ is superior to the others in terms of mean squared error when the model is correctly specified. Wang and Schmidt (2009) derive the distribution of $\hat{u}^e$ and show that $\hat{u}^e$ convergences to the true inefficiency value, u, as the variance of the statistical noise approaches zero. The conditional expectation estimator, $\hat{u}^e$, has been employed in a variety of empirical settings as well: school district inefficiency (Chakraborty et al., 2001), banking efficiency (Mokhtar et al., 2006) and fishing vessel efficiency (Flores-Lagunes et al., 2007), among others. By comparison, formulations of $\hat{u}^m$ have only been extended to the normal-gamma and normal-truncated normal models (Kumbhakar and Lovell, 2003), and its application is relatively rare.

This paper studies the statistical properties of the conditional mode estimator $\hat{u}^m$ and advocates for its use for several reasons. First, building on the arguments of Wang and Schmidt (2009), we prove that $\hat{u}^m$ converges pointwise to u as the variance of the statistical noise v approaches zero, while it converges to the mode of the distribution of u as the variance of v approaches infinity.[4].

Second, we prove that $\hat{u}^m$ achieves near-minimax optimality in the commonly deployed normal-exponential model. That is, the normal-exponential specification of the conditional mode estimator minimizes the worst-case mean squared error. To show this, we reinterpret the conditional mode estimator as a LASSO-type estimator (i.e., an L_1 shrinkage estima-

[4]This result is intuitive, but it is new to the literature to the best of our knowledge.

tor). Our minimax theorem implies that the worst-case risk occurs when many firms are nearly efficient, and the conditional mode estimator minimizes estimation risk in this case by estimating these small inefficiency firms as efficient (as being "zero-mode" firms). Our simulations suggest that this feature enables the mode estimator to be more robust to parametric misspecification than other inefficiency estimators. Thus, the conditional mode estimator may be the preferred choice for estimating technical inefficiency, particularly when misspecification is a concern.

Lastly, we consider a selection rule for an "efficient subset" of firms, and show that the subset has reasonably high probability of containing the most efficient firm in the sample. Our simulation results indicate that the cardinality of the subset increases as the uncertainty in the ranking of inefficiencies across firms increases, allowing the probability of the subset containing the best firm to remain (by and large) constant as the sample size grows. This zero-mode selection rule may be useful for practitioners to identify a "credible" subset of efficient firms when multivariate selection rules (e.g., MCB, Horrace and Schmidt, 2000) are not available for practical reasons (e.g., computational burden).

The rest of the article is organized as follows: Section 2 summarizes closed-form expressions of $\hat{u}^m$ under various distributional assumptions with some being new to the literature, and studies the distribution of $\hat{u}^m$; Section 3 reinterprets $\hat{u}^m$ as a LASSO-type estimator and proves its minimax optimality; Section 4 considers the zero-mode selection rule to identify a subset of maximally efficient firms and examines its statistical properties; Section 5 analyzes the inefficiency of U.S. electric utility firms; and Section 6 concludes. All proofs are included in an Appendix.

2 Formulations and the Distribution of $\hat{u}^m$

2.1 Closed-Form Expressions of $\hat{u}^m$

Most formulations of $\hat{u}^m$ in the literature are derived under the assumption that the two-sided error v is normally distributed (e.g., Jondrow et al., 1982; Kumbhakar and Lovell, 2003; Nguyen, 2010, among others). Under this assumption, the conditional distribution $f(u|\varepsilon)$ often follows a normal distribution truncated below 0, and the conditional mode in this case is simply either the pre-truncated mean or zero, whichever is larger. We now summarize formulations of the conditional mode estimator for several parametric specifications of the model, which may be a useful resource for practitioners.

The upper panel of Table 1 summarizes closed-form expressions of $\hat{u}^m$ when v is normally distributed. First, we observe that $\hat{u}^m$ is weakly monotonic in the composed error ε, while the conditional expectation estimator $\hat{u}^e$ is strictly monotonic, which precludes ties. In other words, the efficiency rankings based on $\hat{u}^e$ do not change from the rankings of ε, while those based on $\hat{u}^m$ can be weakly different from the rankings of ε due to possible ties in inefficiency estimates. Ondrich and Ruggiero (2001) show that the rank correlation between ε and $\hat{u}^e$ in SFMs is unity when the distribution of v is strictly log-concave. The normal distribution is strictly log-concave, so their result applies.

We also observe that, as opposed to $\hat{u}^e$ (which only generates positive estimates), $\hat{u}^m$ may produce zero-inefficiency estimates, which may correspond to maximally efficient firms. This feature may be useful when analyzing highly efficient markets, and the percentage of the zero-mode firms may be a naive estimate of overall market efficiency. Selection of efficient firms in this manner and near-minimax optimality related to zero-mode estimates are discussed in the next sections.

[=== Table 1 here ===]

Some papers consider a Laplace distribution for v. Recently, Horrace and Parmeter (2018)[5] analyze SFMs under Laplace-exponential and Laplace-truncated Laplace distributions, and their simulation results show that the Laplace-exponential model often outperforms the normal-exponential model when the distribution of v is misspecified. The lower panel of Table 1 includes closed-form expressions of $\hat{u}^m$ under Laplace-uniform, Laplace-half normal, Laplace-exponential and Laplace-truncated Laplace distributions, which are new to the literature. The formulas show that $\hat{u}^m$ under a Laplace assumption is determined jointly by ε and the scale parameters of the distribution. This complication comes from the absolute value sign in the Laplace distribution function. Another distinct feature is that in certain cases, $\hat{u}^m$ is not a point estimate, but an interval. For instance, in case of Laplace-truncated Laplace model, when $\sigma_u = \sigma_v$, where σ_u and σ_v are the scale parameters of the Laplace distributions for v and u (respectively), $\hat{u}^m$ can be any point between 0 and μ, if ε is non-negative, where μ is the location parameter of the Laplace distribution for u.

2.2 The Distribution of $\hat{u}^m$

In this section, we derive the distribution of $\hat{u}^m$ under two most commonly employed specifications of the composed error: normal-half normal and normal-exponential. In particular, we investigate how the distribution changes as the variance of v, σ_v^2, approaches zero or infinity. Wang and Schmidt (2009) show that $\hat{u}^e$ converges to u when σ_v^2 approaches zero, while it converges to $E(u)$ when σ_v^2 approaches infinity. This section examines the limiting behavior of $\hat{u}^m$ under similar situations.[6]

[5] Corrections to some typos in the math formulas of the paper are included in the Appendix.

[6] As Wang and Schmidt (2009) mention, studying the distributions of $\hat{u}^e$ or $\hat{u}^m$ is not for inference about u. For such inference, the conditional distribution $f(u|\varepsilon)$ is used as in Horrace and Schmidt (1996). The goal of the analysis here is to understand the distribution of $\hat{u}^m$ that is expected in practice when $\hat{u}^m$ is used.

As Jondrow et al. (1982) show, under normal-half normal assumption, $\hat{u}^m$ is

$$\hat{u}^m = h(\varepsilon) = \begin{cases} -\varepsilon \left(\frac{\sigma_u^2}{\sigma^2}\right), & \text{if } \varepsilon \leq 0 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where $\sigma^2 = \sigma_u^2 + \sigma_v^2$. When ε is non-positive, $\hat{u}^m$ decreases with ε, which allows for a change of variable such that

$$\varepsilon = h^{-1}(\hat{u}^m) = g(\hat{u}^m), \tag{3}$$

$$f_{\hat{u}^m}(\hat{u}^m) = f_\varepsilon(g(\hat{u}^m)) \left| \frac{\partial g(\hat{u}^m)}{\partial \hat{u}^m} \right| . \tag{4}$$

When v and u are distributed as normal and half normal (NHN), respectively, Aigner et al. (1977) prove that $f_\varepsilon(\varepsilon) = \frac{2}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right)\left[1 - \Phi\left(\varepsilon\eta\sigma^{-1}\right)\right]$, where $\eta = \sigma_u/\sigma_v$. Since $g(\hat{u}^m) = -\frac{\sigma^2}{\sigma_u^2}\hat{u}^m$, the probability density function of $\hat{u}^m$ is given by

$$f_{\hat{u}^m}^{NHN}(\hat{u}^m) = F_{\hat{u}^m}^{NHN}(0)\delta(\hat{u}^m) + \frac{2\sigma}{\sigma_u^2}\phi\left(\frac{\sigma}{\sigma_u^2}\hat{u}^m\right)\Phi\left(\frac{\sigma}{\sigma_u\sigma_v}\hat{u}^m\right)\mathbf{I}\{\hat{u}^m > 0\}, \tag{5}$$

where $\delta(.)$ is a Dirac's delta function and $F_{\hat{u}^m}^{NHN}(0) = 1 - F_\varepsilon^{NHN}(0)$. We can see that $\hat{u}^m$ is a mixed random variable whose distribution is truncated below 0 and continuous over the positive support. Similarly, the density function of $\hat{u}^m$ of the normal-exponential (NE) model is

$$f_{u^m}^{NE}(\hat{u}^m) = F_{\hat{u}_m}^{NE}(0)\,\delta(\hat{u}^m) + \frac{1}{\sigma_u}\exp\left\{-\frac{\hat{u}^m}{\sigma_u} - \frac{\sigma_v^2}{2\sigma_u^2}\right\}\Phi\left(\frac{\hat{u}^m}{\sigma_v}\right)\mathbf{I}\{\hat{u}^m > 0\}, \tag{6}$$

where $F_{\hat{u}^m}^{NE}(0) = 1 - F_\varepsilon^{NE}(-\frac{\sigma_v^2}{\sigma_u})$. Then, the following theorem summarizes the limiting behaviors of $\hat{u}^m$.

Theorem 1

(1) As $\sigma_v^2 \to 0$*,* $\hat{u}^m \xrightarrow{p} u$*.*

(2) As $\sigma_v^2 \to 0$*,* $\frac{1}{\sigma_v}(\hat{u}^m - u) \xrightarrow{d} N(0,1)$*.*

(3) As $\sigma_v^2 \to \infty$*,* $\hat{u}^m \xrightarrow{p} 0 = Mode(u)$*.*

(4) As $\sigma_v^2 \to \infty$*,* $\frac{\sigma_v}{\sigma_u^2}\hat{u}^m \xrightarrow{d} z$*,* $f(z) = \frac{1}{2}\delta(z) + \phi(z)\mathbf{I}\{z > 0\}$*, when the model is NHN.*

The proof is in the Appendix. Result (1) indicates that as σ_v^2 approaches 0, $\hat{u}^m$ converges to the true inefficiency value, which is intuitive from the fact that we can observe u directly from ε when $\sigma_v^2 \to 0$. Note that this also implies that the (unconditional) distribution of $\hat{u}^m$ should be identical to the true distribution of u in the limit. Result (2) implies that, for a given u, the distribution of $\hat{u}^m$, when appropriately normalized, is approximately normal when σ_v^2 is small. On the other hand, as σ_v^2 approaches infinity, statistical noise dominates ε, so ε is no longer informative about u. Then, $\hat{u}^m$ simply uses the mode of its true distribution as an estimate, which is zero when the distribution of u is half-normal or exponential. Result (4) shows that, in the case of $\hat{u}^m$ of the normal-half normal model, when $\sigma_v^2 \to \infty$, its limiting distribution for $\hat{u}^m > 0$ is approximately a half-normal, and the rest of the density is concentrated at zero.[7]

We plot the distributions of $\hat{u}^m$ in Figure (1) and (2) for signal-to-noise ratios (i.e., σ_u/σ_v) ranging from 0.1 to 100. The value of σ_u is fixed at 1, so that the results are comparable in scale. The graphs essentially corroborate Theorem (1). One can see that the density function of $\hat{u}^m$ is clearly different from the distribution of u when $\sigma_v = 10$, but converges to the distribution of u as σ_v decreases. This is true for both normal-half normal and normal-exponential cases.

[=== Figure (1) and (2) here ===]

[7]In case of $\hat{u}^m$ of the normal-exponential model, when $\sigma_v^2 \to \infty$, the limiting distribution for $\hat{u}^m > 0$ is approximately a half-normal centered at $-\frac{\sigma_v^2}{\sigma_u}$ and the rest of density is concentrated at zero, which implies the limiting distribution becomes degenerate quickly as $\sigma_v^2 \to \infty$. Thus, we omit the result here.

3 Minimax Optimality of the Mode Estimator

This section proves that $\hat{u}^m$ attains near-minimax optimality in the normal-exponential model. To show this, we first discuss its connection to the Least Absolute Shrinkage and Selection Operator (LASSO; Tibshirani, 1996). It is well known that, in a Bayesian context, the LASSO can be viewed as the posterior mode estimator when independent Laplace priors are imposed on the coefficients in regression models (Tibshirani, 1996; Park and Casella, 2008). Since exponential (half-Laplace) inefficiency may be seen as a prior and the conditional inefficiency distribution may be seen as a posterior distribution, $\hat{u}^m$ may be regarded as a posterior mode estimator under an exponential prior. Then, the Bayesian interpretation of the LASSO suggests $\hat{u}^m$ may also be seen as a LASSO-type estimator.[8] Its closed-form expression clearly shows the connection:

$$\hat{u}_i^m(\lambda) = [-\varepsilon_i - \lambda]_+, \tag{7}$$

where $[z]_+$ returns z if $z > 0$ and 0 otherwise, and $\lambda = \frac{\sigma_v^2}{\sigma_u}$. The parameter λ serves as a tuning parameter, which shrinks inefficiency estimates toward zero as the noise σ_v grows relative to the inefficiency signal, σ_u.[9]. One notable difference between (7) and the usual LASSO-type estimator (i.e., L_1 shrinkage) is that the shrinkage effect in the LASSO is symmetric around the zero, but it is asymmetric in (7) due to the non-negativity constraint on inefficiency.

Donoho and Johnstone (1994) show that L_1 shrinkage achieves near-minimax risk when estimating nonparametric regression functions. In the following subsection, we show that the same optimality is attained by (7).

[8]Similarly, the ridge estimator may be viewed as the posterior mode estimator when independent normal priors imposed on the coefficients, and then a similar connection can be established between the ridge and $\hat{u}^m$ of normal-half normal model.

[9]The conditional expectation estimator can also be viewed as a shrinkage estimator as it shrinks inefficiency toward its mean (Wang and Schmidt, 2009).

3.1 Near-Minimax Optimality

To show near-minimax optimality in the normal-exponential model, we consider the multivariate normal estimation problem discussed in Donoho and Johnstone (1994) with some modifications: we are given n independent observations $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i=1}^n$ where $\varepsilon_i = v_i - u_i$ with $u_i \geq 0$ and $v_i \sim i.i.d.N(0, \sigma_v^2)$,[10] and the objective is to estimate $\boldsymbol{u} = (u_i)_{i=1}^n$ by some estimator, $\hat{\boldsymbol{u}} = (\hat{u}_i)_{i=1}^n$. The quality of the estimator is measured by L_2 risk, i.e., $R(\hat{\boldsymbol{u}}) = E[\sum_{i=1}^n (\hat{u}_i - u_i)^2]$. We first derive the following risk bound of the conditional mode estimator (7):

Theorem 2 *Let $\lambda = \sigma_v (2 \log n)^{1/2}$ in (7). Then,*

$$R(\hat{\boldsymbol{u}}^m(\lambda)) \leq \left(2 \log n + \frac{3}{2}\right) \left(\sum_{i=1}^n \min(u_i^2, \sigma_v^2) + \kappa_n (\log n)^{-3/2} \right) \text{ for all } u_i \geq 0, \tag{8}$$

where $\kappa_n = 5\pi^{-1/2} (\sum_{i=1}^n u_i^2 / n + \sigma_v^2)$.

The proof is in the Appendix. Theorem 2 indicates that the mean squared loss of the conditional mode estimator (7) can be no worse than a factor of $2 \log n$ of $\sum_{i=1}^n \min(u_i^2, \sigma_v^2)$ as $n \to \infty$ for "any" possible value/distribution of $u_i \geq 0$.[11]

Donoho and Johnstone (1994) show that, as $n \to \infty$,

$$\inf_{\hat{\boldsymbol{\theta}}} \sup_{\boldsymbol{\theta} \in \mathbb{R}^n} \frac{R(\hat{\boldsymbol{\theta}})}{\sigma_v^2 + \sum_{i=1}^n \min(\theta_i^2, \sigma_v^2)} \sim 2 \log n \tag{9}$$

in the usual multivariate normal estimation problem such that $\varepsilon_i = \theta_i + v_i$ with $v_i \sim i.i.d.N(0, \sigma_v^2)$ and where the objective is to estimate $\boldsymbol{\theta} = (\theta_i)_{i=1}^n$ using some estimator, $\hat{\boldsymbol{\theta}} = (\hat{\theta}_i)_{i=1}^n$.

The result (9) shows the asymptotic minimax risk expressed in terms of $\sigma_v^2 + \sum_{i=1}^n \min(\theta_i^2, \sigma_v^2)$,

[10] Note that no distributional assumption is imposed on inefficiency in this problem.

[11] More precisely, this statement requires $\sum_{i=1}^n u_i^2 / n < \infty$ as $n \to \infty$ (so that κ_n is finite). This is satisfied if a finite second moment is assumed for u_i, which is standard in the literature.

which, combined with (8), implies the conditional mode estimator (7) achieves near-minimax risk. In other words, it is an estimator that minimizes the worst-case estimation error. The proof of (9) in Donoho and Johnstone (1994) implies that the worst-case risk occurs when most true parameter values (true inefficiency values in our case) are near zero. The conditional mode estimator in this case minimizes estimation risk by estimating small inefficiencies as zero. Note that the risk bound is derived over all possible values/distributions of $u_i \geq 0$, so minimax optimality can be seen as robustness of $\hat{u}^m$ to misspecification of distribution of inefficiency. Thus, the conditional mode estimator may be the preferred choice for estimating technical inefficiency, particularly when misspecification is a concern.

Minimax optimality requires $\lambda = \sigma_v(2\log n)^{1/2}$. Since λ in (7) is calculated based on $\hat{\sigma}_v^2$ and $\hat{\sigma}_u$ estimates in practice, this means that (7) achieves optimality when $\frac{\sigma_v}{\sigma_u} \approx (2\log n)^{1/2}$ (i.e., signal-to-noise ratio goes to zero as $n \to \infty$). This is a reasonable scenario since many firms will attain full efficiency, and departures from efficient performance may become rare or negligible as markets get larger and more competitive. Nonetheless, the simulation results in the next section show that $\hat{u}^m$ exhibits mean squared errors comparable to those of $\hat{u}^e$ even when the signal-to-noise ratio is high, while achieving notable risk savings when the signal-to-noise ratio is low.

3.2 Simulation Results

To demonstrate minimax optimality of the normal-exponential conditional mode estimator, we simulate a data generating process (DGP) that is misspecified in the inefficiency distribution. Let $\varepsilon_i = v_i - u_i$, $i = 1, ..., n$, with $v_i \sim i.i.d.N(0, 1)$. We consider a mixture distribution for inefficiency where $u_i \sim i.i.d.F_{p,k} = p\delta_0 + (1-p)\chi_k^2$; δ_0 is a Dirac mass at 0; and $p \in [0, 1]$ regulates sparsity in the continuous portion of the distribution, given by χ_k^2, a chi-square distribution with k degrees of freedom. For k we randomly select integers from 1 to 10. A larger value of p increases the number of efficient firms in the sample, creating greater sparsity in the number of inefficient firms. Hence, p is a sparsity parameter. This DGP is taken from the fact that maximum risk occurs when most of inefficiency values are at or near zero, as discussed in the previous section. Obviously, the sparsity parameter affects the signal-to-noise ratio of the model as well. A large value of p decreases the signal variance. We set $n = 1,000$.

With all variables generated, we estimate σ_u^2 and σ_v^2 using the normal-half normal (NHN) and normal-exponential (NE) models of Aigner et al. (1977), from which conditional expectation and conditional mode estimates are ultimately computed. Thus, the distribution of u is misspecified in these simulations. Note that the conditional mode estimates from the NE model are (7) (hereafter the minimax estimator) and its performance is our primary interest in this simulation experiment. We repeat this procedure $1,000$ times for each case with $p \in \{0.1, 0.5, 0.9\}$.

We report two types of results in Table 2: average Root Mean Squared Error (RMSE) ($\sqrt{\sum_{i=1}^{n}(\hat{u}_i - u_i)^2/n}$), and average rank correlation between $\hat{u}_i^m$ (or $\hat{u}_i^e$) and u_i using Spearman's rank correlation coefficient.

[=== Table 2 here ===]

When there is little inefficiency sparsity, the RMSEs and rank correlations of the four different estimators are comparable. That is, when $p = 0.1$, the performance of the minimax estimator (in columns 5 and 9) is worst, but performance gaps between the estimators are small. However, as p increases to 0.5 and 0.9, the performance of the minimax estimator significantly improves, while the performances of the other estimators deteriorate in terms of both RMSE and rank correlation, leading to a large disparity between the minimax estimator and the other estimators. As discussed above, the defining feature of the minimax estimator is that it estimates small inefficiencies as exactly zero, allowing it to achieve a notable risk saving when most inefficiencies are near zero. These simulation results clearly show that the minimax estimator (7) minimizes the worst-case risk. This is a strong argument for the conditional mode estimator in the normal-exponential model, and also indicates that it may be preferred for analyzing efficiency in highly competitive markets, where inefficiency may be sparse.

4 Ranking and Selection by the Condition Mode

4.1 Selection of Efficient Firms based on Zero Conditional Mode

Inference on ranked technical efficiency estimates from SFMs has a recent but rich history. Using the conditional distribution of inefficiency, Horrace and Schmidt (1996) and Wheat et al. (2014) develop univariate prediction intervals for inference on u. Simar and Wilson (2009) consider univariate bootstrap inference. Horrace and Schmidt (2000) propose multiple comparisons for the fixed-effect version of the model, and Horrace (2005) and Flores-Lagunes et al. (2007) develop multivariate inference in the parametric SFM.

In particular, Flores-Lagunes et al. (2007) detail selection procedures for identifying a minimal cardinality subset of firms which contains the maximally efficient firm at a pre-specified confidence level. Using the conditional inefficiency density $f(u|\varepsilon)$ and its cumulative distribution function, $F(u|\varepsilon)$, the method calculates the probability of each firm j being most efficient in the sample as

$$\hat{\mathrm{P}}_j = \Pr\{u_j \le u_i \;\forall\; i \neq j|\varepsilon_1, \dots, \varepsilon_n\} = \int_0^\infty f(u_j|\varepsilon_j) \prod_{i\neq j} [1 - F(u_i|\varepsilon_i)]\, du. \tag{10}$$

Note that the magnitude of this probability is independent of whether we chose to estimate inefficiency using the conditional expectation estimator *or* the conditional mode estimator. It only uses the information in the conditional distribution of inefficiency. In particular, the ranks of the $\hat{\mathrm{P}}_j$ may not correspond to the ranks of either $\hat{u}_j^m$ or $\hat{u}_j^e$.

Then, the probability that any subset of firms contains the maximally efficient firm can be computed by summing the probabilities, $\hat{\mathrm{P}}_j$, for only those firms in the subset of interest.[12] Therefore, the technique can be used to calculate this probability sum for the zero-mode firms identified by the conditional mode estimator. The goal of this section is to perform simulations on properly specified normal-half normal and normal exponential models in (1) to examine the probabilities that the zero-mode subsets contain the maximally efficient (best) firm in the sample.

The empirical relevance of the exercise is to understand when practitioners can simply use the zero-mode selection rule (without calculating the computationally intensive probabilities in (10)) and still feel confident that this *ad hoc* selection criterion is credible (i.e., it has reasonably high probability of identifying a subset of firms that contains the "best" firm). We note that Horrace et al. (2022b) use this "zero-mode" selection rule for determining efficient schools in New York City, but their paper doesn't explore the statistical properties of the zero-mode rule under various specifications of the model.[13]

4.2 Simulation Results

We consider the following panel DGP with time invariant u: $\varepsilon_{it} = v_{it} - u_i$, $i = 1, ..., n$, $t = 1, ..., T$, where v_{it} and u_i are generated from either $v_{it} \sim i.i.d.N(0, \sigma_v^2)$, and $u_i \sim i.i.d.|N(0, \sigma_u^2)|$, or $v_{it} \sim i.i.d.N(0, \sigma_v^2)$, and $u_i \sim i.i.d.\mathrm{Exp}(1/\sigma_u)$. We always maintain $\sigma_v^2 + \sigma_u^2 = 1$, and η denotes signal-to-noise ratio (σ_u/σ_v). The total number of iterations for each case is $M = 1,000$. Let $N = \{1, 2, ..., n\}$ be the set of all firm indices in the sample,

[12]Note that $\sum_{j=1}^{n} \hat{\mathrm{P}}_j = 1$. Flores-Lagunes et al. (2007) are actually concerned with pre-specifying the probability sum at a level like 0.95, and then populating the subset with firms with the largest values of $\hat{\mathrm{P}}_j$ to achieve the pre-specified probability, producing a minimal cardinality subset at the pre-specified probability.

[13]For SFMs that explicitly allow for multiple efficient firms, see Kumbhakar et al. (2013) and Horrace et al. (2022a).

and

$$N^m = \{i : \hat{u}_i^m = 0\} \subseteq N \tag{11}$$

be the subset of zero-mode firms. Let the ranked true inefficiencies be $u_{[1]} \leq u_{[2]} \leq ... \leq u_{[N]}$. Then, the probability that the zero-mode subset contains the best firm is

$$\hat{\mathrm{P}}^m = \Pr\{[1] \in N^m | \varepsilon_1, \ldots, \varepsilon_n\} = \sum_{j \in N^m} \hat{\mathrm{P}}_j \, . \tag{12}$$

Then, the empirical question is, how large is $\hat{\mathrm{P}}^m$ on average for our various simulation designs? Let the ranked inefficiency estimates based on the conditional expectation be $\hat{u}^e_{(1)} \leq \hat{u}^e_{(2)} \leq ... \leq \hat{u}^e_{(N)}$. We also calculate the probability that the firm with the smallest conditional expectation estimate is the true best firm

$$\hat{\mathrm{P}}^{\mathrm{e}} = \Pr\{[1] = (1) | \varepsilon_1, \ldots, \varepsilon_n\}. \tag{13}$$

Generally speaking, $\hat{\mathrm{P}}^{\mathrm{e}}$ will be a decreasing function of the sample size, so we are also interested in understanding how $\hat{\mathrm{P}}^{\mathrm{m}}$ responds to the sample size.

We report the average $\hat{\mathrm{P}}^e$, the average $\hat{\mathrm{P}}^m$ and the average cardinality of N^m, denoted $|N^m|$, over the $M = 1,000$ iterations. To assess accuracy of average estimated probabilities $\hat{\mathrm{P}}^e$ and $\hat{\mathrm{P}}^m$, we also calculate their empirical analogs,

$$\begin{aligned} \tilde{\mathrm{P}}^e &\equiv \frac{1}{M} \sum_{k=1}^{M} \mathbb{1}_k \left\{[1] = (1)\right\}, \\ \tilde{\mathrm{P}}^m &\equiv \frac{1}{M} \sum_{k=1}^{M} \mathbb{1}_k \left\{[1] \in N^m\right\}, \end{aligned}$$

where $\mathbb{1}_k$ is an indicator function for the k^{th} simulation sample. We expect average $\hat{\mathrm{P}}^e$ and average $\hat{\mathrm{P}}^m$ to be similar to $\tilde{\mathrm{P}}^e$ and $\tilde{\mathrm{P}}^m$, respectively.

The results are reported in Table 3. The first panel reports the results when the number of firms n increases from 5 to 100, for fixed $T = 1$ and $\eta = 1$. We can see that, as n increases, average $\hat{\mathrm{P}}^e$ significantly decreases, while average $\hat{\mathrm{P}}^m$ remains relatively constant in the range of $0.4 \sim 0.5$ under the normal-half normal distribution and $0.80 \sim 0.85$ under the normal-exponential distribution. That is, the zero-mode subset has reasonably high probability of containing the best firm ($0.80 \sim 0.85$) when the normal-exponential model is the true model. Note also that average cardinality, $|N^m|$, increases proportionally to n.

[=== Table 3 here ===]

Second panel shows the results when η^2 changes from 0.1 to 10 while fixing $n = 50$ and $T = 1$. As η increases, i.e., signal increases, average $\hat{\mathrm{P}}^e$ increases, but average $\hat{\mathrm{P}}^m$ remains largely constant under normal-half normal distribution and decreases under normal-exponential distribution. Even with the decrease, we still observe a significant gap between average $\hat{\mathrm{P}}^e$ and $\hat{\mathrm{P}}^m$. The results in the last panel, where the number of time periods T increases from 1 to 100 while fixing $n = 50$ and $\eta = 1$, are overall similar to those in the second panel. This is because, under the time-invariant inefficiency setting, an increase in T leads to an increase in signal-to-noise ratio. In all cases, average $\hat{\mathrm{P}}^e$ and average $\hat{\mathrm{P}}^m$ are reasonably close to $\tilde{\mathrm{P}}^e$ and $\tilde{\mathrm{P}}^m$.

5 Empirical Example

We analyze the technical efficiency of 123 U.S. electric utility firms using $\hat{u}^m$ and the zero-mode selection rule considered in Section 4. The dataset used in this section was previously analyzed by Greene (1990) and Nguyen (2010). We consider the following production function specification:

$$\ln Q_i = \ _0 + \ _1 \ln L_i + \ _2 \ln K_i + \ _3 \ln F_i + v_i - u_i. \tag{14}$$

where Q is output, L is labor, K is capital and F is fuel. We estimated the parameters using the normal-exponential model: $(\hat{}_0, \hat{}_1, \hat{}_2, \hat{}_3) = (8.497, -0.127, 0.089, 1.103)$, $(\hat{\sigma}_v, \ \hat{\sigma}_u) = (0.088, 0.129)$. Hence, the estimated signal-to-noise ratio $\hat{\sigma}_u/\hat{\sigma}_v$ in this data is 1.4659.

Tables 4, 5 and 6 include conditional expectation (column 3) and conditional mode (column 4) estimates, and the probability of each firm being the most efficient firm calculated by (10) (column 5), where the firms are ranked by the conditional expectation estimates. Column 6 contains the cumulative version of the probabilities in column 4, which represents the probability that the subset of a given firm and all the higher ranked firms contains the most efficient firm in the sample. For instance, the value in the row 3 and column 6 of table 4 indicates the probability that the most efficient firm is among S'westernP.S, NortheastUtil, Orange&Rockln is 0.104.

There are forty firms with a zero conditional mode estimate. This subset of firms has 59.4% probability of containing the most efficient firm, implying that if we naively selected these firms as candidates for the most efficient firm simply by inspecting their conditional mode estimates, we still have a nearly 60% chance to have included the best firm in the subset without calculating the probabilities in column 5. This reassures us that selection based on the zero conditional mode estimate may be a convenient and reliable way to identify a subset with high probability of containing the most efficient firm. The distributions of the conditional expectation and conditional mode estimates are plotted in Figure 3, where we can see a mass of zero estimates in the graphs of the conditional mode estimates.

6 Conclusions

One of the main purposes of stochastic frontier model is to measure firm-specific technical inefficiency, for which Jondrow et al. (1982) propose two estimators - the conditional expectation and conditional mode estimators. While there are numerous papers studying and applying the conditional expectation estimator, the conditional mode estimator has been largely overlooked in the literature.

This paper attempts to fill the gap of research on the conditional mode. We theoretically and empirically demonstrated the merits of the conditional mode estimator, two of which may be particularly important in practice. First, zero conditional mode estimates can be used as a simple selection rule that chooses efficient firms with a reasonably high probability of containing the most efficient firm in the sample, particularly when the model is normal-exponential. Second, it exhibits near-minimax optimality when normal-exponential model is applied. Our theoretical and empirical results indicate that the conditional mode estimator is particularly suited for analyzing competitive markets, where most firms may be near the frontier and identification of multiple efficient firms is desirable.

REFERENCES

Aigner, D., Lovell, C., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1):21 – 37.

Chakraborty, K., Biswas, B., and Lewis, W. C. (2001). Measurement of Technical Efficiency in Public Education: A Stochastic and Nonstochastic Production Function Approach. *Southern Economic Journal*.

Donoho, D. L. and Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *biometrika*, 81(3):425–455.

Flores-Lagunes, A., Horrace, W. C., and Schnier, K. E. (2007). Identifying technically efficient fishing vessels: a non-empty, minimal subset approach. *Journal of Applied Econometrics*, 22(4):729–745.

Greene, W. H. (1990). A gamma-distributed stochastic frontier model. *Journal of Econometrics*, 46(1):141 – 163.

Greene, W. H. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics*, 126(2):269–303.

Horrace, W. C. (2005). On ranking and selection from independent truncated normal distributions. *Journal of Econometrics*, 126(2):335 – 354. Current developments in productivity and efficiency measurement.

Horrace, W. C., Jung, H., and Lee, Y. (2022a). LASSO for stochastic frontier models with many efficient firms. *Working Paper, https://surface.syr.edu/cpr/416*.

Horrace, W. C. and Parmeter, C. F. (2018). A laplace stochastic frontier model. *Econometric Reviews*, 37(3):260–280.

Horrace, W. C., Rothbart, M. W., and Yang, Y. (2022b). Technical efficiency of public middle schools in new york city. *Economics of Education Review*, 86:102216.

Horrace, W. C. and Schmidt, P. (1996). Confidence statements for efficiency estimates from stochastic frontier models. *Journal of Productivity Analysis*, 7(2-3):257–282.

Horrace, W. C. and Schmidt, P. (2000). Multiple comparisons with the best, with economic applications. *Journal of Applied Econometrics*, 15(1):1–26.

Jondrow, J., Lovell, C. K., Materov, I. S., and Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19(2):233 – 238.

Kumbhakar, S. C. and Lovell, C. K. (2003). *Stochastic frontier analysis*. Cambridge university press.

Kumbhakar, S. C., Parmeter, C. F., and Tsionas, E. G. (2013). A zero inefficiency stochastic frontier model. *Journal of Econometrics*, 172(1):66 – 76.

Meeusen, W. and van den Broeck, J. (1977). Efficiency estimation from cobb-douglas production functions with composed error. *International Economic Review*, 18(2):435–44.

Mokhtar, H. S. A., Abdullah, N., and Al-Habshi, S. M. (2006). Efficiency of islamic banking in malaysia: A stochastic frontier approach. *Journal of Economic Cooperation*, 27(2):37–70.

Nguyen, N. B. (2010). *Estimation of technical efficiency in stochastic frontier analysis*. PhD thesis, Bowling Green State University.

Olson, J. A., Schmidt, P., and Waldman, D. M. (1980). A monte carlo study of estimators of stochastic frontier production functions. *Journal of Econometrics*, 13(1):67–82.

Ondrich, J. and Ruggiero, J. (2001). Efficiency measurement in the stochastic frontier model. *European Journal of Operational Research*, 129(2):434–442.

Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686.

Simar, L. and Wilson, P. W. (2009). Inferences from cross-sectional, stochastic frontier models. *Econometric Reviews*, 29(1):62–98.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.

Waldman, D. M. (1984). Properties of technical efficiency estimators in the stochastic frontier model. *Journal of econometrics*, 25(3):353–364.

Wang, W. S. and Schmidt, P. (2009). On the distribution of estimated technical efficiency in stochastic frontier models. *Journal of Econometrics*, 148(1):36–45.

Wheat, P., Greene, W., and Smith, A. (2014). Understanding prediction intervals for firm specific inefficiency scores from parametric stochastic frontier models. *Journal of Productivity Analysis*, 42(1):55–65.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.

Table 1: Closed-Form Formulas for the Conditional Mode

When $v \sim N(0, \sigma_v^2)$		
$f(u)$	u^m	
Uniform on $[0, A]$ with $f(u) = \frac{1}{A}$	0	$\varepsilon \geq 0$
	$-\varepsilon$	$-A \leq \varepsilon < 0$
	A	$\varepsilon < -A$
Doubly truncated normal on $[0, A]$	0	$\frac{-\sigma_u^2\varepsilon + \mu\sigma_v^2}{\sigma_v^2 + \sigma_u^2} < 0$
with $f(u) = \frac{\phi\left(\frac{u-\mu}{\sigma_u}\right)}{\sigma_u\left(\Phi\left(\frac{A-\mu}{\sigma_u}\right) - \Phi\left(\frac{-\mu}{\sigma_u}\right)\right)}$	$\frac{-\sigma_u^2\varepsilon + \mu\sigma_v^2}{\sigma_v^2 + \sigma_u^2}$	$0 \leq \frac{-\sigma_u^2\varepsilon + \mu\sigma_v^2}{\sigma_v^2 + \sigma_u^2} < A$
	A	$\frac{-\sigma_u^2\varepsilon + \mu\sigma_v^2}{\sigma_v^2 + \sigma_u^2} \geq A$
Exponential with $f(u) = \exp\left(-\frac{u}{\sigma_u}\right)/\sigma_u$	0	$\varepsilon > -\sigma_v^2/\sigma_u$
	$-\varepsilon - \sigma_v^2/\sigma_u$	$\varepsilon \leq -\sigma_v^2/\sigma_u$
When $v \sim$ *Laplace with* $f(v) = \exp\left(-\frac{\lvert v\rvert}{\sigma_v}\right)/2\sigma_v$		
$f(u)$	u^m	
Uniform on $[0, A]$ with $f(u) = \frac{1}{A}$	0	$\varepsilon > 0$
	$-\varepsilon$	$-A < \varepsilon \leq 0$
	A	$\varepsilon \leq -A$
Half normal with $f(u) = 2\phi\left(\frac{u}{\sigma_u}\right)/\sigma_u$	0	$\varepsilon \geq 0$
	$-\varepsilon$	$-\sigma_u^2/\sigma_v \leq \varepsilon < 0$
	σ_u^2/σ_v	$\varepsilon < -\sigma_u^2/\sigma_v$
Exponential with $f(u) = \exp\left(-\frac{u}{\sigma_u}\right)/\sigma_u$	0	$\varepsilon \geq 0$
	0	$\varepsilon < 0$ and $\sigma_u < \sigma_v$
	$[0, -\varepsilon]$	$\varepsilon < 0$ and $\sigma_u = \sigma_v$
	$-\varepsilon$	$\varepsilon < 0$ and $\sigma_u > \sigma_v$
Truncated Laplace	0	$\varepsilon \geq 0$ and $\sigma_u > \sigma_v$
with $f(u) = \frac{\exp(-\lvert u-\mu\rvert/\sigma_u)}{2\sigma_u(1-0.5\exp(-\mu/\sigma_u))}$, $\mu > 0$	$[0, \mu]$	$\varepsilon \geq 0$ and $\sigma_u = \sigma_v$
	μ	$\sigma_u < \sigma_v$
	$[\mu, -\varepsilon]$ or $[-\varepsilon, \mu]$	$\varepsilon < 0$ and $\sigma_u = \sigma_v$
	$-\varepsilon$	$\varepsilon < 0$ and $\sigma_u > \sigma_v$

Truncated Laplace refers to Laplace distribution truncated below at 0 with a positive pre-truncated mean. Truncated Laplace with a non-positive pre-truncated mean is an exponential distribution.

Table 2: Minimax Optimality of the Mode Estimator

	RMSE				Rank Correlation			
Sparsity	[1]	[2]	[3]	[4]	[1]	[2]	[3]	[4]
$p = 0.1$	0.973	0.947	0.980	1.003	0.90	0.90	0.90	0.90
$p = 0.5$	1.526	1.172	1.078	0.894	0.83	0.84	0.83	0.86
$p = 0.9$	1.738	1.646	0.774	0.555	0.45	0.47	0.45	0.69

[1] $\hat{u}^e$ of normal-half normal model (NHN); [2] $\hat{u}^m$ of NHN; [3] $\hat{u}^e$ of normal-exponential model (NE); [4] $\hat{u}^m$ of NE.

Table 3: Selection of Efficient Firms by the Conditional Mode Estimator

	Normal-Half Normal					Normal-Exponential				
	n									
	5	10	20	50	100	5	10	20	50	100
$\hat{\mathrm{P}}^m$	0.405	0.475	0.488	0.496	0.497	0.882	0.955	0.964	0.966	0.968
$\tilde{\mathrm{P}}^m$	0.393	0.488	0.483	0.462	0.504	0.819	0.820	0.845	0.839	0.845
$\lvert N^m \rvert$	1.221	6.168	12.378	25.063	50.018	2.720	13.350	26.992	53.715	107.509
$\hat{\mathrm{P}}^e$	0.363	0.107	0.061	0.033	0.018	0.520	0.208	0.127	0.076	0.043
$\tilde{\mathrm{P}}^e$	0.362	0.116	0.061	0.026	0.020	0.403	0.106	0.045	0.039	0.022
	η^2									
	0.1	0.5	1	5	10	0.1	0.5	1	5	10
$\hat{\mathrm{P}}^m$	0.493	0.491	0.488	0.476	0.471	1.000	0.996	0.964	0.769	0.685
$\tilde{\mathrm{P}}^m$	0.491	0.485	0.483	0.472	0.463	1.000	0.923	0.845	0.644	0.591
$\lvert N^m \rvert$	19.955	15.080	12.378	6.590	4.844	49.788	37.019	26.992	11.122	7.571
$\hat{\mathrm{P}}^e$	0.032	0.048	0.061	0.109	0.141	0.232	0.137	0.127	0.142	0.161
$\tilde{\mathrm{P}}^e$	0.032	0.046	0.061	0.098	0.128	0.023	0.034	0.045	0.082	0.107
	T									
	1	5	10	50	100	1	5	10	50	100
$\hat{\mathrm{P}}^m$	0.488	0.476	0.467	0.436	0.400	0.964	0.761	0.678	0.536	0.476
$\tilde{\mathrm{P}}^m$	0.483	0.476	0.496	0.438	0.411	0.845	0.664	0.588	0.492	0.469
$\lvert N^m \rvert$	12.378	6.625	4.887	2.287	1.530	26.992	10.979	7.406	3.054	2.096
$\hat{\mathrm{P}}^e$	0.061	0.108	0.136	0.247	0.317	0.127	0.142	0.163	0.249	0.300
$\tilde{\mathrm{P}}^e$	0.061	0.101	0.149	0.246	0.329	0.045	0.099	0.147	0.231	0.284

Unless otherwise specified, we fix $n = 50$, $T = 1$ and $\eta^2 = 1$. $\hat{\mathrm{P}}^m$ and $\tilde{\mathrm{P}}^m$ are the average calculated and empirical probabilities that the zero-mode subset contains the best firm. $\hat{\mathrm{P}}^e$ and $\tilde{\mathrm{P}}^e$ are the average calculated and empirical probabilities that the firm with the smallest conditional expectation estimate is the best firm. $|N^m|$ is the average cardinality of the zero-mode subset.

Table 4: Inefficiency of U.S. Electric Utility Firms

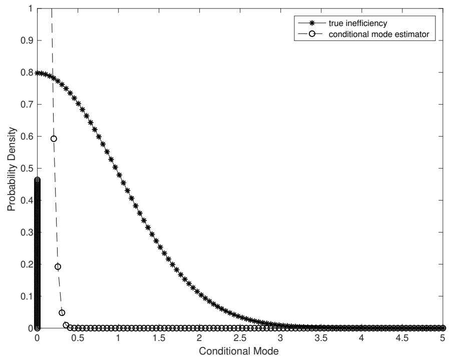
Rank	Name	$\hat{u}^e$	$\hat{u}^m$	Probability	Cumulative Probability
1	S'westernP.S.	0.017	0.000	0.048	0.048
2	NortheastUtil.	0.023	0.000	0.035	0.083
3	Orange&Rockln.	0.036	0.000	0.021	0.104
4	DaytonPwr.&Lt.	0.037	0.000	0.021	0.125
5	BostonEdison	0.038	0.000	0.020	0.144
6	NewMex.Elec.Ser.	0.041	0.000	0.018	0.162
7	MontanaPower	0.042	0.000	0.018	0.180
8	WestTunasUtil.	0.043	0.000	0.017	0.198
9	SierraPac.Pwr.	0.043	0.000	0.017	0.215
10	ToledoEdison	0.043	0.000	0.017	0.232
11	Ctrl.HudsonG.&E.	0.044	0.000	0.017	0.249
12	PacicP&L	0.045	0.000	0.017	0.266
13	HawaiianElec.	0.045	0.000	0.017	0.282
14	LouisvilleG.&E.	0.048	0.000	0.015	0.298
15	Ctrl.III.Pub.Ser.	0.049	0.000	0.015	0.312
16	BaiigorHydro.	0.050	0.000	0.014	0.327
17	Wisc.Pub.Ser.	0.052	0.000	0.014	0.340
18	Wisc.Pwr.&Light	0.056	0.000	0.013	0.353
19	NevadaPower	0.057	0.000	0.012	0.366
20	Indy.Power&L.	0.057	0.000	0.012	0.378
21	NewEnglandEl.	0.058	0.000	0.012	0.390
22	Ctrl.Tel.&Util.	0.058	0.000	0.012	0.402
23	So.Car.EI.&Gas	0.059	0.000	0.012	0.414
24	ElPasoElec.	0.060	0.000	0.012	0.426
25	Atl.CityElec.	0.061	0.000	0.011	0.437
26	KentuckyUtils.	0.061	0.000	0.011	0.448
27	UtahPower&Lt.	0.061	0.000	0.011	0.460
28	DelmarvaP.&L.	0.062	0.000	0.011	0.471
29	MauiElectric	0.062	0.000	0.011	0.482
30	P.S.Co.ofN.H.	0.063	0.000	0.011	0.493
31	Ark.Mo.Power	0.063	0.000	0.011	0.504
32	FloridaPower	0.065	0.000	0.010	0.514
33	CommunityP.S.	0.066	0.000	0.010	0.525
34	CentralLa.Pwr.	0.066	0.000	0.010	0.535
35	ClevelandEl.I.	0.066	0.000	0.010	0.545
36	CentralKansas	0.067	0.000	0.010	0.555
37	TucsonGas&E.	0.069	0.000	0.010	0.565
38	SavannahE.&P.	0.070	0.000	0.010	0.575
39	NiagaraMohawk	0.071	0.000	0.009	0.584
40	DukePowerCo.	0.071	0.000	0.009	0.594
41	KansasGas&El.	0.072	0.003	0.009	0.603

Table 5: Inefficiency of U.S. Electric Utility Firms

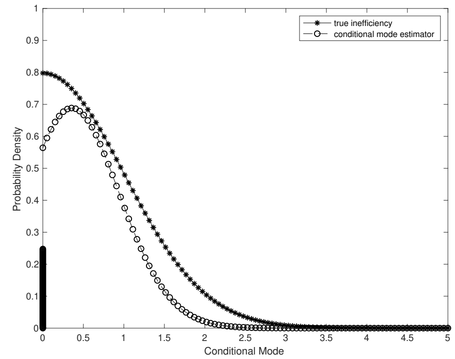
Rank	Name	$\hat{u}^e$	$\hat{u}^m$	Probability	Cumulative Probability
42	IowaPub.Ser.	0.072	0.005	0.009	0.612
43	SanDiegoG.&E.	0.073	0.005	0.009	0.621
44	Balt.Gas&El.	0.073	0.006	0.009	0.630
45	IowaSouthern	0.075	0.012	0.009	0.639
46	CarolinaP.&L.	0.077	0.016	0.008	0.647
47	BlackHillsP&L	0.077	0.016	0.008	0.656
48	Cinci.Gas&El.	0.078	0.018	0.008	0.664
49	LongIs.Light	0.078	0.019	0.008	0.673
50	Ariz.Pub.Ser.	0.078	0.020	0.008	0.681
51	EmpireDist.El.	0.078	0.020	0.008	0.689
52	Cent.MainePwr.	0.079	0.020	0.008	0.697
53	So.Ind.G.&E.	0.079	0.022	0.008	0.705
54	Ark.Power&Lt.	0.080	0.025	0.008	0.713
55	No.Ind.Pub.Ser.	0.081	0.026	0.008	0.721
56	MinnesotaP.&L.	0.081	0.027	0.008	0.729
57	InterstatePwr.	0.083	0.030	0.008	0.737
58	Pub.Ser.Colo.	0.083	0.030	0.008	0.745
59	NewOrleansP.S.	0.084	0.033	0.007	0.752
60	OhioEdisonCo.	0.085	0.035	0.007	0.759
61	Mo.PublicSer.	0.087	0.040	0.007	0.767
62	MadisonGas&E.	0.088	0.041	0.007	0.774
63	S'westernEl.Pr.	0.089	0.043	0.007	0.781
64	CentralPwr.&L.	0.090	0.046	0.007	0.787
65	LouisianaP.&L.	0.093	0.051	0.007	0.794
66	TampaElectric	0.093	0.052	0.006	0.800
67	IllinoisPower	0.094	0.054	0.006	0.807
68	Pub.Scr.NewMex.	0.096	0.057	0.006	0.813
69	Penn.Pwr.&Lt.	0.096	0.058	0.006	0.819
70	NYStateEl.&Gas	0.099	0.064	0.006	0.825
71	Vir.Elec&Pwr.	0.100	0.064	0.006	0.831
72	Nrth.Sts.Pwr.	0.100	0.064	0.006	0.837
73	TexasPower&L.	0.101	0.066	0.006	0.843
74	East.Utl.Ass.	0.101	0.067	0.006	0.848
75	KansasPwr.&L.	0.104	0.072	0.006	0.854
76	Okla.Gas&Elec.	0.105	0.074	0.005	0.859
77	Miss.Power&L.	0.107	0.077	0.005	0.865
78	ConsumersPwr.	0.107	0.078	0.005	0.870
79	IowaPwr.&Light	0.108	0.079	0.005	0.875
80	GeneralPub.U.	0.110	0.082	0.005	0.880
81	PotomacEl.Pr.	0.111	0.084	0.005	0.885
82	SouthernCo.	0.112	0.086	0.005	0.890

Table 6: Inefficiency of U.S. Electric Utility Firms

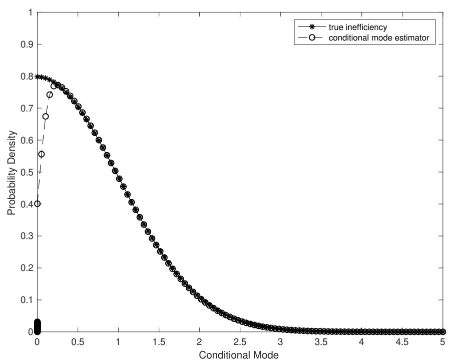
Rank	Name	$\hat{u}^e$	$\hat{u}^m$	Probability	Cumulative Probability
83	DuquesneLight	0.113	0.087	0.005	0.895
84	Pub.Ser.El.&G.	0.113	0.087	0.005	0.900
85	DallasPwr.&L.	0.115	0.090	0.005	0.904
86	UnionElec.Co.	0.115	0.090	0.005	0.909
87	Pac.Gas&Elec.	0.117	0.094	0.005	0.914
88	HoustonLt.&Pr.	0.117	0.094	0.005	0.918
89	MontDak.Utils.	0.118	0.094	0.005	0.923
90	RochesterG.&E.	0.119	0.096	0.004	0.927
91	Pub.Ser.Okla.	0.119	0.096	0.004	0.932
92	Kan.CityP.&L.	0.119	0.097	0.004	0.936
93	UpperPen.Pwr.	0.121	0.099	0.004	0.940
94	OtterTailPwr.	0.121	0.100	0.004	0.945
95	AlleghenyPr.	0.122	0.101	0.004	0.949
96	Amer.Elec.Pr.	0.131	0.113	0.004	0.953
97	Pub.Ser.OfInd.	0.131	0.114	0.004	0.957
98	Phila.Elect.	0.132	0.115	0.004	0.960
99	LakeSup.Dist.Pr.	0.136	0.121	0.004	0.964
100	UnitedIII.Co.	0.139	0.124	0.003	0.967
101	Colms&So.Ohio	0.140	0.127	0.003	0.971
102	GulfStatesUtl.	0.141	0.128	0.003	0.974
103	NewEng.G.&E.Ass.	0.145	0.132	0.003	0.977
104	TexasElec.Ser.	0.148	0.137	0.003	0.980
105	Common.Edison	0.152	0.141	0.003	0.983
106	FloridaPwr.&L.	0.154	0.144	0.003	0.986
107	Wisc.Elec.Pwr.	0.162	0.154	0.002	0.988
108	St.JosephL&P	0.168	0.161	0.002	0.990
109	So.Cal.Edison	0.168	0.161	0.002	0.993
110	IowaElec.L.&Pwr.	0.170	0.163	0.002	0.995
111	DetroitEdison	0.170	0.163	0.002	0.997
112	HiloElec.Light	0.205	0.202	0.001	0.999
113	Consol.Edison	0.234	0.233	0.001	0.999
114	NewportElec.	0.330	0.330	0.000	1.000
115	MainePub.Ser.	0.355	0.355	0.000	1.000
116	CitizensUtils.	0.362	0.362	0.000	1.000
117	FitchburgG.&E.	0.373	0.373	0.000	1.000
118	UnitedGas.I.	0.414	0.414	0.000	1.000
119	Mt.CarmelPub.	0.493	0.493	0.000	1.000
120	IowaIII.G.&E.	0.519	0.519	0.000	1.000
121	N'westernP.S.	0.702	0.702	0.000	1.000
122	Cal.Pac.Util	0.806	0.806	0.000	1.000
123	Ctrl.Ver.Pub.Ser.	1.294	1.294	0.000	1.000



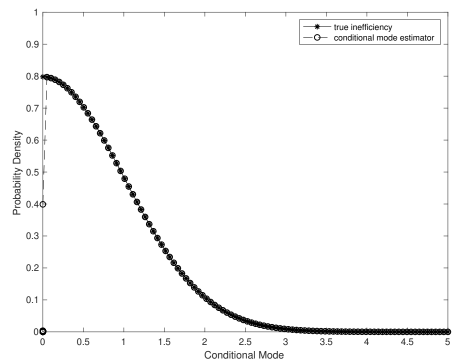
(a) $\sigma_v = 10$, $\sigma_u = 1$.



(b) $\sigma_v = 1$, $\sigma_u = 1$.

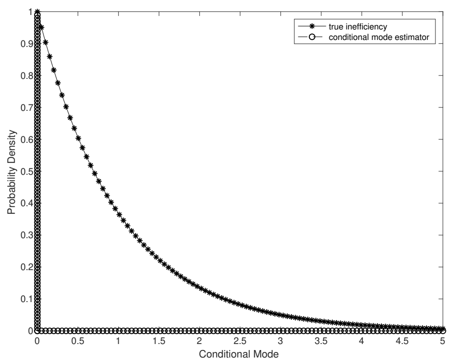


(c) $\sigma_v = 0.1$, $\sigma_u = 1$.

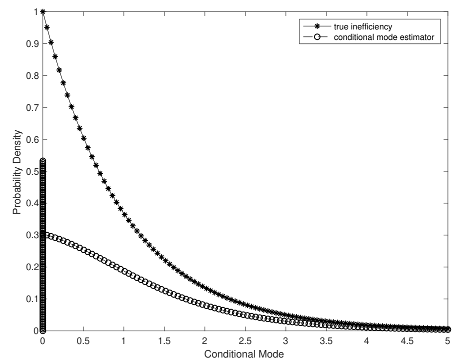


(d) $\sigma_v = 0.01$, $\sigma_u = 1$.

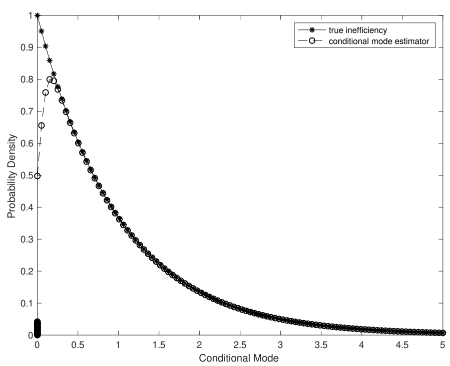
Figure 1: Distributions of $\hat{u}^m$ under Normal-Half Normal Distribution



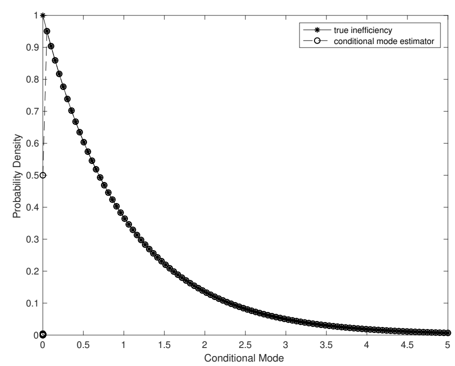
(a) $\sigma_v = 10$, $\sigma_u = 1$.



(b) $\sigma_v = 1$, $\sigma_u = 1$.

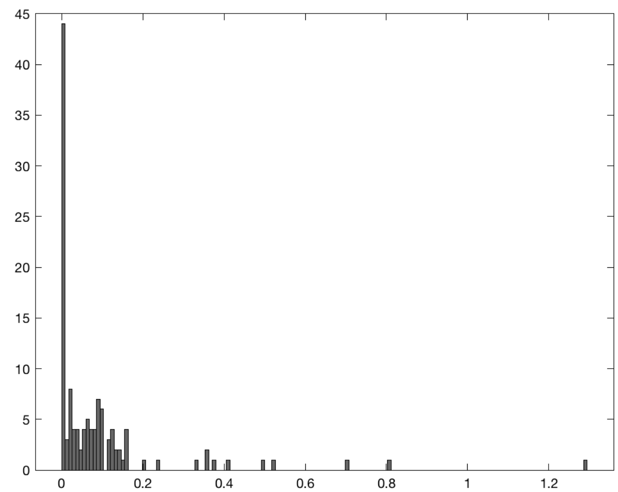


(c) $\sigma_v = 0.1$, $\sigma_u = 1$.

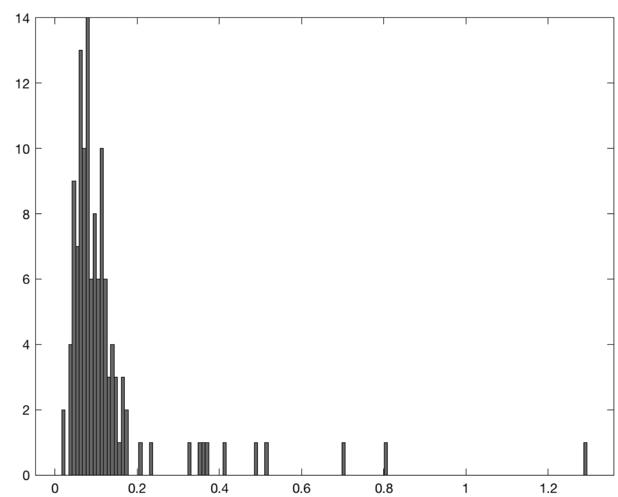


(d) $\sigma_v = 0.01$, $\sigma_u = 1$.

Figure 2: Distributions of $\hat{u}^m$ under Normal-Exponential Distribution



(a) Conditional Mode



(b) Conditional Expectation

Figure 3: Histograms of Technical Inefficiency Estimates of Electric Utility Firms

Appendix

Some Results on Conditional Mode under Laplace-Half Normal Distribution

When v is distributed Laplace with a mean of zero, $f_v(v) = \frac{1}{2\sigma_v}\exp(-\frac{|v|}{\sigma_v})$, and u is distributed half normal, $f_u(u) = \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_u}\exp(-\frac{u^2}{2\sigma_u^2})$, we have $f_{v,u}(\varepsilon+u,u) = \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v}\exp(-\frac{|\varepsilon+u|}{\sigma_v} - \frac{u^2}{2\sigma_u^2})$. Integrating out u gives

$$f(\varepsilon) = \begin{cases} \frac{1}{\sigma_v}\exp\left(-\frac{\varepsilon}{\sigma_v} + \frac{\sigma_u^2}{2\sigma_v^2}\right)\left[1 - \Phi(\frac{\sigma_u}{\sigma_v})\right] & \varepsilon \geq 0 \\ \frac{1}{\sigma_v}\exp(\frac{\sigma_u^2}{2\sigma_v^2})\left[\exp(\frac{\varepsilon}{\sigma_v})\left\{\Phi(-\frac{\varepsilon}{\sigma_u} - \frac{\sigma_u}{\sigma_v}) - \Phi(\frac{-\sigma_u}{\sigma_v})\right\} + \exp(-\frac{\varepsilon}{\sigma_v})\left\{1 - \Phi(-\frac{\varepsilon}{\sigma_u} + \frac{\sigma_u}{\sigma_v})\right\}\right] & \varepsilon < 0 \end{cases}$$

Since $f_u(u|\varepsilon) \propto f_u(u)f_v(\varepsilon+u)$, finding the conditional mode is equivalent to finding the mode of $f_u(u)f_v(\varepsilon+u)$ when ε is given.

When $u+\varepsilon > 0$, we can write

$$f_u(u)f_v(\varepsilon+u) = \frac{1}{\sqrt{2\pi}\sigma_u}\exp\left(-\frac{(u+\frac{\sigma_u^2}{\sigma_v})^2}{2\sigma_u^2}\right) \times \left[\frac{1}{\sigma_v}\exp(-\frac{\varepsilon}{\sigma_v} + \frac{\sigma_u^2}{2\sigma_v^2})\right].$$

When $u+\varepsilon < 0$, $0 < u < -\varepsilon$,

$$f_u(u)f_v(\varepsilon+u) = \frac{1}{\sqrt{2\pi}\sigma_u}\exp\left(-\frac{(u-\frac{\sigma_u^2}{\sigma_v})^2}{2\sigma_u^2}\right) \times \left[\frac{1}{\sigma_v}\exp(\frac{\varepsilon}{\sigma_v} + \frac{\sigma_u^2}{2\sigma_v^2})\right].$$

Proof of Theorem 1

Case I: $\sigma_v^2 \to 0$ under normal-half normal

As $\sigma_v^2 \to 0$, the distribution of v collapses into a Dirac Delta function at 0, which leads to $\varepsilon \equiv v - u \xrightarrow{p} -u$. Then, since $\frac{\sigma_u^2}{\sigma^2} \to 1$ as $\sigma_v^2 \to 0$, where $\sigma^2 = \sigma_v^2 + \sigma_u^2$, it follows that $\hat{u}^m \equiv -\frac{\sigma_u^2}{\sigma^2}\varepsilon \xrightarrow{p} u$. Furthermore, due to the non-negativity constraint $u \geq 0$, this implies that the probability to observe negative $\hat{u}^m$ becomes zero as $\sigma_v^2 \to 0$.

Also, note that $\hat{u}^m \xrightarrow{p} u$ implies that the unconditional distribution of $\hat{u}^m$ should be identical to the distribution of u in the limit.

Also, when u is fixed, it can be shown that $(\hat{u}^m - u)/\sigma_v \to -\frac{v}{\sigma_v} \sim N(0,1)$ as $\sigma_v^2 \to 0$, since $\left(\frac{\sigma_u^2}{\sigma^2} - 1\right)\frac{u}{\sigma_v} = -\frac{\sigma_v}{\sigma^2}u \xrightarrow{p} 0$ as $\sigma_v^2 \to 0$.

Case II: $\sigma_v^2 \to \infty$ under normal-half normal:

Since $\frac{v}{\sigma_v^2} \sim N(0, 1/\sigma_v^2)$ and $\frac{u}{\sigma_v^2} \sim N(0, \sigma_u^2/\sigma_v^4)$, it can be shown that $\hat{u}^m = -\frac{\sigma_u^2}{1+\sigma_u^2/\sigma_v^2}\frac{v-u}{\sigma_v^2} \xrightarrow{p} 0$ as $\sigma_v^2 \to \infty$ for a fixed σ_u^2. Also, from the fact that $\frac{\sigma_v}{\sigma_u^2}\hat{u}^m = \frac{\sigma_v^2}{\sigma_u^2+\sigma_v^2}\left(\frac{u}{\sigma_v} - \frac{v}{\sigma_v}\right)$, where $\frac{u}{\sigma_v} \xrightarrow{p} 0$ and $\frac{\sigma_v^2}{\sigma_u^2+\sigma_v^2} \to 1$ as $\sigma_v^2 \to \infty$, it follows that $\frac{\sigma_v}{\sigma_u^2}\hat{u}^m \xrightarrow{d} -\frac{v}{\sigma_v} \sim N(0,1)$ for $\hat{u}^m > 0$ and half of the probability mass is concentrated at $\hat{u}^m = 0$.

Case III: $\sigma_v^2 \to 0$ under normal-exponential:

Similarly as in Case I, as $\sigma_v^2 \to 0$, $\varepsilon \xrightarrow{p} -u$ and $\frac{\sigma_v^2}{\sigma_u} \to 0$. Therefore, $\hat{u}^m = -\varepsilon - \sigma_v^2/\sigma_u \xrightarrow{p} u$.

Also, for a given u, as $\sigma_v^2 \to 0$, $(\hat{u}^m - u)/\sigma_v \to -\frac{v}{\sigma_v} \sim N(0,1)$ since $\frac{\sigma_v}{\sigma_u} \to 0$.

Case IV: $\sigma_v^2 \to \infty$ under normal-exponential:

Note that $\hat{u}^m = \sigma_v\left(\frac{u}{\sigma_v} - \frac{v}{\sigma_v} - \frac{\sigma_v}{\sigma_u}\right)$, where $\frac{u}{\sigma_v} \xrightarrow{p} 0$, $\frac{v}{\sigma_v} = O_p(1)$ and $\frac{\sigma_v}{\sigma_u} \to \infty$ as $\sigma_v^2 \to \infty$, from which it immediately follows that $\hat{u}^m \xrightarrow{p} 0$ as $\sigma_v^2 \to \infty$.

Proof of Theorem 2

We follow the proof of Theorem 3 in Zou (2006), but account for the positivity constraint on inefficiency parameters. We prove for the univariate case here and the multivariate case follows by summation.

Let $\varepsilon \sim N(-u, \sigma_v^2)$ where $u \geq 0$, and $\hat{u}^m(\lambda) = [-\varepsilon - \lambda]_+$ with $\lambda = \sigma_v \sqrt{2 \log n}$. We first expand the mean squared error of $\hat{u}^m(\lambda)$ such that

$$
\begin{aligned}
E[(\hat{u}^m(\lambda) - u)^2] &= E[(\hat{u}^m(\lambda) + \varepsilon)^2] + E[(-\varepsilon - u)^2] - 2E[\hat{u}^m(\lambda)(u + \varepsilon)] - 2E[\varepsilon(u + \varepsilon)] \\
&= E[(\hat{u}^m(\lambda) + \varepsilon)^2] + \sigma_v^2 - 2E[\hat{u}^m(\lambda)(u + \varepsilon)] - 2\sigma_v^2 \\
&= E[(\hat{u}^m(\lambda) + \varepsilon)^2] - 2\sigma_v^2 E\left[\frac{\partial \hat{u}^m(\lambda)}{\partial \varepsilon}\right] - \sigma_v^2
\end{aligned}
$$

where we have used the Stein's lemma (Stein, 1981) in the last equality, i.e., $E[\hat{u}^m(\lambda)(u+\varepsilon)] = \sigma_v^2 E\left[\frac{\partial \hat{u}^m(\lambda)}{\partial \varepsilon}\right]$. Note that

$$
\begin{cases}
(\hat{u}^m(\lambda) + \varepsilon)^2 = \varepsilon^2 \text{ and } \frac{\partial \hat{u}^m(\lambda)}{\partial \varepsilon} = 0, \text{ if } \varepsilon \geq -\lambda \\
(\hat{u}^m(\lambda) + \varepsilon)^2 = \lambda^2 \text{ and } \frac{\partial \hat{u}(\lambda)}{\partial \varepsilon} = -1, \text{ if } \varepsilon < -\lambda
\end{cases}
$$

Thus, we have

$$
E[(\hat{u}^m(\lambda) - u)^2] = E[\varepsilon^2 \cdot I(\varepsilon \geq -\lambda)] + E[(\lambda^2 + 2\sigma_v^2) \cdot I(\varepsilon < -\lambda)] - \sigma_v^2. \qquad (15)
$$

Note that

$$
\begin{aligned}
E[\varepsilon^2 \cdot I(\varepsilon \geq -\lambda)] &= E[\varepsilon^2 \cdot I(-\lambda \leq \varepsilon < \lambda)] + E[\varepsilon^2 \cdot I(\lambda \leq \varepsilon)] \\
&\leq \lambda^2 P(-\lambda \leq \varepsilon < \lambda) + \frac{u^2}{2n\sqrt{\pi \log n}} + \frac{1}{2}\sigma_v^2,
\end{aligned}
$$

since

$$
\begin{aligned}
E[\varepsilon^2 \cdot I(\lambda \leq \varepsilon)] &= u^2 P(v \geq u + \lambda) - 2uE[v \cdot I(v \geq u + \lambda)] + E[v^2 \cdot I(v \geq u + \lambda)] \\
&\leq \frac{u^2}{2n\sqrt{\pi \log n}} + \frac{1}{2}\sigma_v^2,
\end{aligned}
$$

where we used the fact that $P(v \geq u + \lambda) \leq \Phi\left(-\sqrt{2 \log n}\right) \leq \int_{-\infty}^{-\sqrt{2 \log n}} \frac{x}{-\sqrt{2 \log n}} \phi(x) dx = \frac{1}{\sqrt{2 \log n}} \int_{-\infty}^{-\sqrt{2 \log n}} \phi'(x) dx = \frac{1}{\sqrt{2 \log n}} \phi(-\sqrt{2 \log n}) = \frac{1}{2n\sqrt{\pi \log n}}$, $E[v \cdot I(v \geq u + \lambda)] > 0$, and $E[v^2 \cdot I(v \geq u + \lambda)] \leq \frac{1}{2}\sigma_v^2$.

Then, it follows that

$$
\begin{aligned}
E[(\hat{u}^m(\lambda) - u)^2] &\leq \lambda^2 P(-\lambda \leq \varepsilon \leq \lambda) + (\lambda^2 + 2\sigma_v^2) P(\varepsilon < -\lambda) + \frac{u^2}{2n\sqrt{\pi \log n}} - \frac{1}{2}\sigma_v^2 \\
&= \lambda^2 P(\varepsilon \leq \lambda) + 2\sigma_v^2 P(\varepsilon < -\lambda) + \frac{u^2}{2n\sqrt{\pi \log n}} - \frac{1}{2}\sigma_v^2 \\
&\leq \lambda^2 + \frac{u^2}{2n\sqrt{\pi \log n}} + \frac{3}{2}\sigma_v^2 \\
&= \left(\frac{\lambda^2}{\sigma_v^2} + \frac{3}{2}\right)\left(\sigma_v^2 + \frac{u^2}{2n\sqrt{\pi \log n}(\lambda^2/\sigma_v^2 + 3/2)}\right).
\end{aligned}
\tag{16}
$$

From (15), we also have

$$
\begin{aligned}
E[(\hat{u}^m(\lambda) - u)^2] &= E[\varepsilon^2] + E[(\lambda^2 + 2\sigma_v^2 - \varepsilon^2) \cdot I(\varepsilon < -\lambda)] - \sigma_v^2 \\
&\leq u^2 + 2\sigma_v^2 P(\varepsilon < -\lambda),
\end{aligned}
$$

where the inequality is due to $E[(\lambda^2 - \varepsilon^2) \cdot I(\varepsilon < -\lambda)] < 0$. Then, it can be verified that

$P(\varepsilon < -\lambda) \leq 5\Phi\left(-\sqrt{2\log n}\right) + \frac{\phi(0)}{2\lambda\sigma_v}u^2$ for any $u \geq 0$ and $n \geq 2$,[14] from which it follows that

$$
\begin{aligned}
E[(\hat{u}^m(\lambda) - u)^2] &\leq u^2 + 2\sigma_v^2 \left(5\Phi\left(-\sqrt{2\log n}\right) + \frac{\phi(0)}{2\lambda\sigma_v}u^2\right) \\
&\leq \left(\frac{\lambda^2}{\sigma_v^2} + 1\right)\left(u^2 + \frac{5\sigma_v^2}{n\sqrt{\pi\log n}(\lambda^2/\sigma_v^2 + 1)}\right),
\end{aligned}
\tag{17}
$$

where we used the fact $\frac{\phi(0)}{\lambda\sigma_v} \leq \frac{\lambda^2}{\sigma_v^4}$ for any $n \geq 2$.

Then, equations (16) and (17) imply

$$
E[(\hat{u}^m(\lambda) - u)^2] \leq \left(\frac{\lambda^2}{\sigma_v^2} + \frac{3}{2}\right)\left(\min(u^2, \sigma_v^2) + \frac{5(u^2 + \sigma_v^2)}{n\sqrt{\pi}(\log n)^{3/2}}\right).
$$

Math Correction in Horrace and Parmeter (2018)

- In equation(5), under $\varepsilon < 0$, $\theta \neq \ $, $f_\varepsilon(\varepsilon, \mu \leq 0) = \frac{1}{2\ \theta}[(\lambda_+ - \lambda_-)e^{\varepsilon/\theta} + \lambda_- e^{\varepsilon/\gamma}]$. Under $\varepsilon < 0$ and $\theta = \ $, $f_\varepsilon(\varepsilon, \mu \leq 0) = \frac{1}{2\ \theta}(\lambda_+ - \varepsilon)e^{\varepsilon/\theta}$.

- In the equation below equation (8), $f_u(u|\varepsilon) = \frac{c(\mu_*)}{4\ \theta f_\varepsilon(0)}e^{-\frac{|u-\mu_*|}{\theta} - \frac{u}{\ }}$.

- The log-likelihood function under section 3.3 is $\ln L(\varepsilon_i|\ ,\ ,\theta) = const. - \sum_{i:\varepsilon_i \geq 0}[\ln(\ + \theta) + \varepsilon_i/\gamma] + \sum_{i:\varepsilon_i<0}\ln[\frac{1}{\ -\theta}(e^{\varepsilon_i/\gamma} - e^{\varepsilon_i/\theta}) + \frac{1}{\ +\theta}e^{\varepsilon_i/\theta}]$.

[14] Let $g(u) = \Phi\left(\frac{u-\lambda}{\sigma_v}\right)\left(= P(\varepsilon < -\lambda)\right)$. Note that $g(u)$ is monotonically increasing with u such that $g'(u) = \frac{1}{\sigma_v}\phi\left(\frac{u-\lambda}{\sigma_v}\right)$. Then, we consider a quadratic function $z(u) = a + \frac{1}{2}bu^2$ that $z(u) \geq g(u)$ for any $u \geq 0$ and $n \geq 2$. Since the derivative of $g(u)$ increases until $u = \lambda$ and then decreases whereas the derivative of $z(u)$ is monotonically increasing, one sufficient condition for such $z(u)$ is that $z(0) \geq g(0)$, $z(\lambda) \geq g(\lambda)$ and $z'(\lambda) \geq \frac{1}{\sigma_v}\phi(0)$. It can be verified that the condition is satisfied when $a = 5\Phi\left(-\sqrt{2\log n}\right)$ and $b = \frac{\phi(0)}{\lambda\sigma_v}$.