2
2006

This is to certify that the
dissertation entitled

THREE ESSAYS ON ECONOMETRICS

presented by

MYUNGSUP KIM

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Economics

Peter Schmidt
Major Professor's Signature

JULY 14, 2005
Date

THREE ESSAYS ON ECONOMETRICS

By

Myungsup Kim

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

2005

ABSTRACT

THREE ESSAYS ON ECONOMETRICS

By

Myungsup Kim

Consider a simple stochastic frontier model explaining the output of a firm by $y = x'\beta + v - u$. While v represents random shocks outside the control of producers, u represents technical inefficiency in the production process.

In the first chapter, we wish to test whether technical inefficiency depends on observable characteristics of the firm. It is well known that two-step procedures, in which the second step is the regression of an inefficiency measure on firm characteristics, do not properly estimate the effects of firm characteristics on inefficiency. In this chapter we show that this regression also does not lead to a valid test of the hypothesis of no effect. A valid test of the hypothesis of no effect can be constructed by using an adjustment to the variance matrix of the estimated coefficients in the second step regression. Unfortunately the form of this adjustment is not distribution free. We show that this test is the LM test in the specific case that technical inefficiency is exponential and the alternative is a scaled exponential distribution. We also consider tests based on nonlinear least squares. These tests do not depend on a distributional assumption. There are some technical complications involved due to the non-identification of some of the parameters under the null. We perform an extensive set of simulations to compare the size and power characteristics of these tests and other similar tests, including the Wald test based on a one-step estimate of the entire model.

In the second chapter, we study the construction of confidence intervals for efficiency levels of individual firms in stochastic frontier models with panel data. The focus is on bootstrapping and related methods. We start with a survey of various versions of the bootstrap. Then we offer some simple alternatives based on standard

methods when one acts as if the identity of the best firm is known. Monte Carlo simulations indicate that these simple alternatives work better than the percentile bootstrap but perhaps not as well as the bias-adjusted and accelerated bootstrap. None of the methods yields very accurate confidence intervals except when the time-series sample size is large enough, or the error variance is small enough, that the identity of the best firm is clear. We also present empirical results for two well-known data sets.

In the last chapter, we consider the problem of testing the null hypothesis that a series is stationary against the unit root alternative. A standard test for this null hypothesis is the KPSS test, which is based on cumulations of deviations from the means of the series. A paper by de Jong, Amsler, and Schmidt (2002) constructs a "robust" version of the KPSS test by using an indicator of whether the observation is above or below the sample median. This test, called the indicator KPSS test, is robust in that it does not require existence of moments of the series, yet the asymptotic distribution of the indicator KPSS statistic is the same as that of the KPSS statistic. However, in this chapter we allow a non-zero level for the series under consideration, but not a deterministic trend. The purpose of this chapter is to extend the indicator KPSS statistic to the case of a deterministic trend. The relevant indicator in this setting is whether the residual is positive or negative in a least absolute deviations regression of the series on a time trend. This chapter shows that, under the null of trend-stationarity, the indicator KPSS statistic with a time trend has the same limiting distribution as the KPSS statistic with a time trend.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Professor Peter Schmidt, who taught my first Econometrics course with passion. His willingness to motivate and support my work has made it possible for me to develop my skills as a researcher. Needless to say, his wise advice and continuing counsel have been essential for the course of my studies. Also, I am enormously grateful to Professor Robert M. de Jong for his patient support and extraordinary encouragement of my learning of methodological tools. I owe special thanks to Professor Jeffrey M. Wooldridge for invaluable comments. I would like to thank the other members of my committee: Professor Christine E. Amsler and Professor Robert J. Myers. I greatly appreciate their input and time to this dissertation.

I want to thank my wife, Jiyoung, for being my partner and best friend with her unending love, sacrifice and support. I am also greatly indebted to my parents and my parents-in-law who have always been very supportive of my pursuit of education.

My special thanks also go to my fellow graduate students for their friendships and the staff in the Department of Economics for their jovial smile and helpful nature that have always greeted any problem or deadline.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Valid Tests of Whether Technical Inefficiency Depends on Firm Characteristics

1.1 Introduction

In this chapter we consider the stochastic frontier model

$$y_i = x_i'\beta + v_i - u_i, \quad u_i \geq 0. \tag{1.1}$$

The frontier is $y_i^* = x_i'\beta + v_i$ and u_i represents technical inefficiency. We follow the literature in assuming that the x_i are "fixed" and the v_i are i.i.d. normal. Now we ask whether u_i depends on some variables z_i, which could be characteristics of the firm or measures of the environment in which it operates. Specifically, we wish to test the hypothesis that u_i does not depend on z_i.

One way to do this is to assume a specific model of the alternative hypothesis that shows how the z_i affect the u_i. For example, we could assume:

$$u_i = \exp(z_i'\delta) \cdot u_i^\circ, \tag{1.2}$$

where the u_i° are i.i.d. according to some specific distribution, like exponential or half-normal. Now we can estimate δ by MLE and do a Wald test of the hypothesis that $\delta = 0$, which corresponds to the hypothesis that z_i does not affect u_i. In the frontiers literature this would correspond to what is called a "one-step" procedure (e.g., see Wang and Schmidt (2002)). Models of the form of (1.2) have been considered by Reifschneider and Stevenson (1991), Caudill and Ford (1993), Caudill, Ford, and Gropper (1995), Wang and Schmidt (2002) and Alvarez, Amsler, Orea, and Schmidt (2005), among others. We will follow the literature and call the multiplicative decomposition of u_i (as a function of z_i times a random variable that does not depend on z_i) the "scaling property."

An objection to this type of procedure is that it depends fundamentally on the alternative chosen. Under the null the scaling function $\exp(z_i'\delta)$ really does not exist and so there are many more or less equally plausible alternatives. Partly for this reason, one could consider a "two-step procedure" in which Step 1 would be to estimate the model ignoring the z_i to obtain efficiency measures $\hat{u}_i$, and Step 2 would be a regression of $\hat{u}_i$ on z_i (or some function of z_i). It is well known (Wang and Schmidt (2002)) that when z_i does affect u_i, there are serious biases in both steps, so two-step procedures are not recommended. However, under the null that z_i does not affect u_i, these biases do not arise, and it is not known whether a two-step procedure provides a valid test of this null hypothesis. One contribution of this chapter is to show that a two-step procedure that uses a standard t or F test in the second step does not yield an asymptotically valid test. However, the test becomes valid if we use a corrected variance matrix for the second-step coefficients. Unfortunately, the form of this correction is distribution-specific.

This raises the question of whether a test based on such a corrected two-step procedure entails a loss of power. We do not have a full answer to this question. We do show that, in the case that the alternative is the scaled exponential distribution, the LM test of $\delta = 0$ is asymptotically equivalent to the corrected version of the two-step procedure. Therefore at least in this case the two-step procedure entails no loss in asymptotic local power.

If we assume the scaling property, as in (1.2) above, the stochastic frontier model can also be estimated by nonlinear least squares. Testing whether $\delta = 0$ based on nonlinear least squares involves some technical difficulties, because the mean of u_i° is identified separately from the overall intercept under the alternative but not under the null. We show how to deal with these difficulties and obtain an asymptotically valid test.

In the last section of the chapter, we report the results of an extensive set of simulations that investigate the size and power of these tests.

1.2 Two-Step Procedures

We consider the stochastic frontier model (1.1). As stated in the Introduction, we treat the x_i as fixed and we assume that the v_i are i.i.d. $N(0, \sigma_v^2)$. We also assume that the u_i° are i.i.d. with some specific distribution, such as exponential or half-normal, that is known up to some parameters. Finally, the z_i variables whose influence on u_i we wish to test are independent of v_i and u_i°. For the purposes of this section, these assumptions could be weakened somewhat, but we would need the stronger set subsequently, so we simply make them here.

To motivate the tests considered here, suppose that u_i were observed. Then we could regress u_i on z_i and test the hypothesis that the coefficients equal zero by standard methods. More precisely, the regression would have to include an intercept because $E(u_i)$ is not equal to zero, and we would do an F-test on the coefficients other than the intercept.

Now let ψ equal the unknown parameters of the problem. These would be β, σ_v^2 and whatever parameters there are in the distribution of u_i°. Step 1 of the two-step procedure results in an estimate $\hat{\psi}$ which should be consistent and asymptotically normal (subject to the usual regularity conditions). We then obtain an estimate of u_i, say $\hat{u}_i(\hat{\psi})$. In the stochastic frontier model, $\hat{u}_i$ is the expected value of u_i conditional on $\epsilon_i \equiv v_i - u_i$, evaluated at the sample estimates, as suggested by Jondrow, Lovell, Materov, and Schmidt (1982). It should be noted that, even if ψ were known, $\hat{u}_i(\psi)$ would be $E(u_i|\epsilon_i)$ which is different from u_i. However, $\hat{u}_i(\psi)$ is a function of ϵ_i, which is i.i.d. and independent of z_i. So, if we regressed $u_i(\psi)$ on intercept and z_i, an F-test of the significance of the coefficients of z_i should be asymptotically valid. The question is whether this is still true when $u_i(\psi)$ is replaced by $u_i(\hat{\psi})$. Unfortunately, the answer is no. A valid test must account for the estimation error in $\hat{\psi}$.

To show this, we could consider a regression of $\hat{u}_i(\hat{\psi})$ on intercept and z_i. However, it is simpler to demean the $\hat{u}_i$ by switching our attention to $b_i(\psi) = E(u_i|\epsilon_i) - E(u_i)$, with $\hat{b}_i = b_i(\hat{\psi})$ being the corresponding estimate evaluated at the first-step estimates $\hat{\psi}$. So now we simply wish to test whether $\gamma = 0$ in the regression:

$$\hat{b}_i = z_i'\gamma + \nu_i. \tag{1.3}$$

Our test statistic will be $\hat{\gamma}'[\widehat{\text{Var}(\hat{\gamma})}]^{-1}\hat{\gamma}$, where $\hat{\gamma}$ is the least squares estimate from (1.3), and this should be asymptotically χ^2, if $\text{Var}(\hat{\gamma})$ is properly calculated.

This is a "generated dependent variable" problem that can be analyzed by methods similar to those used for the "generated regressor" problem (e.g., Wooldridge (2002)[pp. 139-141]). We have $b_i = b_i(\psi) = f(y_i, x_i, \psi)$ and $\hat{b}_i = b_i(\hat{\psi}) = f(y_i, x_i, \hat{\psi})$.

By the Mean Value Theorem,

$$\hat{b}_i = b_i + \nabla_\psi f(y_i, x_i, \ddot{\psi})'(\hat{\psi} - \psi) \tag{1.4}$$

where $\ddot{\psi}$ is between ψ and $\hat{\psi}$. Therefore

$$\begin{aligned}
\sqrt{N}\hat{\gamma} &= \left(\frac{1}{N}\sum_{i=1}^{N} z_i z_i'\right)^{-1} \frac{1}{\sqrt{N}}\sum_{i=1}^{N} z_i \hat{b}_i \\
&= \left(\frac{1}{N}\sum_{i=1}^{N} z_i z_i'\right)^{-1} \left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} z_i b_i \right. \\
&\quad \left. + \frac{1}{N}\sum_{i=1}^{N} z_i \nabla_\psi f(y_i, x_i, \ddot{\psi})' \sqrt{N}(\hat{\psi} - \psi)\right] \\
&= \left(\frac{1}{N}\sum_{i=1}^{N} z_i z_i'\right)^{-1} \left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} z_i b_i \right. \\
&\quad \left. + \frac{1}{N}\sum_{i=1}^{N} z_i \nabla_\psi f(y_i, x_i, \psi)' \sqrt{N}(\hat{\psi} - \psi)\right] + o_p(1).
\end{aligned} \tag{1.5}$$

From the last line of equation (1.5), we can see immediately that the term involving the estimation error in ψ will be relevant unless $\mathrm{E}[z_i \nabla_\psi f(y_i, x_i, \psi)] = 0$. (In this exceptional case, $N^{-1}\sum_{i=1}^{N} z_i \nabla_\psi f(y_i, x_i, \psi) \xrightarrow{p} 0$ and the last term vanishes. Otherwise it does not.)

To proceed further, we use the same device as in Wooldridge (2002), and assume that

$$\sqrt{N}(\hat{\psi} - \psi) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} r_i(\psi) + o_p(1), \tag{1.6}$$

where $\mathrm{E}\, r_i(\psi) = 0$. We will be more specific about the form of $r_i(\psi)$, below. Then

$$\sqrt{N}\hat{\gamma} = \left(\frac{1}{N}\sum_{i=1}^{N} z_i z_i'\right)^{-1} \left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} z_i b_i + \frac{1}{N}\sum_{i=1}^{N} z_i \nabla_\psi f(y_i, x_i, \psi)' \frac{1}{\sqrt{N}}\sum_{i=1}^{N} r_i(\psi)\right] + o_p(1). \tag{1.7}$$

It follows from a central limit theorem applied to (1.7) that

$$\sqrt{N}\hat{\gamma} \to N(0, B^{-1}AB^{-1}) \tag{1.8}$$

where

$$B = \mathrm{E}\, z_i z_i', \tag{1.9a}$$

$$A = \mathrm{E}[(z_i b_i + G r_i)(z_i b_i + G r_i)'], \tag{1.9b}$$

$$G = \mathrm{E}\, z_i \nabla_\psi f(y_i, x_i, \psi)'. \tag{1.9c}$$

Also, all of these quantities can be consistently estimated by the corresponding sample quantities: $\hat{B} = N^{-1}\sum_{i=1}^{N} z_i z_i'$, $\hat{A} = N^{-1}\sum_{i=1}^{N}[(z_i\hat{b}_i + \hat{G}\hat{r}_i)(z_i\hat{b}_i + \hat{G}\hat{r}_i)']$, $\hat{G} = N^{-1}\sum_{i=1}^{N} z_i \nabla_\psi f(y_i, x_i, \hat{\psi})'$.

The remaining detail is an expansion for r_i. The first-step MLE $\hat{\psi}$ satisfies $\sum_{i=1}^{N} s_i(\hat{\psi}) = 0$, where $s_i(\psi)$ is the score function for observation i. (That is, $s_i(\psi)$ is the derivative with respect to ψ of the i^{th} observation's contribution to the log likelihood). Then another Mean Value Theorem expansion yields

$$0 = \sum_{i=1}^{N} s_i(\hat{\psi}) = \sum_{i=1}^{N} s_i(\psi) + \sum_{i=1}^{N} \nabla_\psi s_i(\ddot{\psi})(\hat{\psi} - \psi), \tag{1.10}$$

where $\ddot{\psi}$ is between $\hat{\psi}$ and ψ. So

$$\begin{aligned}\sqrt{N}(\hat{\psi}-\psi) &= \left[-\frac{1}{N}\sum_{i=1}^{N}\nabla_\psi s_i(\psi)\right]^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}s_i(\psi)+o_p(1)\\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathcal{I}^{\circ -1}s_i(\psi)+o_p(1)\end{aligned} \tag{1.11}$$

where

$$\mathcal{I}^\circ = \mathrm{E}\, s_i(\psi)s_i(\psi)' = -\,\mathrm{E}\,\nabla_\psi s_i(\psi) = \lim_{N\to\infty}\frac{1}{N}\mathcal{I}, \tag{1.12}$$

and

$$\mathcal{I} = \mathrm{E}(\nabla_\psi \ln\mathrm{L})(\nabla_\psi \ln\mathrm{L})' = -\,\mathrm{E}\,\nabla_\psi^2 \ln\mathrm{L}\,. \tag{1.13}$$

$\mathcal{I}$ is the information matrix for the first-step MLE problem with a log-likelihood of lnL, and $\mathcal{I}^\circ$ is the limiting information matrix. In terms of the score, $\mathcal{I} = \sum_{i=1}^{N}\mathrm{E}\, s_i(\psi)s_i(\psi)' = -\sum_{i=1}^{N}\mathrm{E}\,\nabla_\psi s_i(\psi)$. Therefore, in (1.6) and the subsequent expressions above, $r_i(\psi) = \mathcal{I}^{\circ -1}s_i(\psi)$. In terms of sample quantities, $\hat{r}_i = \hat{\mathcal{I}}^{\circ -1}s_i(\psi)$ where $\hat{\mathcal{I}}^\circ = N^{-1}\sum_{i=1}^{N}s_i(\hat{\psi})s_i(\hat{\psi})'$.

We note two things. First, the standard (naive) test of $\gamma = 0$ that ignores the effect of estimation error in $\hat{b}_i$ corresponds to omitting the terms corresponding to Gr_i in (1.9b). This test will be invalid unless $G = 0$. Since $G = \mathrm{E}\, z_i\nabla_\psi f(y_i, x_i, \psi)$, this condition will hold if z_i is independent of x_i as well as of v_i and u_i. However, it will generally fail if z_i and x_i are correlated. Second, the "correct" test is not difficult. However, unsurprisingly, the form of the correction depends on the distribution of u_i, since that influences the nature of the first-step MLE problem. There is no simple, distribution-free correction.

1.3 The Scaled Exponential Case

In this section we consider the special case that u_i follows a scaled exponential distribution. That is, $u_i = \exp(z_i'\delta) \cdot u_i^\circ$, as in (1.2), where u_i° is distributed as exponential with parameter λ. We will derive the LM test of the hypothesis $\delta = 0$, and show that it is asymptotically equivalent to the (corrected) two-step procedure of the last section. This shows that there is at least one case in which the two-step procedure does not entail any loss of (local) power, compared to the usual Wald-likelihood ratio-LM trinity of tests.

For the normal-scaled exponential model we consider, the pdf of the composite error ($\epsilon_i = v_i - u_i$) is:

$$f(\epsilon_i) = \frac{1}{\lambda \exp(z_i'\delta)} \cdot \exp\Big(\frac{\epsilon_i}{\lambda \exp(z_i'\delta)} + \frac{\sigma_v^2}{2\lambda^2 \exp(2z_i'\delta)}\Big) \cdot \Big(1 - \Phi\big(\frac{\epsilon_i}{\sigma_v} + \frac{\sigma_v}{\lambda \exp(z_i'\delta)}\big)\Big) \tag{1.14}$$

where Φ is the cumulative distribution function of the standard normal distribution. Note that under the null of $\delta = 0$, $\mathrm{E}(\epsilon_i) = -\mathrm{E}(u_i) = -\lambda$ and $\mathrm{Var}(\epsilon_i) = \sigma_v^2 + \lambda^2$. Also, the distribution of u_i given ϵ_i is $N(-\epsilon_i - \sigma_v^2/(\lambda \exp(z_i'\delta)),\ \sigma_v^2)^+$ where "+" represents truncation on the left at zero.

From (1.14), it follows that the log-likelihood function $\mathrm{lnL}(\delta, \beta, \sigma_v^2, \lambda^2) = \mathrm{lnL}(\theta)$ is given by:

$$\begin{aligned} \mathrm{lnL}(\theta) = &-\sum_{i=1}^{N} \ln(\lambda \exp(z_i'\delta)) + \sum_{i=1}^{N} \frac{\epsilon_i}{\lambda \exp(z_i'\delta)} \\ &+ \sum_{i=1}^{N} \frac{\sigma_v^2}{2\lambda^2 \exp(2z_i'\delta)} + \sum_{i=1}^{N} \ln\left(1 - \Phi\left(\frac{\epsilon_i}{\sigma_v} + \frac{\sigma_v}{\lambda \exp(z_i'\delta)}\right)\right). \end{aligned} \tag{1.15}$$

The generic form of the LM statistic is

$$\text{LM} = \nabla_\theta \ln\text{L}(\tilde{\theta})' \cdot \mathcal{I}^{-1}(\tilde{\theta}) \cdot \nabla_\theta \ln\text{L}(\tilde{\theta}). \tag{1.16}$$

Here $\tilde{\theta}$ is the MLE subject to the restriction $\delta = 0$; $\mathcal{I}(\tilde{\theta})$ is the information matrix evaluated at $\theta = \tilde{\theta}$; and $\nabla_\theta \ln\text{L}(\tilde{\theta})$ is the score function, $\nabla_\theta \ln\text{L}(\theta)$, evaluated at $\theta = \tilde{\theta}$. If we partition $\theta = (\delta', \psi')$, where $\psi = (\beta', \sigma_v^2, \lambda^2)'$, then

$$\nabla_\theta \ln\text{L}(\theta) = \begin{pmatrix} \nabla_\delta \ln\text{L}(\theta) \\ \nabla_\psi \ln\text{L}(\theta) \end{pmatrix}, \quad \mathcal{I}(\theta) = \begin{pmatrix} \mathcal{I}_{\delta\delta} & \mathcal{I}_{\delta\psi} \\ \mathcal{I}_{\psi\delta} & \mathcal{I}_{\psi\psi} \end{pmatrix}. \tag{1.17}$$

It is a standard result that $\nabla_\theta \ln\text{L}(\tilde{\theta})$ is equal to zero for those elements of θ that are unrestricted. That is, $\nabla_\psi \ln\text{L}(\tilde{\theta}) = 0$. Therefore

$$\begin{aligned} \text{LM} &= \nabla_\delta \ln\text{L}(\tilde{\theta})' \cdot [\mathcal{I}^{-1}(\tilde{\theta})]_{\delta\delta} \cdot \nabla_\delta \ln\text{L}(\tilde{\theta}) \\ &= \nabla_\delta \ln\text{L}(\tilde{\theta})' \cdot [\tilde{\mathcal{I}}_{\delta\delta} - \tilde{\mathcal{I}}_{\delta\psi}\tilde{\mathcal{I}}_{\psi\psi}^{-1}\tilde{\mathcal{I}}_{\psi\delta}]^{-1} \cdot \nabla_\delta \ln\text{L}(\tilde{\theta}) \end{aligned} \tag{1.18}$$

where $\tilde{\mathcal{I}}_{**}$ stands for the *,* block of $\mathcal{I}$, evaluated at $\theta = \tilde{\theta}$.

A straightforward calculation reveals that

$$\nabla_\delta \ln\text{L}(\tilde{\theta}) = \sum_{i=1}^{N} z_i \left(\frac{\tilde{\sigma_v}\tilde{\xi_i}}{\tilde{\lambda}} - \frac{\tilde{\epsilon_i}}{\tilde{\lambda}} - \frac{\tilde{\sigma}_v^2}{\tilde{\lambda}^2} - 1 \right) \tag{1.19}$$

where $\tilde{\xi_i} = (\phi(\tilde{\epsilon_i}/\tilde{\sigma_v} + \tilde{\sigma_v}/\tilde{\lambda}))(1 - \Phi(\tilde{\epsilon_i}/\tilde{\sigma_v} + \tilde{\sigma_v}/\tilde{\lambda}))^{-1}$ and ϕ is the pdf of the standard normal distribution. Note that

$$\nabla_\delta \ln\text{L}(\tilde{\theta}) = \frac{1}{\tilde{\lambda}} \sum_{i=1}^{N} z_i \left(\tilde{\sigma_v}\tilde{\xi_i} - \tilde{\epsilon_i} - \frac{\tilde{\sigma}_v^2}{\tilde{\lambda}} - \tilde{\lambda} \right) = \frac{1}{\tilde{\lambda}} \sum_{i=1}^{N} z_i \tilde{b}_i, \tag{1.20}$$

where $\tilde{b}_i = (\tilde{\sigma_v}\tilde{\xi_i} - \tilde{\epsilon_i} - \tilde{\sigma}_v^2/\tilde{\lambda} - \tilde{\lambda})$ is $(\text{E}(u_i|\epsilon_i) - \text{E}(u_i)) \equiv b_i$ evaluated at $\tilde{\theta}$. (This follows because $\text{E}(u_i|\epsilon_i) = \sigma_v(\xi_i - (\epsilon_i/\sigma_v + \sigma_v/\lambda))$ while $\text{E}(u_i) = \lambda$.) Note that apart from the scalar $1/\tilde{\lambda}$, $\nabla_\delta \ln\mathrm{L}(\tilde{\theta})$ equals the numerator of $\sqrt{N}\hat{\gamma} = (N^{-1}\sum_{i=1}^N z_i z_i')^{-1}$ $N^{-1/2}\sum_{i=1}^N z_i\tilde{b}_i$. So the LM test must be asymptotically equivalent to a properly constructed test based on the two-step estimator $\hat{\gamma}$. Some further algebraic details of this equivalence are given in the Appendix. Basically, the naive test that ignores the effects of estimation error in $\tilde{b}_i$ would correspond to omitting the terms $\tilde{\mathcal{I}}_{\delta\psi}\tilde{\mathcal{I}}_{\psi\psi}^{-1}\tilde{\mathcal{I}}_{\psi\delta}$ in (1.18). These terms correspond to the same correction as was created by the terms Gr_i in (1.9b) above.

This section's result (that the LM test is asymptotically equivalent to a properly constructed test based on a two-step procedure) holds for the case that u_i is exponential with a scaling factor of the form $\exp(z_i'\delta)$. So far as we can determine, it does not hold for the scaled half-normal case. If it does not, then in the half-normal we would expect the LM test to be better (in the sense of asymptotic local power) than the two-step test of the last section. An interesting question for further research is whether we can identify a class of distributions for which a result like the present one holds.

1.4 A Test Based on Nonlinear Least Squares

In this section we continue to assume that the stochastic frontier model (1.1) is correct. We further assume that the scaling property (1.2), with an exponential scaling function, holds, so $u_i = \exp(z_i'\delta)\cdot u_i^\circ$. However, now we do not make any specific distributional assumption about the u_i°. We simply assume that they are i.i.d. and independent of x_i, z_i and v_i.

Let $\mu \equiv \mathrm{E}(u_i^\circ) = \mathrm{E}(u_i^\circ|x_i, z_i)$. Then

$$\mathrm{E}(y_i|x_i, z_i) = x_i'\beta - \mu\cdot\exp(z_i'\delta), \tag{1.21}$$

or equivalently

$$y_i = x_i'\beta - \mu \cdot \exp(z_i'\delta) + w_i \tag{1.22}$$

where $\mathrm{E}(w_i|x_i, z_i) = 0$. This model can be estimated consistently by nonlinear least squares, as has been noted by Simar, Lovell, and Vanden Eeckaut (1994), Wang and Schmidt (2002) and others. This raises the question of whether we can test the hypothesis $\delta = 0$ based on the nonlinear least squares regression.

There is a non-trivial problem because the parameter μ is not identified (separately from the intercept in the regression) when $\delta = 0$. To see this clearly, we explicitly distinguish the intercept from the rest of x_i: $x_i' = (1, x_i^{*\prime})$, $\beta' = (\alpha, \beta^{*\prime})$ so that (1.22) becomes

$$y_i = \alpha + x_i^{*\prime}\beta^* - \mu \cdot \exp(z_i'\delta) + w_i. \tag{1.23}$$

Alternatively we can write this as

$$y_i = (\alpha - \mu) + x_i^{*\prime}\beta^* + \mu(1 - \exp(z_i'\delta)) + w_i. \tag{1.24}$$

From (1.24) it is clear that $(\alpha - \mu)$ is identified, but μ is identified only when $\delta \neq 0$.

In cases such as this, in which some parameters ("nuisance parameters") are not identified under the null hypothesis, standard tests like the Wald test or the likelihood ratio test are not asymptotically valid. A standard reference on this problem is Hansen (1996). A Wald test in this context would consist of estimating δ and then testing whether it is significantly different from zero, using a statistic of the form $\hat{\delta}'[\mathrm{Var}(\hat{\delta})]^{-1}\hat{\delta}$, where $\hat{\delta}$ is the NLLS estimate and $\mathrm{Var}(\hat{\delta})$ is the asymptotic variance matrix of $\hat{\delta}$. Such a test is not valid in this context because the usual $\mathrm{Var}(\hat{\delta})$ that would be valid when $\delta \neq 0$ is not valid when $\delta = 0$, because of the non-identification of μ.

It is interesting that for our problem (though not for general problems) an asymptotically valid test can be derived from the LM (or score) test principle. We follow the discussion in Wooldridge (2002)[pp. 363-369]. Let the NLLS criterion function be

$$Q_N(\theta) = \frac{1}{N}\sum_{i=1}^{N} q(\omega_i, \theta) = \frac{1}{N}\sum_{i=1}^{N}(y_i - x_i'\beta + \mu \exp(z_i'\delta))^2, \tag{1.25}$$

where θ represents β, μ and δ, and ω_i represents y_i, x_i and z_i. Then the LM or score test is based on the quantity $\nabla_\delta Q_N(\tilde{\theta})$, that is, on the derivative of $Q_N(\theta)$ with respect to δ, evaluated at the restricted estimates $\tilde{\theta}$. We might expect this approach to fail here because $\tilde{\mu}$ is not well defined. However, this turns out not to matter. Doing the appropriate calculation,

$$\nabla_\delta Q_N(\theta) = \frac{2}{N}\sum_{i=1}^{N}(y_i - x_i'\beta + \mu \exp(z_i'\delta))(\mu \exp(z_i'\delta) z_i) \tag{1.26}$$

and therefore (since $\tilde{\delta} = 0$):

$$\nabla_\delta Q_N(\tilde{\theta}) = \frac{2}{N}\sum_{i=1}^{N}(y_i - (\tilde{\alpha} - \tilde{\mu}) - x_i^{*\prime}\tilde{\beta}^*)(\tilde{\mu} z_i) = \frac{2}{N}\sum_{i=1}^{N}\tilde{w}_i(\tilde{\mu} z_i). \tag{1.27}$$

Here $\tilde{\beta} = ((\tilde{\alpha} - \tilde{\mu}), \tilde{\beta}^{*\prime})'$ is just the coefficient in a regression of y on X, and $\tilde{w}_i = y_i - x_i'\tilde{\beta}$. In matrix form, the sum in (1.27) is equal to $y'M_X(\tilde{\mu}Z)$, where $M_X = I - X(X'X)^{-1}X'$ is the projection orthogonal to X. Note that if we regressed y on $[X, \tilde{\mu}Z]$, the coefficients of $\tilde{\mu}Z$ would be $[(\tilde{\mu}Z)'M_X(\tilde{\mu}Z)]^{-1}(\tilde{\mu}Z)'M_X y$, so that the sum in (1.27) is equal to the random (numerator) portion of this coefficient. Therefore the LM statistic will be equivalent to an F-statistic for the significance of the coefficients (say, ζ) of $(\tilde{\mu} z_i)$ in the regression

$$y_i = x_i'\beta + (\tilde{\mu} z_i)'\zeta + error_i. \tag{1.28}$$

Now, the essential point is that this F-statistic is invariant to any non-zero value of $\tilde{\mu}$. That is, $\tilde{\mu}$ is just a scale factor for z_i, and changing $\tilde{\mu}$ is like changing the units of measurements of z_i. It does not affect the value of the F-statistic. (If we double $\tilde{\mu}$, this will cause $\hat{\zeta}$ to be divided by two, and Var($\hat{\zeta}$) to be divided by four, so the scale factor "two" cancels from the test statistic.) So we can just set $\tilde{\mu} = 1$, and calculate the LM statistic as the F-statistic for the significance of the coefficients of z_i in a regression of y_i on $[x_i, z_i]$.

This is an intuitively reasonable result because, under the null hypothesis being tested, $\mathrm{E}(y|X, Z)$ does not depend on Z.

An interesting and relevant fact is that the same test statistic would result if we replaced the exponential scaling function $\exp(z_i'\delta)$ by any scaling function $g(z_i'\delta)$, where g is monotonic and differentiable at zero. The same derivation as above leads us to a regression of y_i on x_i and $\mu g'(0) z_i$, or equivalently a regression of y_i on x_i and z_i. This is relevant because it suggests that the OLS-based test may have reasonable power against a variety of alternatives (different scaling functions), whereas the power properties of the MLE-based tests when the scaling function is misspecified are not at all clear.

We note that, if v_i and u_i° are i.i.d., the error in (1.27) is homoskedastic under the null hypothesis. Nevertheless it is possible to consider a heteroskedasticity-robust test. We simply have to use the heteroskedasticity-robust variance matrix of White (1980). See Wooldridge (2002)[pp. 55-58] for details.

Another thing to note is the following. The test above is the F-test for the significance of the coefficients of z in a regression of y on x and z. This is the same as the F-test for the significance of the coefficients of z in a regression of $\tilde{w}$ on x and z, where as in (1.27) above $\tilde{w} = y - x\tilde{\beta}$. It is essential that x be included in this regression, even though x is orthogonal to $\tilde{w}$. If we regressed $\tilde{w}$ on z only and did an F-test, this test would not be valid, even asymptotically.

1.5 Simulations: Experimental Design

We wish to perform simulations to investigate the size and power properties of the tests derived in the previous sections. The data generating process for our simulations will be as follows:

$$\begin{aligned} y_i &= \alpha + \beta x_i + v_i - \exp(\delta z_i) \cdot u_i^\circ \\ &= \alpha + \beta x_i - \lambda \exp(\delta z_i) + w_i, \quad i = 1, \cdots, N, \end{aligned} \tag{1.29}$$

where $w_i = v_i - \exp(\delta z_i)(u_i - \lambda)$. All random draws are independent over i. The explanatory variables x_i and z_i are both scalars, and $(x_i, z_i)'$ is standard bivariate normal with correlation ρ. The v_i are distributed as $N(0, \sigma_v^2)$ and the u_i° are distributed as exponential with parameter λ. The random variables $(x_i, z_i)'$, v_i and u_i° are mutually independent.

The set of parameters is therefore $\alpha, \beta, \delta, \sigma_v^2, \lambda, \rho$ and N. We chose a "base case" set of parameters as follows:

$$\alpha = 0, \beta = 0, \delta = 0, \sigma_v^2 = 1, \lambda = 1, \rho = 0.5, N = 200. \tag{1.30}$$

We will then change these parameter values, as described below, in our experiments.

We consider the following tests.

WALD. For the WALD test we estimate (1.29) by MLE and then test whether $\hat{\delta}$

is significantly different from zero. Specifically, the WALD statistic is given by

$$\text{WALD} = \left[\hat{\delta}^2 \left(\hat{\mathcal{I}}_{\delta\delta} - \hat{\mathcal{I}}_{\delta\varphi}\hat{\mathcal{I}}_{\varphi\varphi}^{-1}\hat{\mathcal{I}}_{\varphi\delta}\right)\right] \tag{1.31}$$

where the notation is the same as in Section 1.3. Two different versions of the WALD statistic are computed. WALD-OPG uses the OPG (outer product of the gradient) estimate of the information matrix, while WALD-HES uses the negative Hessian estimate of the information matrix.

LM. This is the LM statistic discussed in Section 1.3. The statistic is given by

$$\text{LM} = \left[\frac{1}{\bar{\lambda}}\sum_{i=1}^{N}\tilde{b}_i z_i \left(\tilde{\mathcal{I}}_{\delta\delta} - \tilde{\mathcal{I}}_{\delta\varphi}\tilde{\mathcal{I}}_{\varphi\varphi}^{-1}\tilde{\mathcal{I}}_{\varphi\delta}\right)^{-1} \frac{1}{\bar{\lambda}}\sum_{i=1}^{N}\tilde{b}_i z_i\right]. \tag{1.32}$$

Once again we have different versions, depending on how the information matrix is estimated. LM-OPG and LM-HES are analogous to WALD-OPG and WALD-HES.

GDV. This is the "generated dependent variable" test discussed in Section 1.2. More specifically,

$$\begin{aligned} \text{GDV} &= \sqrt{N}\hat{\gamma}[\widehat{\text{Var}(\sqrt{N}\hat{\gamma})}]^{-1}\sqrt{N}\hat{\gamma} \\ &= \left[\sum_{i=1}^{N}\tilde{b}_i z_i \left(\sum_{i=1}^{N}\left(\tilde{b}_i z_i + \tilde{G}\tilde{\mathcal{I}}^{-1}s_i(\tilde{\varphi})\right)^2\right)^{-1} \sum_{i=1}^{N}\tilde{b}_i z_i\right]. \end{aligned} \tag{1.33}$$

Here $\tilde{\mathcal{I}}$ is the negative Hessian form of the information matrix for the first-step MLE, as in (1.11) above. We also consider the test BADGDV, which is the invalid test based on regression (1.3) above and which ignores the estimation error in $\hat{\psi}$.

OLS. This is the set of tests discussed in Section 1.4. OLS refers to the standard F-test for significance of the coefficients of z_i in a regression of y_i on $(1, x_i, z_i)$. This reduces to a t-test in the present case since z_i is scalar. We use the critical values based on the standard normal distribution rather than the t-distribution but for our values of N this makes essentially no difference. OLS-H is the heteroskedasticity-robust version of the test. BADOLS is the invalid test based on the t-statistic for the significance of the coefficient of z_i when $\tilde{b}_i$ is regressed on z_i (without intercept or x_i in the regression), as discussed at the end of Section 1.4. BADOLS-H is the heteroskedasticity-robust version of BADOLS.

The number of replications in the experiment was 10,000, except for a few cases noted below.

The outputs of the experiments are as follows. For each of the parameter estimates, we calculated their mean, standard deviation, and MSE. For the MLE of the full model (needed for the WALD test calculations), the parameters estimated are α, β, δ, σ_v^2 and λ. For the MLE of the model subject to the restriction $\delta = 0$ (needed for the LM and GDV test calculations), the parameters estimated are α, β, σ_v^2 and λ. Note that, in the output tables, we report the mean, standard deviation, and MSE of the estimates of λ^2, not λ, for an easier comparison with the estimates of σ_v^2. For the NLLS estimates under the restriction that $\delta = 0$ (which is just OLS of y_i on x_i, and is needed for the OLS test calculations), the parameters estimated are $\eta = \alpha - \lambda$, β and $\sigma_w^2 = \sigma_v^2 + \lambda^2$.

We also calculated the mean, standard deviation and MSE of the technical efficiency estimates for the MLE and the restricted MLE. The technical efficiency of firm i is $TE_i = \exp(-u_i)$ and the technical efficiency estimate (Battese and Coelli (1988)) is

$$\begin{aligned} TE_i &= \mathrm{E}(\exp(-u_i)|\epsilon_i) \\ &= \frac{\Phi\left(-\sigma_v - \epsilon_i/\sigma_v - \sigma_v/(\lambda \exp(\delta z_i))\right)}{\Phi\left(-\epsilon_i/\sigma_v - \sigma_v/(\lambda \exp(\delta z_i))\right)} \exp\left(\frac{\sigma_v^2}{2} + \epsilon_i + \frac{\sigma_v^2}{\lambda \exp(\delta z_i)}\right). \end{aligned} \quad (1.34)$$

Here $\epsilon_i = v_i - u_i = y_i - \alpha - \beta x_i$ and $\widehat{TE}_i$ is the expression (1.34) evaluated at the MLE estimates. By the law of iterated expectations, $\mathrm{E}(TE_i) = \mathrm{E}\exp(-u_i)$. However, for the calculation of MSE we average the squared deviations of $\widehat{TE}_i$ for $TE_i = \exp(-u_i)$, not from $\mathrm{E}\exp(-u_i)$. The mean, standard deviation and MSE for $\widehat{TE}_i$ are calculated by averaging across observations ($i = 1, \cdots, N$) as well as across replications. We also report the correlation of $\widehat{TE}_i$ and TE_i. The is the average across replications of the correlation coefficient for a given replication.

For the tests, we calculated the proportion of rejections, which is interpreted as size (if $\delta = 0$) or power (if $\delta \neq 0$). The size (or power) is calculated in four ways. Size1 uses all 10,000 replications. Size2 drops replications in which there was a numerical failure in the calculation of the WALD or LM statistics, due to outliers in the estimates. Outliers are defined as $|\hat{\delta}| \geq 16$, $\hat{\sigma}_v \leq 10^{-7}$ or $\hat{\sigma}_v \geq 37$, and $\hat{\lambda} \leq 10^{-7}$ or $\hat{\lambda} \geq 37$. Size3 drops observations with negative LM statistics. These may occur when the maximization algorithm fails to reach the global maximum. Finally, Size4 drops any replication dropped by either the Size2 or the Size3 calculation.

We also report the mean and standard deviations of the test statistics. This calculation was done over the same set of replications used to calculated Size4.

Many of the replications discarded in Size2 and Size4 are ones in which the variance parameters (σ_v^2 and λ^2) and δ are poorly estimated. Very small values of λ^2 tended to go with very large values of δ, as the likelihood calculation seemed to try to accommodate the presence of the one-sided error $\exp(\delta z_i) \cdot u_i^\circ$ by balancing a small variance of u_i° with a large value of $\exp(\delta z_i)$. In these cases the variance of $\hat{\delta}$ is also hard to calculate, and it is just not clear whether or not they constitute evidence against the null that $\delta = 0$. Dropping these cases primarily reduces the number of rejections for the WALD tests. However, except for a few parameter values (e.g. very large σ_v^2), not enough replications were dropped to make much difference.

1.6 Simulation Results: Size

In this section, we investigate the size of the tests. Therefore all of the cases considered have $\delta = 0$ so that the null hypothesis is true. All of the tests except BADGDV, BADOLS and BADOLS-H (which we will call the BAD tests for short) are valid asymptotically but we are interested in how substantial their size distortions may be in finite samples.

1.6.1 Base case

We first consider the base case: $\alpha = \beta = \delta = 0$, $\sigma_v^2 = \lambda = 1$, $\rho = 0.5$, and $N = 200$. The results are given in Table 1.1.

The results for the point estimates are fairly unremarkable. There is little or no evidence of finite-sample bias. The restricted MLE's are better than the unrestricted MLE's, in terms of standard deviation and MSE, but the differences are quite small.

The sizes of the various tests differ fairly substantially from each other. All of the BAD tests are indeed bad, in the sense of size substantially less than 5%. However, some of the asymptotically valid tests also have sizes that are substantially different from 5%. The WALD tests are substantially undersized. Conversely, the LM-OPG test rejects too often. The LM-HES, GDV and OLS tests have size fairly close to 5%, and the OLS-H test is only slightly worse than those three.

1.6.2 Effects of changing α or β

Changes in α or β would not be expected to change the results, and this is true in the following sense. We did one simulation with the same parameters as in the base case except that $\alpha = 1$, and another simulation with the same parameters except that $\beta = 1$. These changes did not change the size of any of the tests, and the only effect on the point estimates was to change the mean value of $\hat{\alpha}$ or $\hat{\beta}$ by one.

1.6.3 Effects of changing N

Next we considered parameter values that were the same as in the base case, except that we changed N to $N = 500$ (Table 1.2) and $N = 1000$ (Table 1.3).

When we increase N, we reduce the standard deviation and MSE of the various parameter estimates, as expected. However, it is notable that we do not increase the precision of the technical efficiency estimates except perhaps trivially. To understand why, recall that the technical efficiency estimate is the expectation of $\exp(-u)$ conditional on $(v - u)$, evaluated at the estimated values of the parameters. The variance of this estimate depends on (i) "intrinsic variability," by which we mean the variance of $\exp(-u)$ conditional on $(v - u)$, which does not depend on N, and (ii) "sampling error," by which we mean the variance of the parameter estimates, which does depend on N. Apparently even for $N = 200$ sampling error is quite small relative to intrinsic variability.

As would be expected, increasing N does not reduce the size distortions of the BAD tests, but it does improve the asymptotically valid tests. For $N = 500$ we have the same pattern of size distortions as we observed for $N = 200$, but they are much smaller. Also the various types of numerical failures that distinguish Size1 from Size2, Size3 and Size4 have largely disappeared. For $N = 1000$ all of the asymptotically valid tests have reasonably accurate size; the worst is LM-OPG with size of 5.77%. The good news in this statement is that the tests behave as they should asymptotically. The bad news is that $N = 1000$ would be a very large sample size indeed for the type of efficiency measurement exercise that is considered here.

1.6.4 Effects of changing ρ

Now we consider changes in ρ, the correlation between x and z. The question of interest is whether strong correlation between x and z creates difficulties (akin to multicollinearity) in estimation and whether this affects the tests. In the base case we had $\rho = 0.5$, and now we keep the rest of the base case parameters but consider $\rho = -0.5$ (Table 1.4), $\rho = 0$ (Table 1.5) and $\rho = 0.9$ (Table 1.6). We also considered $\rho = 0.25$ and $\rho = 0.75$, and those results are in a supplementary set of tables.

In terms of the point estimates based on MLE, the value of ρ makes little difference. When $\rho = 0.9$ the standard deviation and MSE of β and δ do increase, but not by very much. The value of ρ does not matter very much for any of the asymptotically valid tests, and in fact the results for the OLS and OLS-H tests do not change at all. For the BAD tests, it makes more difference, as asymptotic theory would suggest. For $\rho = 0$ the BAD tests are asymptotically valid, and they have approximately correct size, while for $\rho = 0.9$ the BAD tests have size of nearly zero.

1.6.5 Effects of changing λ

Next we consider a change in λ, the parameter of exponential distribution of the one-sided error u°. In Table 1.7 we report the results for $\lambda = 3$, whereas the base case had $\lambda = 1$.

Since the overall error in the model is $v - u$, where v is normal noise, increasing λ effectively decreases the relative importance of the noise, and should make inference about u or about the effect of z on u more reliable. Comparing Table 1.7 to Table 1.1, we see that this is true. With the larger value of λ, the sizes of the asymptotically valid tests (other than GDV) all become closer to 5%. The effects of this change on the point estimates were less clear, in part because when $\lambda = 3$ there were more outliers.

1.6.6 Effects of changing σ_v^2

Now we change σ_v^2 to 9, as opposed to its base case value of 1, holding the other parameters the same. The results are in Table 1.8. This is a pure increase in statistical noise and it should make all of the estimates and tests worse. Comparing Table 1.8

to Table 1.1, that turns out to be true for all of the estimates, and for most of the tests. Among the asymptotically valid tests, the WALD tests and the GDV test are very seriously affected. They give very few rejections. There is relatively little effect of this change on the size of the LM-OPG and LM-HES tests or the OLS and OLS-H tests, however.

It is also notable that the number of replications dropped in the size calculations is very large with the higher value of σ_v^2. The data are close enough to normal that the maximization process was difficult. As a curiosity we ran the Schmidt and Lin (1984) test of the hypothesis of no one-sided error, and we could reject this hypothesis (at the 5% level) only 1,086 times out of 10,000.

1.7 Simulation Results: Power

In this section we investigate the power of the various tests. We therefore set δ to some non-zero value. An immediate problem that arises is that it is not meaningful to compare the power of tests if their sizes are very different. One possibility is to consider size-adjusted power, but this has the disadvantage that then we are no longer investigating the power of a procedure that is feasible outside the simulation setting. An alternative possibility, which we follow, is to investigate power using a sample size sufficiently large that size distortions are not a serious problem. Therefore for all of our simulations in this section we will set $N = 1000$. Our "base case" is therefore the set of parameters for the simulations reported in Table 1.3, and we now change δ from 0 to 0.05, 0.10 and 0.15, where these values were chosen to yield power that moved through a reasonable part of the range between zero and one. These results are given in Tables 1.9, 1.10 and 1.11.

Changing δ has very little effect on any of the point estimates, other than the mean of $\hat{\delta}$, and we will not discuss the estimation results further.

Power increases as δ increases, for obvious reasons. If we compare the WALD, LM and GDV tests their powers are quite similar. Fine distinctions are hard to make because even with $N = 1000$ their sizes were slightly different in Table 1.3. These tests are all asymptotically valid, and they all have the same asymptotic local power, so it is not surprising that their powers should be similar for $N = 1000$. A more interesting comparison is between their power and the power of the OLS-based tests (OLS and OLS-H). The OLS-based tests do not make use of the assumption that the u_i° are exponential, and the failure to exploit this fact ought to make them less powerful than the WALD, LM and GDV tests. This turns out to be true, with the difference in power being non-trivial but not huge. For example, for $\delta = 0.1$, compare 0.51 for OLS to 0.64 for LM-HES.

We also did some additional simulations with $\rho = 0.9$, so that the variables x and z are more highly correlated than in the cases just considered (which had $\rho = 0.5$). Table 1.12 gives the results for $\delta = 0.1$ and $\rho = 0.9$, and the results for $\delta = 0.05$ and 0.15 are in our supplemental set of tables. Comparing Table 1.12 to Table 1.10, we can see that the higher value of ρ results in substantially lower powers for all of the tests. Among the asymptotically valid tests, the loss in power is much larger for the OLS-based tests than for the WALD, LM or GDV tests. These differences are certainly non-trivial. For example, the power of the LM-HES test changes from 0.64 to 0.48 when ρ changes from 0.5 to 0.9, while the power of the OLS test changes from 0.51 to 0.17.

The low power of the OLS-based tests occurs because of multicollinearity in the OLS regression when x and z are highly correlated. The coefficient of z is poorly estimated and it is hard to reject the hypothesis that it is zero. The MLE-based tests do a better job of exploiting the nonlinearity of the relationship between y, x and z and suffer less when x and z are highly correlated. How much this matters, in an empirical setting, obviously will depend on how different the variables in z are from those in x.

Finally, we did some simulations in which the tests are exactly as above, and are therefore based on the assumption that the true scaling function is $\exp(\delta z_i)$, when in fact this is not the true scaling function. For these simulations, we have $u_i = \phi(\delta z_i)(1-\Phi(\delta z_i))^{-1}u_i^\circ$, where ϕ is the standard normal density, Φ is the standard normal cdf, and u_i° is exponential with parameter $\lambda = 1$. So, in the data generating process, the scaling function is the inverse Mill's ratio, $\phi(\delta z_i)(1 - \Phi(\delta z_i))^{-1}$.

Under the null, $\delta = 0$ and u_i is exponential with parameter $\sqrt{2/\pi}$. So our tests based on the exponential scaling function correctly encompass the null, and the only question is power. For the MLE-based tests, their power properties when the scaling function is misspecified are certainly not clear. For the OLS-based tests, however, we saw that the same statistic resulted from the score test principle for any monotonic differentiable scaling function $g(\delta z_i)$. As a result, we might expect our OLS test to have better power properties relative to the MLE-based tests when the MLE-based tests are based on the wrong scaling function.

Table 1.13 gives the simulation results with $\delta = 0.1$. These simulations have $N = 1000$, and are based on 2000 replications. The surprising aspect of these results is the good performance of the MLE-based methods. The parameter estimates look quite reasonable, despite the misspecification of the model. Similarly the MLE-based tests are more powerful than the OLS-based tests, despite the arguments of the previous paragraph. These optimistic results deserve attention in future research.

1.8 Simulation Results: Robustness

In this section we investigate the effects of misspecification of the distribution of the one-sided error term. Specifically, we will consider the properties of the tests based on the MLE that assumes an exponential error, when in fact the distribution of u° is either truncated normal or gamma.

We note at the outset that this issue does not arise with our OLS-based tests. These do not rely on any distributional assumptions on the errors, and they are asymptotically valid for any error distribution with finite variance (so that the central limit theorem applies).

The MLE-based tests, on the other hand, will generally be invalid when the error distribution is misspecified. Fundamentally this is simply because the likelihood is then misspecified. To be more specific, consider the LM test or the GDV test based on the normal-exponential model, as discussed in Sections 1.2 and 1.3 above. These fundamentally depend on the quantity $\sum_{i=1}^{N} z_i \hat{b}_i$ where $\hat{b}_i$ is an estimate of $b_i = \mathrm{E}(u_i|\epsilon_i) - \mathrm{E}(u_i)$, with $\epsilon_i = v_i - u_i$. The precise form of b_i depends on the assumption that v_i is normal and u_i° is exponential. If in fact u_i° is not exponential, then $\mathrm{E}(b_i) \neq 0$ and we cannot expect the test to be valid. A secondary but still relevant issue is that the asymptotic variance of $\sum_{i=1}^{N} z_i \hat{b}_i$, which also figures into the test statistic, also depends on the distributional assumption for u_i° being correct. See Section 1.2 above.

We emphasize that the lack of robustness of the MLE-based tests to distributional misspecification is not just a finite-sample issue. This problem persists even asymptotically.

The lack of validity of the MLE-based tests should show up in simulations as incorrect size when the null hypothesis is true. The question then is how serious this problem is. Greene (1990) has argued that the rankings of estimated inefficiencies are often not sensitive to distributional assumptions on the one-sided error. Also, the exponential distribution shares same features with other one-sided distributions. The half-normal distribution, like the exponential, has a mode at zero. The gamma(g_1, g_2) distribution with $g_1 = 1$ is exponential, and for $0 < g_1 < 1$ it has a shape similar to the exponential.

In the simulations of this section we have $\alpha = \beta = \delta = 0$, $\sigma_v^2 = 1$ and $N = 1000$.

The number of replication is 2000.

1.8.1 Normal-truncated normal

Here the distribution of u_i° is $N(\mu, \sigma^2)^+$, that is, truncated normal. Table 1.14 gives our results for the case that $\mu = 0$ and $\sigma^2 = \pi/2$. This is the half-normal distribution with mean equal to one. This choice makes the distribution somewhat comparable to the exponential distribution with parameter one, as in Table 1.3 above. However, the truncated normal with $\mu = 0$ and $\sigma^2 = \pi/2$ has variance equal to 0.57. (A truncated normal, unlike an exponential does not have its mean equal to its standard deviation.) We also considered three other cases: (i) $\mu = 0$, $\sigma^2 = \pi/(\pi - 2)$, for which the variance equals one but the mean equals 1.32; (ii) $\mu = 0$, $\sigma^2 = 1$; (iii) $\mu = 1$, $\sigma^2 = 1$. The results for these three cases are in our supplemental set of tables.

In Table 1.14 we see that the OLS-based tests appear to have proper size, while the MLE-based tests exhibit significant size distortions. For MLE there are also considerable biases in the parameter estimates. The WALD and GDV tests are undersized, while the LM-OPG test rejects too often. This same pattern occurs for all four cases that we considered but the extent of the size distortions varied considerably over choices of μ and σ_u^2.

Comparing Table 1.14 to Table 1.3, we also see that there are many more replications dropped when the distribution is misspecified. Obviously the data do not always fit the likelihood well and numerical problems occur.

1.8.2 Normal-gamma

Now the distribution of u_i° is gamma(g_1, g_2). The results in Table 1.15 and Table 1.16 are similar to those in Table 1.14 for the normal-truncated normal case. The OLS-based tests have more or less proper size, while the MLE-based tests do not. The LM-OPG test rejects too often, and this is true across all of our (g_1, g_2) values. The

WALD, LM-HES and GDV tests also show significant size distortions, and sometimes reject too seldom and sometimes too often, depending the value of (g_1, g_2). The MLE parameter estimates show clear biases. However, unlike the truncated normal case, not many replications were dropped here. The exponential model fits the data better in the normal-gamma case than in the normal-truncated normal case. Interestingly, that does not mean that it leads to more robust inference in the former case than in the latter.

1.9 Concluding Remarks

In this chapter we have considered tests of the hypothesis that observable firm characteristics do not affect technical efficiency. We do this in the context of a specific model in which the one-sided errors are exponential. Under the null they are i.i.d. while under the alternative they are scaled by a function $\exp(z_i'\delta)$, where z_i are the firm characteristics whose influence we are testing.

In this context we can estimate the model by MLE and test whether $\delta = 0$, which is the WALD test. We can also use an LM test. We show that a simple two-step test is not valid. (Here step one is to estimate technical efficiency for each firm. Step two is to regress these estimates on z_i and test whether the coefficients are zero.) This test can be made valid by correcting the asymptotic variance matrix for the second-step estimates. This correction is distribution-specific. When technical efficiency is exponential, we show that the corrected two-step test is asymptotically equivalent to the LM test.

We can also derive a valid test from the score test principle applied to the nonlinear least squares problem. This takes the form of an F-test of the significance of the coefficients of z_i in an OLS regression of output on z_i and the inputs. This test does not require a distributional assumption and it would be the same for any scaling function of the form $g(z_i'\delta)$, where g is monotonic and differentiable at zero. The OLS-based test therefore has good robustness properties, but it may be expected to have lower power than the MLE-based tests when the model for MLE is correctly specified.

We perform a number of simulations to investigate the size and power properties of the tests we have suggested. The OLS-based tests do turn out to have good robustness properties and the MLE-based tests do turn out to be more powerful when the model is correctly specified. The loss in power for the OLS-based tests is especially large when the inputs and the firm characteristics z_i are highly correlated. The MLE-based tests show significant differences among themselves when the sample size is not very large. The WALD tests reject too seldom and the LM-OPG test rejects too often. The LM test using the Hessian (LM-HES) and the corrected two-step test (GDV) are generally most reliable. The MLE-based tests perform reasonably well if the scaling function is misspecified but they do not have proper size if the distribution of inefficiency is misspecified.

These results provide some guidance for empirical work. If the researcher's interest is not in the inefficiencies themselves, but just in testing whether they depend on firm characteristics (like firm size, state versus private ownership, etc.) then the OLS-based tests would be natural, unless these firm characteristics are very strongly correlated with the inputs. However, if the researcher is going to estimate firm-level efficiencies in any case, then a distributional assumption will ultimately be needed, and MLE-based tests may as well be used. Among these tests the LM test using the Hessian or the corrected two-step test would be preferred.

1.10 Output Tables

Table 1.1: (BASE CASE) $\alpha = \beta = \delta = 0$, $\sigma_v^2 = \lambda = 1$, $\rho = 0.5$, $N = 200$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.0001	0.1157	0.0134	
	$\hat{\alpha}$	-0.0342	0.1874	0.0363	
	$\hat{\beta}$	0.0000	0.1004	0.0101	
	$\hat{\sigma}_v^2$	1.0072	0.2279	0.0520	
	$\hat{\lambda}^2$	0.9594	0.3524	0.1258	
	$\widehat{TE}$	0.5148	0.1769	0.0555	0.6131
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	-0.0214	0.1817	0.0335	
	$\tilde{\beta}$	0.0001	0.0932	0.0087	
	$\tilde{\sigma}_v^2$	0.9985	0.2240	0.0502	
	$\tilde{\lambda}^2$	0.9919	0.3461	0.1199	
	$\widetilde{TE}$	0.5099	0.1767	0.0547	0.6194
Restricted NLLS	$\tilde{\eta}$	-1.0004	0.0994	0.0099	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	0.0007	0.1004	0.0101	
$\eta = -1, \beta = 0, \sigma_w^2 = 2$)	$\tilde{\sigma}_w^2$	2.0019	0.2691	0.0724	

STATISTICS	Size1	Size2	Size3	Size4	Mean	s.d.
WALD-OPG	0.0211	0.0213	0.0214	0.0215	-0.0027	0.8669
WALD-HES	0.0298	0.0300	0.0296	0.0297	-0.0045	0.9345
LM-OPG	0.0788	0.0783	0.0766	0.0763	1.2115	1.7972
LM-HES	0.0523	0.0518	0.0515	0.0509	1.0561	3.3408*
GDV	0.0466	0.0467	0.0471	0.0471	1.0067	1.3462
BADGDV	0.0363	0.0355	0.0350	0.0344	0.8564	1.2163
OLS	0.0495	0.0490	0.0481	0.0475	-0.0011	0.9973
OLS-H	0.0575	0.0572	0.0561	0.0558	-0.0003	1.0216
BADOLS	0.0237	0.0235	0.0233	0.0230	-0.0010	0.8634
BADOLS-H	0.0266	0.0265	0.0262	0.0260	-0.0014	0.8806
Rep. dropped	0	73	121	171		

* due to outliers

Table 1.2: (Change of N) $N = 500$, $\alpha = \beta = \delta = 0$, $\sigma_v^2 = \lambda = 1$, $\rho = 0.5$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.0000	0.0649	0.0042	
	$\hat{\alpha}$	-0.0125	0.1108	0.0124	
	$\hat{\beta}$	-0.0005	0.0627	0.0039	
	$\hat{\sigma}_v^2$	1.0024	0.1405	0.0197	
	$\hat{\lambda}^2$	0.9849	0.2171	0.0473	
	$\widehat{TE}$	0.5050	0.1790	0.0522	0.6215
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	-0.0074	0.1089	0.0119	
	$\tilde{\beta}$	-0.0005	0.0583	0.0034	
	$\tilde{\sigma}_v^2$	0.9986	0.1390	0.0193	
	$\tilde{\lambda}^2$	0.9985	0.2146	0.0461	
	$\widetilde{TE}$	0.5032	0.1792	0.0520	0.6227
Restricted NLLS	$\tilde{\eta}$	-1.0007	0.0632	0.0040	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	-0.0005	0.0629	0.0040	
$\eta = -1, \beta = 0, \sigma_w^2 = 2$)	$\tilde{\sigma}_w^2$	2.0023	0.1681	0.0283	

STATISTICS	Size1	Size2	Size3	Size4	Mean	s.d.
WALD-OPG	0.0361	0.0361	0.0361	0.0361	-0.0005	0.9386
WALD-HES	0.0434	0.0434	0.0434	0.0434	0.0003	0.9741
LM-OPG	0.0611	0.0611	0.0611	0.0612	1.0967	1.5831
LM-HES	0.0503	0.0503	0.0502	0.0502	0.9876	1.4005
GDV	0.0502	0.0502	0.0502	0.0502	1.0127	1.3969
BADGDV	0.0355	0.0355	0.0355	0.0355	0.8613	1.2250
OLS	0.0513	0.0513	0.0513	0.0513	0.0005	0.9941
OLS-H	0.0538	0.0538	0.0538	0.0538	0.0008	1.0043
BADOLS	0.0245	0.0245	0.0245	0.0245	0.0003	0.8604
BADOLS-H	0.0254	0.0254	0.0254	0.0254	0.0005	0.8681
Rep.dropped	0	2	8	9		

Table 1.3: (Change of N) $N = 1000$, $\alpha = \beta = \delta = 0$, $\sigma_v^2 = \lambda = 1$, $\rho = 0.5$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.0006	0.0450	0.0020	
	$\hat{\alpha}$	-0.0070	0.0757	0.0058	
	$\hat{\beta}$	0.0007	0.0447	0.0020	
	$\hat{\sigma}_v^2$	1.0023	0.0965	0.0093	
	$\hat{\lambda}^2$	0.9907	0.1509	0.0228	
	$\widehat{TE}$	0.5026	0.1796	0.0514	0.6230
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	-0.0046	0.0750	0.0057	
	$\tilde{\beta}$	0.0005	0.0419	0.0018	
	$\tilde{\sigma}_v^2$	1.0006	0.0961	0.0092	
	$\tilde{\lambda}^2$	0.9971	0.1499	0.0225	
	$\widetilde{TE}$	0.5018	0.1796	0.0514	0.6235
Restricted NLLS	$\tilde{\eta}$	-1.0004	0.0448	0.0020	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	0.0005	0.0449	0.0020	
$\eta = -1, \beta = 0, \sigma_w^2 = 2$)	$\tilde{\sigma}_w^2$	2.0001	0.1196	0.0143	

STATISTICS	Size1	Size2	Size3	Size4	Mean	s.d.
WALD-OPG	0.0448	0.0448	0.0448	0.0448	0.0141	0.9748
WALD-HES	0.0485	0.0485	0.0485	0.0485	0.0146	0.9916
LM-OPG	0.0577	0.0577	0.0577	0.0577	1.0545	1.5021
LM-HES	0.0513	0.0513	0.0513	0.0513	1.0006	1.4092
GDV	0.0522	0.0522	0.0522	0.0522	1.0136	1.4101
BADGDV	0.0377	0.0377	0.0377	0.0377	0.8789	1.2469
OLS	0.0490	0.0490	0.0490	0.0490	-0.0143	0.9970
OLS-H	0.0488	0.0488	0.0488	0.0488	-0.0145	1.0008
BADOLS	0.0243	0.0243	0.0243	0.0243	-0.0124	0.8636
BADOLS-H	0.0253	0.0253	0.0253	0.0253	-0.0123	0.8661
Rep. dropped	0	0	1	1		

Table 1.4: (Change of ρ) $\rho = -0.5$, $\alpha = \beta = \delta = 0$, $\sigma_v^2 = \lambda = 1$, $N = 200$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.0012	0.1157	0.0134	
	$\hat{\alpha}$	-0.0338	0.1859	0.0357	
	$\hat{\beta}$	-0.0003	0.0994	0.0099	
	$\hat{\sigma}_v^2$	1.0071	0.2273	0.0517	
	$\hat{\lambda}^2$	0.9596	0.3518	0.1254	
	$\widehat{TE}$	0.5147	0.1769	0.0554	0.6133
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	-0.0210	0.1808	0.0331	
	$\tilde{\beta}$	0.0001	0.0933	0.0087	
	$\tilde{\sigma}_v^2$	0.9984	0.2244	0.0504	
	$\tilde{\lambda}^2$	0.9922	0.3458	0.1196	
	$\widetilde{TE}$	0.5098	0.1767	0.0546	0.6194
NLLS under the null	$\tilde{\eta}$	-1.0002	0.0995	0.0099	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	0.0008	0.1004	0.0101	
$\eta = -1, \beta = 0, \sigma_w^2 = 2$)	$\tilde{\sigma}_w^2$	2.0023	0.2689	0.0723	

STATISTICS	Size1	Size2	Size3	Size4	Mean	s.d.
WALD-OPG	0.0175	0.0176	0.0177	0.0178	0.0092	0.8595
WALD-HES	0.0302	0.0304	0.0302	0.0304	0.0093	0.9313
LM-OPG	0.0776	0.0773	0.0758	0.0756	1.2179	1.7921
LM-HES	0.0515	0.0503	0.0515	0.0502	1.0036	1.4506
GDV	0.0467	0.0470	0.0472	0.0474	1.0081	1.3418
BADGDV	0.0367	0.0360	0.0355	0.0349	0.8546	1.2196
OLS	0.0495	0.0490	0.0481	0.0477	-0.0007	0.9973
OLS-H	0.0575	0.0571	0.0560	0.0557	0.0003	1.0217
BADOLS	0.0239	0.0236	0.0234	0.0231	-0.0002	0.8629
BADOLS-H	0.0283	0.0282	0.0276	0.0275	0.0013	0.8808
Rep. dropped	0	82	133	184		

Table 1.5: (Change of ρ) $\rho = 0$, $\alpha = \beta = \delta = 0$, $\sigma_v^2 = \lambda = 1$, $N = 200$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.0007	0.1062	0.0113	
	$\hat{\alpha}$	-0.0339	0.1858	0.0357	
	$\hat{\beta}$	0.0001	0.0934	0.0087	
	$\hat{\sigma}_v^2$	1.0072	0.2271	0.0516	
	$\hat{\lambda}^2$	0.9605	0.3509	0.1247	
	$\widehat{TE}$	0.5145	0.1767	0.0556	0.6150
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	-0.0214	0.1812	0.0333	
	$\tilde{\beta}$	0.0001	0.0933	0.0087	
	$\tilde{\sigma}_v^2$	0.9987	0.2243	0.0503	
	$\tilde{\lambda}^2$	0.9916	0.3459	0.1197	
	$\widetilde{TE}$	0.5099	0.1767	0.0546	0.6194
Restricted NLLS	$\tilde{\eta}$	-1.0003	0.0994	0.0099	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	0.0008	0.1004	0.0101	
$\eta = -1, \beta = 0, \sigma_w^2 = 2$)	$\tilde{\sigma}_w^2$	2.0020	0.2690	0.0723	

STATISTICS	Size1	Size2	Size3	Size4	Mean	s.d.
WALD-OPG	0.0207	0.0208	0.0209	0.0210	0.0034	0.8710
WALD-HES	0.0293	0.0295	0.0293	0.0295	0.0019	0.9376
LM-OPG	0.0770	0.0763	0.0754	0.0748	1.2030	1.7534
LM-HES	0.0515	0.0502	0.0514	0.0500	1.0164	1.6641
GDV	0.0452	0.0453	0.0455	0.0455	1.0128	1.3481
BADGDV	0.0517	0.0507	0.0502	0.0492	0.9824	1.3752
OLS	0.0495	0.0485	0.0483	0.0474	-0.0013	0.9963
OLS-H	0.0575	0.0568	0.0563	0.0556	-0.0002	1.0211
BADOLS	0.0499	0.0489	0.0487	0.0478	-0.0013	0.9963
BADOLS-H	0.0557	0.0550	0.0545	0.0539	-0.0002	1.0159
Rep. dropped	0	70	118	162		

Table 1.6: (Change of ρ) $\rho = 0.9$, $\alpha = \beta = \delta = 0$, $\sigma_v^2 = \lambda = 1$, $N = 200$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	-0.0012	0.1414	0.0200	
	$\hat{\alpha}$	-0.0321	0.1876	0.0362	
	$\hat{\beta}$	-0.0008	0.1217	0.0148	
	$\hat{\sigma}_v^2$	1.0051	0.2280	0.0520	
	$\hat{\lambda}^2$	0.9593	0.3544	0.1272	
	$\widehat{TE}$	0.5149	0.1781	0.0560	0.6079
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	-0.0200	0.1812	0.0332	
	$\tilde{\beta}$	0.0000	0.0934	0.0087	
	$\tilde{\sigma}_v^2$	0.9970	0.2234	0.0499	
	$\tilde{\lambda}^2$	0.9946	0.3459	0.1196	
	$\widetilde{TE}$	0.5094	0.1769	0.0546	0.6195
Restricted NLLS	$\tilde{\eta}$	-1.0004	0.0995	0.0099	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	0.0007	0.1005	0.0101	
$\eta = -1, \beta = 0, \sigma_w^2 = 2$)	$\tilde{\sigma}_w^2$	2.0030	0.2695	0.0726	

STATISTICS	Size1	Size2	Size3	Size4	Mean	s.d.
WALD-OPG	0.0188	0.0191	0.0192	0.0195	-0.0132	0.8417
WALD-HES	0.0314	0.0319	0.0316	0.0320	-0.0149	0.9290
LM-OPG	0.0820	0.0815	0.0807	0.0803	1.2490	1.8683
LM-HES	0.0561	0.0558	0.0558	0.0552	1.0876	6.7456*
GDV	0.0405	0.0408	0.0414	0.0415	0.9791	1.2808
BADGDV	0.0108	0.0106	0.0109	0.0108	0.5739	0.8467
OLS	0.0495	0.0491	0.0486	0.0484	0.0027	0.9991
OLS-H	0.0575	0.0571	0.0565	0.0562	0.0037	1.0230
BADOLS	0.0000	0.0000	0.0000	0.0000	0.0015	0.4345
BADOLS-H	0.0000	0.0000	0.0000	0.0000	0.0008	0.4441
Rep. dropped	0	171	217	346		

* due to outliers

Table 1.7: (Change of λ) $\lambda = 3$, $\alpha = \beta = \delta = 0$, $\sigma_v^2 = 1$, $\rho = 0.5$, $N = 200$ [E(exp($-u$)) = 0.1095]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.0000	0.0768	0.0059	
	$\hat{\alpha}$	-0.0183	0.2100	0.0445	
	$\hat{\beta}$	-0.0010	0.1462	0.0214	
	$\hat{\sigma}_v^2$	0.9873	0.3385	0.1147	
	$\hat{\lambda}^2$	8.9144	1.7316	3.0054	
	$\widehat{TE}$	0.2531	0.2234	0.0330	0.7736
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	-0.0122	0.2502	0.0627	
	$\tilde{\beta}$	-0.0010	0.1411	0.0199	
	$\tilde{\sigma}_v^2$	0.9884	0.9388	0.8814	
	$\tilde{\lambda}^2$	9.0074	1.7342	3.0072	
	$\widetilde{TE}$	0.2521	0.2234	0.0337	0.7740
Restricted NLLS	$\tilde{\eta}$	-2.9982	0.2227	0.0496	
on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	0.0003	0.2224	0.0495	
$\eta = -3, \beta = 0, \sigma_w^2 = 10$	$\tilde{\sigma}_w^2$	9.9914	1.8539	3.4365	

STATISTICS	Size1	Size2	Size3	Size4	Mean	s.d.
WALD-OPG	0.0397	0.0402	0.0397	0.0402	0.0001	0.9366
WALD-HES	0.0422	0.0427	0.0422	0.0427	-0.0002	0.9757
LM-OPG	0.0629	0.0580	0.0629	0.0580	1.0675	1.4782
LM-HES	0.0493	0.0440	0.0493	0.0441	0.9573	1.3053
GDV	0.0545	0.0499	0.0545	0.0499	1.0083	1.3753
BADGDV	0.0443	0.0386	0.0443	0.0386	0.8949	1.2349
OLS	0.0503	0.0440	0.0503	0.0441	-0.0041	0.9753
OLS-H	0.0565	0.0513	0.0565	0.0513	-0.0027	1.0011
BADOLS	0.0226	0.0183	0.0226	0.0183	-0.0034	0.8449
BADOLS-H	0.0259	0.0228	0.0259	0.0228	-0.0030	0.8664
Rep. dropped	0	124	1	125		

Table 1.8: (Change of σ_v^2) $\sigma_v^2 = 9$, $\alpha = \beta = \delta = 0$, $\lambda = 1$, $\rho = 0.5$, $N = 200$ [E(exp($-u$)) = 0.5100]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.0012	0.2333	0.0544	
	$\hat{\alpha}$	0.0486	0.5434	0.2977	
	$\hat{\beta}$	0.0035	0.2494	0.0622	
	$\hat{\sigma}_v^2$	8.5074	1.3725	2.1263	
	$\hat{\lambda}^2$	1.3239	1.1434	1.4120	
	$\widehat{TE}$	0.5221	0.0982	0.1002	0.2302
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	0.0713	0.5544	0.3124	
	$\tilde{\beta}$	0.0036	0.2262	0.0512	
	$\tilde{\sigma}_v^2$	8.5207	1.4162	2.2351	
	$\tilde{\lambda}^2$	1.4226	1.1787	1.5676	
	$\widetilde{TE}$	0.5138	0.0819	0.0965	0.2771
Restricted NLLS	$\tilde{\eta}$	-1.0015	0.2221	0.0493	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	0.0040	0.2254	0.0508	
$\eta = -1, \beta = 0, \sigma_w^2 = 10$)	$\tilde{\sigma}_w^2$	10.0157	1.0331	1.0675	

STATISTICS	Size1	Size2	Size3	Size4	Mean	s.d.
WALD-OPG	0.0011	0.0014	0.0021	0.0024	-0.0040	0.5877
WALD-HES	0.0041	0.0054	0.0040	0.0045	-0.0067	0.7009
LM-OPG	0.0792	0.0834	0.0506	0.0531	0.9325	1.4464
LM-HES	0.0506	0.0530	0.0674	0.0634	1.4862	21.7190*
GDV	0.0177	0.0160	0.0107	0.0114	0.6292	0.8579
BADGDV	0.0317	0.0305	0.0143	0.0133	0.5772	0.8821
OLS	0.0520	0.0503	0.0326	0.0323	0.0122	0.8829
OLS-H	0.0576	0.0558	0.0369	0.0363	0.0148	0.9078
BADOLS	0.0241	0.0234	0.0128	0.0123	0.0107	0.7641
BADOLS-H	0.0287	0.0286	0.0162	0.0163	0.0115	0.7844
Rep. dropped	0	2352	4689	5349		

* due to outliers

Table 1.9: (Change of δ) $\delta = 0.05$, $\alpha = \beta = 0$, $\sigma_v^2 = \lambda = 1$, $\rho = 0.5$, $N = 1000$ $[E(\exp(-u)) = 0.5232]$

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.0520	0.0450	0.0020	
	$\hat{\alpha}$	-0.0048	0.0753	0.0057	
	$\hat{\beta}$	0.0023	0.0442	0.0020	
	$\hat{\sigma}_v^2$	1.0004	0.0979	0.0096	
	$\hat{\lambda}^2$	0.9935	0.1507	0.0227	
	$\widehat{TE}$	0.5023	0.1802	0.0514	0.6240
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	0.0006	0.0743	0.0055	
	$\tilde{\beta}$	-0.0156	0.0413	0.0020	
	$\tilde{\sigma}_v^2$	0.9964	0.0971	0.0094	
	$\tilde{\lambda}^2$	1.0087	0.1497	0.0225	
	$\widetilde{TE}$	0.5001	0.1803	0.0514	0.6240
Restricted NLLS (OLS on	$\tilde{\eta}$	-1.0009	0.0466	0.0022	
$y_i = \eta + \beta x_i + w_i$: $\eta = -1.0013$,	$\tilde{\beta}$	-0.0238	0.0447	0.0020	
$\beta = -0.0250, \sigma_w^2 = 2.0075$)	$\tilde{\sigma}_w^2$	2.0091	0.1197	0.0143	

STATISTICS	Power1	Power2	Power3	Power4	Mean	s.d.
WALD-OPG	0.1960	0.1960	0.1962	0.1962	1.1269	0.9574
WALD-HES	0.2035	0.2035	0.2037	0.2037	1.1489	0.9714
LM-OPG	0.2270	0.2270	0.2272	0.2272	2.4506	2.7265
LM-HES	0.2090	0.2090	0.2092	0.2092	2.3278	2.5553
GDV	0.2115	0.2115	0.2117	0.2117	2.3049	2.4668
BADGDV	0.1710	0.1710	0.1712	0.1712	2.0401	2.2455
OLS	0.1690	0.1690	0.1687	0.1687	-0.9800	1.0053
OLS-H	0.1670	0.1670	0.1672	0.1672	-0.9824	1.0085
BADOLS	0.1000	0.1000	0.1001	0.1001	-0.8492	0.8713
BADOLS-H	0.0985	0.0985	0.0986	0.0986	-0.8496	0.8727
Rep. dropped	0	0	2	2		

The number of replication is 2000.

Table 1.10: (Change of δ) $\delta = 0.1$, $\alpha = \beta = 0$, $\sigma_v^2 = \lambda = 1$, $\rho = 0.5$, $N = 1000$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.1024	0.0456	0.0021	
	$\hat{\alpha}$	-0.0047	0.0754	0.0057	
	$\hat{\beta}$	0.0022	0.0441	0.0020	
	$\hat{\sigma}_v^2$	1.0003	0.0979	0.0096	
	$\hat{\lambda}^2$	0.9935	0.1517	0.0230	
	$\widehat{TE}$	0.5023	0.1814	0.0513	0.6268
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	0.0092	0.0739	0.0055	
	$\tilde{\beta}$	-0.0329	0.0414	0.0028	
	$\tilde{\sigma}_v^2$	0.9901	0.0970	0.0095	
	$\tilde{\lambda}^2$	1.0336	0.1515	0.0241	
	$\widetilde{TE}$	0.4973	0.1817	0.0515	0.6251
Restricted NLLS (OLS on	$\tilde{\eta}$	-1.0047	0.0467	0.0022	
$y_i = \eta + \beta x_i + w_i$: $\eta = -1.0050$,	$\tilde{\beta}$	-0.0490	0.0449	0.0020	
$\beta = -0.0503, \sigma_w^2 = 2.0304$)	$\tilde{\sigma}_w^2$	2.0305	0.1228	0.0151	

STATISTICS	Power1	Power2	Power3	Power4	Mean	s.d.
WALD-OPG	0.6115	0.6118	0.6115	0.6118	2.1879	0.9419
WALD-HES	0.6350	0.6353	0.6350	0.6353	2.2309	0.9407
LM-OPG	0.6580	0.6578	0.6580	0.6578	6.4945	4.8214
LM-HES	0.6410	0.6408	0.6410	0.6408	6.1642	4.5083
GDV	0.6425	0.6423	0.6425	0.6423	5.8979	4.1321
BADGDV	0.5915	0.5913	0.5915	0.5913	5.4355	4.0116
OLS	0.5105	0.5103	0.5105	0.5103	-1.9417	1.0116
OLS-H	0.5060	0.5058	0.5060	0.5058	-1.9357	1.0040
BADOLS	0.3770	0.3767	0.3770	0.3767	-1.6816	0.8769
BADOLS-H	0.3685	0.3682	0.3685	0.3682	-1.6696	0.8681
Rep. dropped	0	1	0	1		

The number of replication is 2000.

Table 1.11: (Change of δ) $\delta = 0.15$, $\alpha = \beta = 0$, $\sigma_v^2 = \lambda = 1$, $\rho = 0.5$, $N = 1000$ [$E(\exp(-u)) = 0.5232$]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.1528	0.0465	0.0022	
	$\hat{\alpha}$	-0.0046	0.0756	0.0057	
	$\hat{\beta}$	0.0021	0.0440	0.0019	
	$\hat{\sigma}_v^2$	1.0000	0.0979	0.0096	
	$\hat{\lambda}^2$	0.9936	0.1532	0.0235	
	$\widehat{TE}$	0.5023	0.1834	0.0511	0.6314
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	0.0228	0.0739	0.0060	
	$\tilde{\beta}$	-0.0498	0.0415	0.0042	
	$\tilde{\sigma}_v^2$	0.9802	0.0968	0.0098	
	$\tilde{\lambda}^2$	1.0744	0.1556	0.0297	
	$\widetilde{TE}$	0.4925	0.1837	0.0516	0.6269
Restricted NLLS (OLS on	$\tilde{\eta}$	-1.0110	0.0471	0.0022	
$y_i = \eta + \beta x_i + w_i$: $\eta = -1.0113$,	$\tilde{\beta}$	-0.0746	0.0454	0.0021	
$\beta = -0.0758, \sigma_w^2 = 2.0693$)	$\tilde{\sigma}_w^2$	2.0666	0.1282	0.0164	

STATISTICS	Power1	Power2	Power3	Power4	Mean	s.d.
WALD-OPG	0.9100	0.9114	0.9104	0.9118	3.1972	0.9177
WALD-HES	0.9255	0.9269	0.9259	0.9273	3.2605	0.8930
LM-OPG	0.9350	0.9349	0.9354	0.9353	13.1080	7.0740
LM-HES	0.9310	0.9309	0.9314	0.9313	12.4001	6.5219
GDV	0.9280	0.9279	0.9284	0.9283	11.3511	5.6386
BADGDV	0.9065	0.9064	0.9069	0.9068	11.0093	5.9053
OLS	0.8220	0.8217	0.8228	0.8226	-2.9061	1.0167
OLS-H	0.8230	0.8227	0.8238	0.8236	-2.8711	0.9920
BADOLS	0.7425	0.7421	0.7432	0.7429	-2.5151	0.8813
BADOLS-H	0.7300	0.7296	0.7307	0.7303	-2.4673	0.8566
Rep. dropped	0	3	2	5		

The number of replication is 2000.

Table 1.12: (Change of δ and ρ) $\delta = 0.1$, $\rho = 0.9$, $\alpha = \beta = 0$, $\sigma_v^2 = \lambda = 1$, $N = 1000$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.1032	0.0541	0.0029	
	$\hat{\alpha}$	-0.0048	0.0757	0.0058	
	$\hat{\beta}$	0.0032	0.0531	0.0028	
	$\hat{\sigma}_v^2$	1.0007	0.0980	0.0096	
	$\hat{\lambda}^2$	0.9926	0.1525	0.0233	
	$\widehat{TE}$	0.5024	0.1815	0.0514	0.6261
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	0.0052	0.0747	0.0056	
	$\tilde{\beta}$	-0.0607	0.0416	0.0054	
	$\tilde{\sigma}_v^2$	0.9932	0.0975	0.0095	
	$\tilde{\lambda}^2$	1.0257	0.1520	0.0238	
	$\widetilde{TE}$	0.4983	0.1811	0.0516	0.6235
Restricted NLLS (OLS on	$\tilde{\eta}$	-1.0047	0.0468	0.0022	
$y_i = \eta + \beta x_i + w_i$:$\eta = -1.0050$,	$\tilde{\beta}$	-0.0890	0.0452	0.0020	
$\beta = -0.0905, \sigma_w^2 = 2.0304$)	$\tilde{\sigma}_w^2$	2.0252	0.1220	0.0149	

STATISTICS	Power1	Power2	Power3	Power4	Mean	s.d.
WALD-OPG	0.4405	0.4429	0.4405	0.4429	1.8212	0.9378
WALD-HES	0.4655	0.4681	0.4655	0.4681	1.8627	0.9343
LM-OPG	0.4915	0.4902	0.4915	0.4902	4.8166	4.0995
LM-HES	0.4835	0.4816	0.4835	0.4816	4.5987	3.8783
GDV	0.4670	0.4656	0.4670	0.4656	4.2959	3.4242
BADGDV	0.2675	0.2680	0.2675	0.2680	2.7913	2.3952
OLS	0.1715	0.1699	0.1715	0.1699	-0.9813	1.0056
OLS-H	0.1705	0.1689	0.1705	0.1689	-0.9835	1.0083
BADOLS	0.0000	0.0000	0.0000	0.0000	-0.4280	0.4382
BADOLS-H	0.0000	0.0000	0.0000	0.0000	-0.4247	0.4355
Rep. dropped	0	11	0	11		

The number of replication is 2000.

Table 1.13: (Change of scaling functions to $\phi(\delta z_i)/(1-\Phi(\delta z_i))$) $\delta = 0.1$, $\alpha = \beta = 0$, $\sigma_v^2 = \lambda = 1$, $\rho = 0.5$, $N = 1000$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.0827	0.0514	0.0029	
	$\hat{\alpha}$	-0.0065	0.0783	0.0062	
	$\hat{\beta}$	0.0021	0.0418	0.0018	
	$\hat{\sigma}_v^2$	1.0012	0.0938	0.0088	
	$\hat{\lambda}^2$	0.6278	0.1206	0.1531	
	$\widehat{TE}$	0.5600	0.1574	0.0627	0.5509
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	0.0043	0.0764	0.0059	
	$\tilde{\beta}$	-0.0230	0.0389	0.0020	
	$\tilde{\sigma}_v^2$	0.9939	0.0928	0.0086	
	$\tilde{\lambda}^2$	0.6503	0.1187	0.1364	
	$\widetilde{TE}$	0.5557	0.1577	0.0623	0.5485
Restricted NLLS	$\tilde{\eta}$	-0.7987	0.0420	0.0423	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	-0.0305	0.0406	0.0026	
$\eta = -1, \beta = 0, \sigma_w^2 = 2$)	$\tilde{\sigma}_w^2$	1.6471	0.0897	0.1550	

STATISTICS	Power4	Mean	s.d.
WALD-OPG	0.3255	1.5395	0.9178
WALD-HES	0.3540	1.5824	0.9317
LM-OPG	0.4045	3.8867	3.6256
LM-HES	0.3740	3.6723	3.6913
GDV	0.3745	3.4652	3.0224
BADGDV	0.3170	3.1357	2.9020
OLS	0.2825	-1.3685	1.0079
OLS-H	0.2855	-1.3710	1.0093
BADOLS	0.1855	-1.1856	0.8737
BADOLS-H	0.1885	-1.1848	0.8731
Rep. dropped	0		

The number of replication is 2000.

Table 1.14: (Change of the distribution of u_i° to $N(0,\pi/2)^+$) $\alpha = \beta = \delta = 0$, $\sigma_v^2 = \lambda = 1$, $\rho = 0.5$, $N = 1000$ [E(exp(−u)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	-0.0001	0.0746	0.0056	
	$\hat{\alpha}$	-0.4192	0.1093	0.1876	
	$\hat{\beta}$	0.0015	0.0423	0.0018	
	$\hat{\sigma}_v^2$	1.2244	0.1108	0.0626	
	$\hat{\lambda}^2$	0.3455	0.1110	0.4407	
	$\widehat{TE}$	0.6359	0.1128	0.0834	0.5383
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	-0.4106	0.1087	0.1804	
	$\tilde{\beta}$	0.0013	0.0383	0.0015	
	$\tilde{\sigma}_v^2$	1.2185	0.1118	0.0602	
	$\tilde{\lambda}^2$	0.3569	0.1148	0.4268	
	$\widetilde{TE}$	0.6320	0.1131	0.0815	0.5465
Restricted NLLS	$\tilde{\eta}$	-0.9993	0.0396	0.0016	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	0.0015	0.0386	0.0015	
$\eta = -1, \beta = 0, \sigma_w^2 = 2$)	$\tilde{\sigma}_w^2$	1.5730	0.0717	0.1874	

STATISTICS	Size1	Size2	Size3	Size4	Mean	s.d.
WALD-OPG	0.0210	0.0213	0.0221	0.0223	-0.0005	0.8788
WALD-HES	0.0270	0.0274	0.0285	0.0287	-0.0003	0.9174
LM-OPG	0.0635	0.0635	0.0622	0.0622	1.0649	1.5782
LM-HES	0.0635	0.0620	0.0611	0.0595	1.0941	2.0656
GDV	0.0370	0.0371	0.0390	0.0388	0.9229	1.2642
BADGDV	0.0320	0.0320	0.0311	0.0308	0.8265	1.2130
OLS	0.0500	0.0503	0.0496	0.0494	0.0159	0.9915
OLS-H	0.0515	0.0518	0.0506	0.0505	0.0163	0.9963
BADOLS	0.0250	0.0249	0.0253	0.0250	0.0142	0.8591
BADOLS-H	0.0250	0.0249	0.0248	0.0244	0.0145	0.8626
Rep. dropped	0	32	103	118		

The number of replication is 2000.

Table 1.15: (Change of the distribution of u_i° to $gamma(0.5,2)$) $\alpha = \beta = \delta = 0$, $\sigma_v^2 = \lambda = 1$, $\rho = 0.5$, $N = 1000$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.0003	0.0413	0.0017	
	$\hat{\alpha}$	0.4262	0.0651	0.1859	
	$\hat{\beta}$	0.0010	0.0444	0.0020	
	$\hat{\sigma}_v^2$	0.7906	0.0825	0.0507	
	$\hat{\lambda}^2$	2.0376	0.2222	1.1259	
	$\widehat{TE}$	0.4149	0.2146	0.0886	0.6763
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	0.4276	0.0658	0.1871	
	$\tilde{\beta}$	0.0010	0.0422	0.0018	
	$\tilde{\sigma}_v^2$	0.7897	0.0831	0.0511	
	$\tilde{\lambda}^2$	2.0450	0.2243	1.1424	
	$\widetilde{TE}$	0.4144	0.2146	0.0888	0.6763
Restricted NLLS	$\tilde{\eta}$	-1.0003	0.0550	0.0030	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	0.0013	0.0541	0.0029	
$\eta = -1, \beta = 0, \sigma_w^2 = 2$)	$\tilde{\sigma}_w^2$	2.9983	0.2504	1.0594	

STATISTICS	Size4	Mean	s.d.
WALD-OPG	0.0860	0.0059	1.1330
WALD-HES	0.0765	0.0069	1.0860
LM-OPG	0.0630	1.1103	1.5402
LM-HES	0.0765	1.1829	1.6295
GDV	0.0585	1.0788	1.4734
BADGDV	0.0495	0.9787	1.3557
OLS	0.0560	-0.0140	1.0078
OLS-H	0.0540	-0.0143	1.0091
BADOLS	0.0235	-0.0120	0.8725
BADOLS-H	0.0230	-0.0122	0.8722
Rep. dropped	0		

The number of replication is 2000.

Table 1.16: (Change of the distribution of u_i° to $gamma(2,0.5)$) $\alpha = \beta = \delta = 0$, $\sigma_v^2 = \lambda = 1$, $\rho = 0.5$, $N = 1000$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.0013	0.0639	0.0041	
	$\hat{\alpha}$	-0.3772	0.0976	0.1518	
	$\hat{\beta}$	0.0006	0.0413	0.0017	
	$\hat{\sigma}_v^2$	1.1015	0.1041	0.0211	
	$\hat{\lambda}^2$	0.3945	0.1088	0.3785	
	$\widehat{TE}$	0.6187	0.1250	0.0703	0.5223
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	-0.3695	0.0984	0.1462	
	$\tilde{\beta}$	0.0003	0.0379	0.0014	
	$\tilde{\sigma}_v^2$	1.0959	0.1050	0.0202	
	$\tilde{\lambda}^2$	0.4058	0.1132	0.3659	
	$\widetilde{TE}$	0.6154	0.1255	0.0690	0.5263
Restricted NLLS	$\tilde{\eta}$	-0.9997	0.0392	0.0015	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	0.0001	0.0385	0.0015	
$\eta = -1, \beta = 0, \sigma_w^2 = 2$)	$\tilde{\sigma}_w^2$	1.5002	0.0708	0.2548	

STATISTICS	Size1	Size2	Size3	Size4	Mean	s.d.
WALD-OPG	0.0300	0.0301	0.0304	0.0304	0.0204	0.9101
WALD-HES	0.0320	0.0321	0.0324	0.0325	0.0178	0.9430
LM-OPG	0.0640	0.0637	0.0633	0.0634	1.0674	1.4957
LM-HES	0.0600	0.0591	0.0592	0.0588	1.0155	1.4688
GDV	0.0415	0.0411	0.0415	0.0416	0.9545	1.2671
BADGDV	0.0335	0.0331	0.0329	0.0330	0.8294	1.1569
OLS	0.0480	0.0476	0.0471	0.0472	-0.0240	0.9990
OLS-H	0.0500	0.0496	0.0491	0.0492	-0.0236	1.0057
BADOLS	0.0235	0.0231	0.0228	0.0228	-0.0211	0.8648
BADOLS-H	0.0245	0.0241	0.0238	0.0238	-0.0206	0.8694
Rep. dropped	0	5	25	28		

The number of replication is 2000.

1.11 Appendix: LM Test for the Scaled Exponential Case

Recall that the LM statistic of (1.18) is:

$$\begin{aligned}
\mathrm{LM} &= \nabla_\delta \ln\mathrm{L}(\tilde{\theta})'[\tilde{\mathcal{I}}_{\delta\delta} - \tilde{\mathcal{I}}_{\delta\psi}\tilde{\mathcal{I}}_{\psi\psi}^{-1}\tilde{\mathcal{I}}_{\psi\delta}]^{-1}\nabla_\delta \ln\mathrm{L}(\tilde{\theta}) \\
&= \left(\frac{1}{\tilde{\lambda}}\sum_{i=1}^{N}\tilde{b}_i z_i\right)' [\tilde{\mathcal{I}}_{\delta\delta} - \tilde{\mathcal{I}}_{\delta\psi}\tilde{\mathcal{I}}_{\psi\psi}^{-1}\tilde{\mathcal{I}}_{\psi\delta}]^{-1} \left(\frac{1}{\tilde{\lambda}}\sum_{i=1}^{N}\tilde{b}_i z_i\right) \\
&= \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{b}_i z_i\right)' \left[\frac{\tilde{\lambda}^2}{N}\left(\tilde{\mathcal{I}}_{\delta\delta} - \tilde{\mathcal{I}}_{\delta\psi}\tilde{\mathcal{I}}_{\psi\psi}^{-1}\tilde{\mathcal{I}}_{\psi\delta}\right)\right]^{-1} \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{b}_i z_i\right).
\end{aligned} \tag{1.35}$$

Now we compare LM with $\sqrt{N}\hat{\gamma}'(\mathrm{Var}(\sqrt{N}\hat{\gamma}))^{-1}\sqrt{N}\hat{\gamma}$ where the asymptotic distribution of $\sqrt{N}\hat{\gamma}$ is derived in (1.8):

$$\begin{aligned}
&\sqrt{N}\hat{\gamma}'(\mathrm{Var}(\sqrt{N}\hat{\gamma}))^{-1}\sqrt{N}\hat{\gamma} \\
&= \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{b}_i z_i\right)' \left(\frac{1}{N}\sum_{i=1}^{N} z_i z_i'\right)^{-1} BA^{-1}B \left(\frac{1}{N}\sum_{i=1}^{N} z_i z_i'\right)^{-1} \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{b}_i z_i\right) \\
&= \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{b}_i z_i\right)' A^{-1} \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{b}_i z_i\right) + o_p(1) \\
&= \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{b}_i z_i\right)' \left(\frac{1}{N}\sum_{i=1}^{N}(z_i b_i + Gr_i)(z_i b_i + Gr_i)'\right)^{-1} \\
&\quad \times \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{b}_i z_i\right) + o_p(1).
\end{aligned} \tag{1.36}$$

In the following we prove LM and $\sqrt{N}\hat{\gamma}'(\mathrm{Var}(\sqrt{N}\hat{\gamma}))^{-1}\sqrt{N}\hat{\gamma}$ have the same asymptotic distribution by showing that the probability limit of $(N^{-1}\sum_{i=1}^{N}(z_i b_i + Gr_i)(z_i b_i + Gr_i)')$ is $[\lambda^2(\mathcal{I}^\circ_{\delta\delta} - \mathcal{I}^\circ_{\delta\psi}\mathcal{I}^{\circ -1}_{\psi\psi}\mathcal{I}^\circ_{\psi\delta})]$ where $\mathcal{I}^\circ$ is the limiting information matrix as defined in (1.12).

From (1.15), the gradient of the log-likelihood, $\nabla_\theta \ln\mathrm{L}(\theta)$ is:

$$\begin{pmatrix} \frac{\partial \ln\mathrm{L}}{\partial \delta} \\ \frac{\partial \ln\mathrm{L}}{\partial \beta} \\ \frac{\partial \ln\mathrm{L}}{\partial \sigma_v^2} \\ \frac{\partial \ln\mathrm{L}}{\partial \lambda^2} \end{pmatrix} \equiv \begin{pmatrix} \sum_{i=1}^N s_i(\delta) \\ \sum_{i=1}^N s_i(\beta) \\ \sum_{i=1}^N s_i(\sigma_v^2) \\ \sum_{i=1}^N s_i(\lambda^2) \end{pmatrix}. \tag{1.37}$$

Specifically,

$$\begin{aligned}
\sum_{i=1}^N s_i(\delta) &= \sum_{i=1}^N \left(\frac{\sigma_v \xi_i}{\lambda \exp(z_i'\delta)} - \frac{\epsilon_i}{\lambda \exp(z_i'\delta)} - \frac{\sigma_v^2}{\lambda^2 \exp(2z_i'\delta)} - 1 \right) z_i, \\
\sum_{i=1}^N s_i(\beta) &= \sum_{i=1}^N \left(\frac{\xi_i}{\sigma_v} - \frac{1}{\lambda \exp(z_i'\delta)} \right) x_i, \\
\sum_{i=1}^N s_i(\sigma_v^2) &= \sum_{i=1}^N \left(\frac{1}{2\lambda^2 \exp(2z_i'\delta)} - \frac{\xi_i}{2\lambda \exp(z_i'\delta)\sigma_v} + \frac{\epsilon_i \xi_i}{2\sigma_v^3} \right), \\
\sum_{i=1}^N s_i(\lambda^2) &= \sum_{i=1}^N \left(\frac{\sigma_v \xi_i}{2\lambda^3 \exp(3z_i'\delta)} - \frac{\epsilon_i}{2\lambda^3 \exp(3z_i'\delta)} - \frac{\sigma_v^2}{2\lambda^4 \exp(4z_i'\delta)} - \frac{1}{2\lambda^2 \exp(2z_i'\delta)} \right),
\end{aligned} \tag{1.38}$$

where

$$\xi_i = \frac{\phi\left(\epsilon_i/\sigma_v + \sigma_v/(\lambda \exp(z_i'\delta))\right)}{1 - \Phi\left(\epsilon_i/\sigma_v + \sigma_v/(\lambda \exp(z_i'\delta))\right)}. \tag{1.39}$$

Let $H(\theta)$ denote the Hessian, $\nabla_\theta^2 \ln\mathrm{L}(\theta) \equiv \partial^2 \ln\mathrm{L}(\theta)/\partial\theta\partial\theta'$, and $\tilde{H}$ denote $H(\theta)$ evaluated at $\theta = \tilde{\theta} = (0', \tilde{\beta}', \tilde{\sigma}_v^2, \tilde{\lambda}^2)'$. We partition $\tilde{H}$ conformably to (1.17) as

$$\tilde{H} = \begin{pmatrix} \tilde{H}_{\delta\delta} & \tilde{H}_{\delta\psi} \\ \tilde{H}_{\psi\delta} & \tilde{H}_{\psi\psi} \end{pmatrix} \tag{1.40}$$

where $\tilde{H}_{**}$ stands for the *,* block of H, evaluated at the estimates, and $\psi = (\beta', \sigma_v^2, \lambda^2)'$. Specifically, each element of $\tilde{H}$ is:

$$
\begin{aligned}
\tilde{H}_{\delta\delta} &= \tilde{\lambda}^{-2} \sum_{i=1}^{N} \left(\widetilde{\mathrm{Var}(u_i|\epsilon_i)} - \tilde{\lambda}^2 \right) z_i z_i', \\
\tilde{H}_{\beta\delta} &= (\tilde{\lambda}\tilde{\sigma}_v^2)^{-1} \sum_{i=1}^{N} \widetilde{\mathrm{Var}(u_i|\epsilon_i)} x_i z_i', \\
\tilde{H}_{\sigma_v^2\delta} &= (2\tilde{\lambda}\tilde{\sigma}_v^2)^{-1} \sum_{i=1}^{N} \left(\left(\frac{\tilde{\epsilon}_i}{\tilde{\sigma}_v^2} - \frac{1}{\tilde{\lambda}} \right) \widetilde{\mathrm{Var}(u_i|\epsilon_i)} + \tilde{\lambda} \right) z_i', \\
\tilde{H}_{\lambda^2\delta} &= (2\tilde{\lambda}^4)^{-1} \sum_{i=1}^{N} \left(\widetilde{\mathrm{Var}(u_i|\epsilon_i)} - \tilde{\lambda}^2 \right) z_i', \\
\tilde{H}_{\beta\beta} &= \tilde{\sigma}_v^{-4} \sum_{i=1}^{N} \left(\widetilde{\mathrm{Var}(u_i|\epsilon_i)} - \tilde{\sigma}_v^2 \right) x_i x_i', \\
\tilde{H}_{\sigma_v^2\beta} &= (2\tilde{\sigma}_v^4)^{-1} \sum_{i=1}^{N} \left(\left(\frac{\tilde{\epsilon}_i}{\tilde{\sigma}_v^2} - \frac{1}{\tilde{\lambda}} \right) \widetilde{\mathrm{Var}(u_i|\epsilon_i)} - \tilde{\epsilon}_i \right) x_i', \\
\tilde{H}_{\lambda^2\beta} &= (2\tilde{\lambda}^3\tilde{\sigma}_v^2)^{-1} \sum_{i=1}^{N} \widetilde{\mathrm{Var}(u_i|\epsilon_i)} x_i', \\
\tilde{H}_{\sigma_v^2\sigma_v^2} &= (4\tilde{\sigma}_v^6)^{-1} \sum_{i=1}^{N} \left(\left(\frac{\tilde{\epsilon}_i}{\tilde{\sigma}_v} - \frac{\tilde{\sigma}_v}{\tilde{\lambda}} \right)^2 \widetilde{\mathrm{Var}(u_i|\epsilon_i)} - \tilde{\epsilon}_i^2 - 2\tilde{\sigma}_v^2 \right), \\
\tilde{H}_{\lambda^2\sigma_v^2} &= (4\tilde{\lambda}^3\tilde{\sigma}_v^2)^{-1} \sum_{i=1}^{N} \left(\left(\frac{\tilde{\epsilon}_i}{\tilde{\sigma}_v^2} - \frac{1}{\tilde{\lambda}} \right) \widetilde{\mathrm{Var}(u_i|\epsilon_i)} + \tilde{\lambda} \right), \\
\tilde{H}_{\lambda^2\lambda^2} &= (4\tilde{\lambda}^6)^{-1} \sum_{i=1}^{N} \left(\widetilde{\mathrm{Var}(u_i|\epsilon_i)} - \tilde{\lambda}^2 \right),
\end{aligned}
\tag{1.41}
$$

where we use

$$
\begin{aligned}
\mathrm{E}(u_i|\epsilon_i) &= \sigma_v \left(\xi_i - \left(\frac{\epsilon_i}{\sigma_v} + \frac{\sigma_v}{\lambda \exp(z_i'\delta)} \right) \right), \\
\mathrm{Var}(u_i|\epsilon_i) &= \sigma_v^2 \left(1 + \left(\frac{\epsilon_i}{\sigma_v} + \frac{\sigma_v}{\lambda \exp(z_i'\delta)} \right) \xi_i - \xi_i^2 \right).
\end{aligned}
\tag{1.42}
$$

Lemma 1. *Let G be defined as in* (1.9c). *Then, with $\delta = 0$,*

$$
\begin{aligned}
G &= \frac{1}{N}\sum_{i=1}^{N} z_i \nabla_\psi f(y_i, x_i, \psi)' + o_p(1) = \frac{\lambda}{N}\frac{1}{\lambda}\sum_{i=1}^{N} z_i \nabla_\psi f(y_i, x_i, \psi)' + o_p(1) \\
&= \frac{\lambda}{N}\nabla_\psi \left(\frac{1}{\lambda}\sum_{i=1}^{N} z_i b_i\right) + o_p(1) = \frac{\lambda}{N}\nabla_\psi \left(\nabla_\delta \ln\mathrm{L}\,|_{\delta=0}\right) + o_p(1) \\
&= \frac{\lambda}{N}\nabla^2_{\psi\delta} \ln\mathrm{L}\,|_{\delta=0} + o_p(1) = \frac{\lambda}{N} H_{\delta\psi,\delta=0} + o_p(1) \\
&= -\lambda\left(-\frac{1}{N}H_{\delta\psi,\delta=0}\right) + o_p(1) = -\lambda \mathcal{I}^{\circ}_{\delta\psi,\delta=0} + o_p(1).
\end{aligned}
\tag{1.43}
$$

Lemma 2. *With $\delta = 0$,*

$$
\begin{aligned}
\frac{1}{\lambda^2 N}\sum_{i=1}^{N} b_i^2 z_i z_i' &= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\lambda}b_i z_i\right)\left(\frac{1}{\lambda}b_i z_i\right)' = \frac{1}{N}\sum_{i=1}^{N} s_i(\delta) s_i(\delta)'|_{\delta=0} \\
&= \mathcal{I}^{\circ}_{\delta\delta,\delta=0} + o_p(1).
\end{aligned}
\tag{1.44}
$$

Lemma 3. *Consider* $\sum_{i=1}^{N} s_i(\psi) = \left(\sum_{i=1}^{N} s_i(\beta)', \sum_{i=1}^{N} s_i(\sigma_v^2), \sum_{i=1}^{N} s_i(\lambda^2)\right)'$ *as given in* (1.38). *Then, with $\delta = 0$, $N^{-1}\sum_{i=1}^{N} s_i(\psi) b_i z_i' = \lambda \mathcal{I}^{\circ}_{\psi\delta,\delta=0} + o_p(1)$.*

Proof. Note that $N^{-1}\sum_{i=1}^{N} s_i(\psi) b_i z_i'$ is equal to:

$$
\begin{pmatrix}
\frac{-1}{\sigma_v^2 N}\sum_{i=1}^{N} \mathrm{Var}(u_i|\epsilon_i) x_i z_i' \\
\qquad + \sum_{i=1}^{N}\Big(\frac{\lambda}{N}(\frac{1}{\lambda} - \frac{\xi_i}{\sigma_v}) - \frac{1}{\lambda N}(\mathrm{E}\,u_i|\epsilon_i - \mathrm{E}\,u_i)\Big) x_i z_i' \\
\frac{-1}{2\sigma_v^2 N}\sum_{i=1}^{N}\Big((\frac{\epsilon_i}{\sigma_v^2} - \frac{1}{\lambda})\,\mathrm{Var}(u_i|\epsilon_i) - \epsilon_i\Big) z_i' \\
\qquad - \frac{\lambda}{N}\sum_{i=1}^{N}\Big(\frac{1}{2\lambda^2} - \frac{\xi_i}{2\lambda\sigma_v} + \frac{\epsilon_i\xi_i}{2\sigma_v^3}\Big) z_i' \\
\frac{1}{2\lambda^3 N}\sum_{i=1}^{N} b_i^2 z_i'
\end{pmatrix}
$$

$$
= \begin{pmatrix} \dfrac{-1}{\sigma_v^2 N} \sum_{i=1}^{N} \mathrm{Var}(u_i|\epsilon_i) x_i z_i' \\ \quad + \dfrac{1}{\sigma_v^2 N} \sum_{i=1}^{N} \Big(2\sigma_v^2 + \dfrac{\sigma_v^4}{\lambda^2} + \dfrac{\sigma_v^2 \epsilon_i}{\lambda} - (\lambda\sigma_v + \dfrac{\sigma_v^3}{\lambda}) \xi_i \Big) x_i z_i' \\ \dfrac{-1}{2\sigma_v^2 N} \sum_{i=1}^{N} \Big((\dfrac{\epsilon_i}{\sigma_v^2} - \dfrac{1}{\lambda}) \mathrm{Var}(u_i|\epsilon_i) - \epsilon_i \Big) z_i' \\ \quad - \dfrac{\lambda}{N} \sum_{i=1}^{N} \Big(\dfrac{1}{2\lambda^2} - \dfrac{\xi_i}{2\lambda\sigma_v} + \dfrac{\epsilon_i \xi_i}{2\sigma_v^3} \Big) z_i' \\ \dfrac{1}{2\lambda^3 N} \sum_{i=1}^{N} b_i^2 z_i' \end{pmatrix} . \tag{1.45}
$$

Now, the second terms in the first two elements of the above vector converge in probability to zero:

$$
\begin{aligned}
&\frac{1}{\sigma_v^2 N} \sum_{i=1}^{N} \left(2\sigma_v^2 + \frac{\sigma_v^4}{\lambda^2} + \frac{\sigma_v^2 \epsilon_i}{\lambda} - \left(\lambda\sigma_v + \frac{\sigma_v^3}{\lambda} \right) \xi_i \right) x_i z_i' = o_p(1), \\
&\frac{\lambda}{N} \sum_{i=1}^{N} \left(\frac{1}{2\lambda^2} - \frac{\xi_i}{2\lambda\sigma_v} + \frac{\epsilon_i \xi_i}{2\sigma_v^3} \right) z_i' = o_p(1).
\end{aligned} \tag{1.46}
$$

This is because, first,

$$
\begin{aligned}
&\mathrm{E}\left[\frac{1}{\sigma_v^2} \left(2\sigma_v^2 + \frac{\sigma_v^4}{\lambda^2} + \frac{\sigma_v^2 \epsilon_i}{\lambda} - \left(\lambda\sigma_v + \frac{\sigma_v^3}{\lambda} \right) \xi_i \right) x_i z_i' \right] \\
&= \frac{1}{\sigma_v^2} \left(2\sigma_v^2 + \frac{\sigma_v^4}{\lambda^2} + \frac{\sigma_v^2}{\lambda} \mathrm{E}(\epsilon_i) - \left(\lambda\sigma_v + \frac{\sigma_v^3}{\lambda} \right) \mathrm{E}(\xi_i) \right) \mathrm{E}(x_i z_i') \\
&= \frac{1}{\sigma_v^2} \left(2\sigma_v^2 + \frac{\sigma_v^4}{\lambda^2} + \frac{\sigma_v^2}{\lambda}(-\lambda) - \left(\lambda\sigma_v + \frac{\sigma_v^3}{\lambda} \right) \frac{\sigma_v}{\lambda} \right) \mathrm{E}(x_i z_i') = 0,
\end{aligned} \tag{1.47}
$$

where, with $\delta = 0$, ϵ_i and ξ_i are functions of only the error terms, v_i and u_i° which are assumed to be independent of x_i and z_i. Note that $\mathrm{E}(\xi_i) = \sigma_v/\lambda$ since

$$\begin{aligned} \lambda = \mathrm{E}(u_i) = \mathrm{E}[\mathrm{E}(u_i|\epsilon_i)] &= \mathrm{E}\left[\sigma_v\left(\xi_i - \left(\frac{\epsilon_i}{\sigma_v} + \frac{\sigma_v}{\lambda}\right)\right)\right] \\ &= \sigma_v\left(\mathrm{E}(\xi_i) + \frac{\lambda}{\sigma_v} - \frac{\sigma_v}{\lambda}\right), \end{aligned} \tag{1.48}$$

which we solve for $\mathrm{E}(\xi_i)$. Secondly,

$$\lambda\,\mathrm{E}\left(\frac{1}{2\lambda^2} - \frac{\xi_i}{2\lambda\sigma_v} + \frac{\epsilon_i\xi_i}{2\sigma_v^3}\right) = \lambda\,\mathrm{E}\,s_i(\sigma_v^2)|_{\delta=0} = 0 \tag{1.49}$$

where $s_i(\sigma_v^2)$ is as defined in (1.38). Then (1.45) is equal to:

$$\begin{aligned} &\begin{pmatrix} -(\sigma_v^2 N)^{-1}\sum_{i=1}^N \mathrm{Var}(u_i|\epsilon_i)x_i z_i' + o_p(1) \\ -(2\sigma_v^2 N)^{-1}\sum_{i=1}^N\Big(\big(\epsilon_i/\sigma_v^2 - 1/\lambda\big)\,\mathrm{Var}(u_i|\epsilon_i) - \epsilon_i\Big)z_i' + o_p(1) \\ (2\lambda^3 N)^{-1}\sum_{i=1}^N b_i^2 z_i' \end{pmatrix} \\ &= -\frac{\lambda}{N}H_{\psi\delta,\delta=0} + o_p(1) = \lambda\left(-\frac{1}{N}H_{\psi\delta,\delta=0}\right) + o_p(1) = \lambda\mathcal{I}^\circ_{\psi\delta,\delta=0} + o_p(1). \end{aligned} \tag{1.50}$$

□

Recall that $r_i = r_i(\psi) = \mathcal{I}^{\circ-1}s_i(\psi) = \mathcal{I}^{\circ-1}_{\psi\psi,\delta=0}s_i(\psi) = \mathcal{I}^{\circ-1}_{\psi\psi}s_i$ as shown in (1.11). Now we are ready to evaluate our main expression:

$$\frac{1}{N}\sum_{i=1}^N\left[(z_i b_i + Gr_i)(z_i b_i + Gr_i)'\right]$$

$$= \frac{1}{N}\sum_{i=1}^N b_i^2 z_i z_i' \tag{1.51a}$$

$$+ \frac{1}{N}\sum_{i=1}^N G r_i r_i' G' \tag{1.51b}$$

$$+ \frac{1}{N}\sum_{i=1}^N z_i b_i r_i' G' \tag{1.51c}$$

$$+\frac{1}{N}\sum_{i=1}^{N} Gr_i b_i z_i'. \tag{1.51d}$$

We will evaluate these term by term. By Lemma 2,

$$\frac{1}{N}\sum_{i=1}^{N} b_i^2 z_i z_i' = \lambda^2 \mathcal{I}^{\circ}_{\delta\delta} + o_p(1). \tag{1.52}$$

Next, we evaluate (1.51b):

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} Gr_i r_i' G' &= \frac{1}{N}\sum_{i=1}^{N}(-\lambda \mathcal{I}^{\circ}_{\delta\psi})(\mathcal{I}^{\circ -1}_{\psi\psi} s_i)(\mathcal{I}^{\circ -1}_{\psi\psi} s_i)'(-\lambda \mathcal{I}^{\circ}_{\psi\delta}) + o_p(1) \\
&= \lambda^2 \mathcal{I}^{\circ}_{\delta\psi}\mathcal{I}^{\circ -1}_{\psi\psi}\left(\frac{1}{N}\sum_{i=1}^{N} s_i s_i'\right)\mathcal{I}^{\circ -1}_{\psi\psi}\mathcal{I}^{\circ}_{\psi\delta} + o_p(1) \\
&= \lambda^2 \mathcal{I}^{\circ}_{\delta\psi}\mathcal{I}^{\circ -1}_{\psi\psi}\mathcal{I}^{\circ}_{\psi\psi}\mathcal{I}^{\circ -1}_{\psi\psi}\mathcal{I}^{\circ}_{\psi\delta} + o_p(1) \\
&= \lambda^2 \mathcal{I}^{\circ}_{\delta\psi}\mathcal{I}^{\circ -1}_{\psi\psi}\mathcal{I}^{\circ}_{\psi\delta} + o_p(1).
\end{aligned} \tag{1.53}$$

Finally, we evaluate (1.51c):

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} z_i b_i r_i' G' &= \frac{1}{N}\sum_{i=1}^{N} z_i b_i s_i' \mathcal{I}^{\circ -1}_{\psi\psi}(-\lambda \mathcal{I}^{\circ}_{\psi\delta}) + o_p(1) \\
&= (\lambda \mathcal{I}^{\circ}_{\delta\psi})\mathcal{I}^{\circ -1}_{\psi\psi}(-\lambda \mathcal{I}^{\circ}_{\psi\delta}) + o_p(1) \qquad \text{(using Lemma 3)} \\
&= -\lambda^2 \mathcal{I}^{\circ}_{\delta\psi}\mathcal{I}^{\circ -1}_{\psi\psi}\mathcal{I}^{\circ}_{\psi\delta} + o_p(1).
\end{aligned} \tag{1.54}$$

And term (1.51d) is exactly the same as this term.

Inserting (1.52), (1.53) and (1.54) into (1.51), we obtain

$$\frac{1}{N}\sum_{i=1}^{N}(z_i b_i + Gr_i)(z_i b_i + Gr_i)' = \lambda^2\left(\mathcal{I}^{\circ}_{\delta\delta} - \mathcal{I}^{\circ}_{\delta\psi}\mathcal{I}^{\circ -1}_{\psi\psi}\mathcal{I}^{\circ}_{\psi\delta}\right) + o_p(1). \tag{1.55}$$

Therefore the expressions inside the inverse in equations (1.35) and (1.36), for the LM test and the GDV test respectively, have the same probability limit.

1.12 Appendix: Supplementary Tables

Supplemental Table 1.17: (Change of ρ) $\rho = 0.25$, $\alpha = \beta = \delta = 0$, $\sigma_v^2 = \lambda = 1$, $N = 200$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.0003	0.1084	0.0118	
	$\hat{\alpha}$	-0.0338	0.1859	0.0357	
	$\hat{\beta}$	0.0001	0.0951	0.0090	
	$\hat{\sigma}_v^2$	1.0069	0.2271	0.0516	
	$\hat{\lambda}^2$	0.9606	0.3510	0.1247	
	$\widehat{TE}$	0.5145	0.1767	0.0553	0.6145
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	-0.0213	0.1811	0.0333	
	$\tilde{\beta}$	0.0001	0.0933	0.0087	
	$\tilde{\sigma}_v^2$	0.9985	0.2241	0.0502	
	$\tilde{\lambda}^2$	0.9918	0.3459	0.1197	
	$\widetilde{TE}$	0.5098	0.1767	0.0546	0.6194
Restricted NLLS	$\tilde{\eta}$	-1.0003	0.0994	0.0099	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	0.0008	0.1004	0.0101	
$\eta = -1, \beta = 0, \sigma_w^2 = 2$)	$\tilde{\sigma}_w^2$	2.0020	0.2691	0.0724	

STATISTICS	Size1	Size2	Size3	Size4	Mean	s.d.
WALD-OPG	0.0212	0.0214	0.0215	0.0216	-0.0005	0.8719
WALD-HES	0.0303	0.0305	0.0305	0.0306	-0.0019	0.9385
LM-OPG	0.0787	0.0780	0.0769	0.0762	1.2060	1.7718
LM-HES	0.0516	0.0507	0.0512	0.0503	1.0830	6.7671*
GDV	0.0465	0.0465	0.0471	0.0470	1.0121	1.3502
BADGDV	0.0484	0.0475	0.0469	0.0461	0.9524	1.3354
OLS	0.0495	0.0487	0.0483	0.0475	0.0008	0.9966
OLS-H	0.0575	0.0569	0.0563	0.0558	0.0016	1.0212
BADOLS	0.0432	0.0425	0.0420	0.0413	0.0007	0.9650
BADOLS-H	0.0490	0.0484	0.0479	0.0473	0.0009	0.9839
Rep. dropped	0	73	124	172		

* due to outliers

Supplemental Table 1.18: (Change of ρ) $\rho = 0.75$, $\alpha = \beta = \delta = 0$, $\sigma_v^2 = \lambda = 1$, $N = 200$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	-0.0006	0.1280	0.0164	
	$\hat{\alpha}$	-0.0327	0.1869	0.0360	
	$\hat{\beta}$	-0.0003	0.1109	0.0123	
	$\hat{\sigma}_v^2$	1.0055	0.2272	0.0516	
	$\hat{\lambda}^2$	0.9601	0.3527	0.1260	
	$\widehat{TE}$	0.5147	0.1776	0.0557	0.6107
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	-0.0199	0.1804	0.0329	
	$\tilde{\beta}$	0.0001	0.0933	0.0087	
	$\tilde{\sigma}_v^2$	0.9968	0.2228	0.0496	
	$\tilde{\lambda}^2$	0.9944	0.3455	0.1194	
	$\widetilde{TE}$	0.5094	0.1769	0.0546	0.6195
Restricted NLLS	$\tilde{\eta}$	-1.0004	0.0995	0.0099	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	0.0007	0.1005	0.0101	
$\eta = -1, \beta = 0, \sigma_w^2 = 2$)	$\tilde{\sigma}_w^2$	2.0025	0.2692	0.0725	

STATISTICS	Size1	Size2	Size3	Size4	Mean	s.d.
WALD-OPG	0.0213	0.0216	0.0215	0.0217	-0.0060	0.8548
WALD-HES	0.0308	0.0312	0.0309	0.0312	-0.0072	0.9304
LM-OPG	0.0794	0.0790	0.0775	0.0775	1.2305	1.8417
LM-HES	0.0529	0.0527	0.0524	0.0523	1.0285	2.1901
GDV	0.0442	0.0446	0.0445	0.0448	0.9949	1.3236
BADGDV	0.0226	0.0224	0.0222	0.0220	0.6994	1.0175
OLS	0.0495	0.0491	0.0483	0.0480	-0.0003	0.9993
OLS-H	0.0575	0.0574	0.0563	0.0563	0.0005	1.0234
BADOLS	0.0027	0.0027	0.0027	0.0028	0.0000	0.6601
BADOLS-H	0.0052	0.0053	0.0052	0.0052	-0.0008	0.6740
Rep. dropped	0	123	161	244		

Supplemental Table 1.19: (Change of δ and ρ) $\delta = 0.05$, $\rho = 0.9$, $\alpha = \beta = 0$, $\sigma_v^2 = \lambda = 1$, $N = 1000$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.0527	0.0534	0.0029	
	$\hat{\alpha}$	-0.0045	0.0751	0.0057	
	$\hat{\beta}$	0.0033	0.0530	0.0028	
	$\hat{\sigma}_v^2$	1.0004	0.0979	0.0096	
	$\hat{\lambda}^2$	0.9932	0.1506	0.0227	
	$\widehat{TE}$	0.5023	0.1803	0.0515	0.6234
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	-0.0003	0.0744	0.0055	
	$\tilde{\beta}$	-0.0295	0.0414	0.0026	
	$\tilde{\sigma}_v^2$	0.9971	0.0975	0.0095	
	$\tilde{\lambda}^2$	1.0069	0.1498	0.0225	
	$\widetilde{TE}$	0.5006	0.1802	0.0515	0.6236
Restricted NLLS (OLS on	$\tilde{\eta}$	-1.0009	0.0466	0.0022	
$y_i = \eta + \beta x_i + w_i$: $\eta = -1.0013$,	$\tilde{\beta}$	-0.0438	0.0448	0.0020	
$\beta = -0.0451, \sigma_w^2 = 2.0075$)	$\tilde{\sigma}_w^2$	2.0081	0.1194	0.0143	

STATISTICS	Power1	Power2	Power3	Power4	Mean	s.d.
WALD-OPG	0.1380	0.1384	0.1381	0.1385	0.9407	0.9431
WALD-HES	0.1520	0.1525	0.1521	0.1525	0.9617	0.9560
LM-OPG	0.1795	0.1795	0.1796	0.1796	2.0105	2.4353
LM-HES	0.1610	0.1610	0.1611	0.1611	1.9042	2.2712
GDV	0.1650	0.1650	0.1651	0.1651	1.8615	2.1505
BADGDV	0.0560	0.0562	0.0560	0.0562	1.1487	1.3851
OLS	0.0750	0.0747	0.0750	0.0748	-0.4981	1.0049
OLS-H	0.0750	0.0747	0.0750	0.0748	-0.5001	1.0104
BADOLS	0.0000	0.0000	0.0000	0.0000	-0.2175	0.4377
BADOLS-H	0.0000	0.0000	0.0000	0.0000	-0.2174	0.4383
Rep. dropped	0	6	1	7		

The number of replication is 2000.

Supplemental Table 1.20: (Change of δ and ρ) $\delta = 0.15$, $\rho = 0.9$, $\alpha = \beta = 0$, $\sigma_v^2 = \lambda = 1$, $N = 1000$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.1536	0.0549	0.0030	
	$\hat{\alpha}$	-0.0045	0.0758	0.0058	
	$\hat{\beta}$	0.0029	0.0530	0.0028	
	$\hat{\sigma}_v^2$	0.9999	0.0980	0.0096	
	$\hat{\lambda}^2$	0.9931	0.1536	0.0236	
	$\widehat{TE}$	0.5023	0.1836	0.0512	0.6308
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	0.0147	0.0746	0.0058	
	$\tilde{\beta}$	-0.0917	0.0419	0.0102	
	$\tilde{\sigma}_v^2$	0.9856	0.0977	0.0097	
	$\tilde{\lambda}^2$	1.0580	0.1548	0.0273	
	$\widetilde{TE}$	0.4945	0.1828	0.0520	0.6232
Restricted NLLS (OLS on $y_i = \eta + \beta x_i + w_i$: $\eta = -1.0113$, $\beta = -0.1365, \sigma_w^2 = 2.0693$)	$\tilde{\eta}$	-1.0110	0.0470	0.0022	
	$\tilde{\beta}$	-0.1353	0.0459	0.0021	
	$\tilde{\sigma}_w^2$	2.0542	0.1264	0.0162	

STATISTICS	Power1	Power2	Power3	Power4	Mean	s.d.
WALD-OPG	0.7830	0.7865	0.7829	0.7864	2.6717	0.9268
WALD-HES	0.8140	0.8177	0.8139	0.8176	2.7330	0.8988
LM-OPG	0.8360	0.8358	0.8359	0.8357	9.3681	5.8643
LM-HES	0.8260	0.8262	0.8259	0.8261	9.0193	5.6024
GDV	0.8125	0.8122	0.8124	0.8121	7.9604	4.5539
BADGDV	0.6310	0.6308	0.6308	0.6307	5.5233	3.5299
OLS	0.3155	0.3154	0.3157	0.3156	-1.4680	1.0098
OLS-H	0.3155	0.3149	0.3157	0.3151	-1.4676	1.0078
BADOLS	0.0005	0.0005	0.0005	0.0005	-0.6396	0.4401
BADOLS-H	0.0000	0.0000	0.0000	0.0000	-0.6271	0.4319
Rep. dropped	0	9	1	10		

The number of replication is 2000.

Supplemental Table 1.21: (Change of the distribution of u_i° to $N(0,1)^+$) $\alpha = \beta = \delta = 0$, $\sigma_v^2 = \lambda = 1$, $\rho = 0.5$, $N = 1000$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.0010	0.0984	0.0097	
	$\hat{\alpha}$	-0.3498	0.1193	0.1366	
	$\hat{\beta}$	0.0014	0.0396	0.0016	
	$\hat{\sigma}_v^2$	1.1488	0.0995	0.0321	
	$\hat{\lambda}^2$	0.2119	0.0933	0.6298	
	$\widehat{TE}$	0.6958	0.0885	0.0835	0.4635
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	-0.3395	0.1159	0.1286	
	$\tilde{\beta}$	0.0011	0.0356	0.0013	
	$\tilde{\sigma}_v^2$	1.1434	0.0996	0.0305	
	$\tilde{\lambda}^2$	0.2218	0.0955	0.6147	
	$\widetilde{TE}$	0.6901	0.0881	0.0807	0.4832
Restricted NLLS	$\tilde{\eta}$	-0.7972	0.0365	0.0425	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	0.0012	0.0357	0.0013	
$\eta = -1, \beta = 0, \sigma_w^2 = 2$)	$\tilde{\sigma}_w^2$	1.3658	0.0612	0.4060	

STATISTICS	Size1	Size2	Size3	Size4	Mean	s.d.
WALD-OPG	0.0125	0.0132	0.0149	0.0153	0.0178	0.8294
WALD-HES	0.0205	0.0217	0.0217	0.0223	0.0183	0.8670
LM-OPG	0.0560	0.0565	0.0490	0.0491	1.0036	1.5132
LM-HES	0.0725	0.0713	0.0657	0.0637	2.0072	34.9627*
GDV	0.0255	0.0269	0.0291	0.0300	0.8246	1.1256
BADGDV	0.0320	0.0328	0.0291	0.0293	0.7639	1.1324
OLS	0.0510	0.0486	0.0434	0.0427	-0.0098	0.9697
OLS-H	0.0540	0.0502	0.0453	0.0446	-0.0093	0.9744
BADOLS	0.0260	0.0254	0.0229	0.0229	-0.0082	0.8404
BADOLS-H	0.0240	0.0232	0.0205	0.0204	-0.0079	0.8439
Rep. dropped	0	107	387	431		

* due to outliers. The number of replication is 2000.

Supplemental Table 1.22: (Change of the distribution of u_i° to $N(0, \pi/(\pi - 2))^+$) $\alpha = \beta = \delta = 0$, $\sigma_v^2 = \lambda = 1$, $\rho = 0.5$, $N = 1000$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	-0.0002	0.0548	0.0030	
	$\hat{\alpha}$	-0.5135	0.1031	0.2744	
	$\hat{\beta}$	0.0016	0.0460	0.0021	
	$\hat{\sigma}_v^2$	1.3480	0.1296	0.1379	
	$\hat{\lambda}^2$	0.6616	0.1488	0.1366	
	$\widehat{TE}$	0.5541	0.1456	0.0774	0.6116
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	-0.5069	0.1047	0.2679	
	$\tilde{\beta}$	0.0017	0.0423	0.0018	
	$\tilde{\sigma}_v^2$	1.3420	0.1311	0.1342	
	$\tilde{\lambda}^2$	0.6744	0.1533	0.1295	
	$\widetilde{TE}$	0.5518	0.1460	0.0764	0.6140
Restricted NLLS	$\tilde{\eta}$	-1.3225	0.0448	0.1060	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	0.0019	0.0433	0.0019	
$\eta = -1, \beta = 0, \sigma_w^2 = 2$)	$\tilde{\sigma}_w^2$	2.0018	0.0938	0.0088	

STATISTICS	Size1	Size2	Size3	Size4	Mean	s.d.
WALD-OPG	0.0295	0.0296	0.0296	0.0297	-0.0025	0.9065
WALD-HES	0.0395	0.0396	0.0396	0.0397	-0.0009	0.9490
LM-OPG	0.0605	0.0607	0.0607	0.0608	1.0705	1.5630
LM-HES	0.0535	0.0536	0.0532	0.0533	0.9746	1.4373
GDV	0.0510	0.0511	0.0512	0.0513	0.9959	1.3894
BADGDV	0.0390	0.0391	0.0391	0.0392	0.8596	1.2448
OLS	0.0480	0.0481	0.0481	0.0483	0.0098	1.0040
OLS-H	0.0500	0.0501	0.0502	0.0503	0.0099	1.0085
BADOLS	0.0260	0.0261	0.0261	0.0261	0.0089	0.8698
BADOLS-H	0.0270	0.0271	0.0271	0.0271	0.0091	0.8730
Rep. dropped	0	5	6	11		

The number of replication is 2000.

Supplemental Table 1.23: (Change of the distribution of u_i° to $N(1,1)^+$) $\alpha = \beta = \delta = 0$, $\sigma_v^2 = \lambda = 1$, $\rho = 0.5$, $N = 1000$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.0000	0.0992	0.0098	
	$\hat{\alpha}$	-0.8053	0.1340	0.6665	
	$\hat{\beta}$	0.0015	0.0431	0.0019	
	$\hat{\sigma}_v^2$	1.3787	0.1197	0.1577	
	$\hat{\lambda}^2$	0.2470	0.1141	0.5800	
	$\widehat{TE}$	0.6807	0.0910	0.1564	0.5174
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	-0.7959	0.1294	0.6502	
	$\tilde{\beta}$	0.0015	0.0391	0.0015	
	$\tilde{\sigma}_v^2$	1.3738	0.1193	0.1539	
	$\tilde{\lambda}^2$	0.2565	0.1144	0.5658	
	$\widetilde{TE}$	0.6755	0.0884	0.1471	0.5412
Restricted NLLS	$\tilde{\eta}$	-1.2871	0.0405	0.0840	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	0.0015	0.0392	0.0015	
$\eta = -1, \beta = 0, \sigma_w^2 = 2$)	$\tilde{\sigma}_w^2$	1.6312	0.0736	0.1414	

STATISTICS	Size1	Size2	Size3	Size4	Mean	s.d.
WALD-OPG	0.0055	0.0058	0.0069	0.0070	0.0118	0.7864
WALD-HES	0.0120	0.0127	0.0119	0.0122	0.0148	0.8308
LM-OPG	0.0605	0.0602	0.0469	0.0480	0.9462	1.3146
LM-HES	0.0745	0.0755	0.0695	0.0698	1.2467	4.0045*
GDV	0.0195	0.0195	0.0188	0.0192	0.7776	0.9973
BADGDV	0.0285	0.0280	0.0200	0.0205	0.7140	0.9702
OLS	0.0540	0.0512	0.0438	0.0423	-0.0224	0.9365
OLS-H	0.0550	0.0517	0.0444	0.0423	-0.0228	0.9407
BADOLS	0.0215	0.0211	0.0144	0.0141	-0.0196	0.8114
BADOLS-H	0.0215	0.0211	0.0138	0.0135	-0.0202	0.8146
Rep. dropped	0	106	402	439		

* due to outliers. The number of replication is 2000.

Supplemental Table 1.24: (Change of the distribution of u_i° to $gamma(0.5, \sqrt{2})$)
$\alpha = \beta = \delta = 0$, $\sigma_v^2 = \lambda = 1$, $\rho = 0.5$, $N = 1000$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.0004	0.0446	0.0020	
	$\hat{\alpha}$	0.3384	0.0667	0.1190	
	$\hat{\beta}$	0.0009	0.0412	0.0017	
	$\hat{\sigma}_v^2$	0.8493	0.0812	0.0293	
	$\hat{\lambda}^2$	1.0968	0.1470	0.0309	
	$\widehat{TE}$	0.4905	0.1903	0.0837	0.6173
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	0.3389	0.0660	0.1192	
	$\tilde{\beta}$	0.0008	0.0389	0.0015	
	$\tilde{\sigma}_v^2$	0.8494	0.0810	0.0292	
	$\tilde{\lambda}^2$	1.0996	0.1452	0.0310	
	$\widetilde{TE}$	0.4902	0.1902	0.0838	0.6175
Restricted NLLS	$\tilde{\eta}$	-0.7075	0.0449	0.0876	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	0.0009	0.0440	0.0019	
$\eta = -1, \beta = 0, \sigma_w^2 = 2$)	$\tilde{\sigma}_w^2$	1.9974	0.1367	0.0187	

STATISTICS	Size4	Mean	s.d.
WALD-OPG	0.0710	0.0081	1.0704
WALD-HES	0.0620	0.0089	1.0491
LM-OPG	0.0630	1.0989	1.5310
LM-HES	0.0655	1.1158	1.5362
GDV	0.0525	1.0550	1.4346
BADGDV	0.0425	0.9360	1.3003
OLS	0.0495	-0.0156	0.9961
OLS-H	0.0520	-0.0155	0.9984
BADOLS	0.0225	-0.0132	0.8622
BADOLS-H	0.0215	-0.0132	0.8623
Rep. dropped	0		

The number of replication is 2000.

Supplemental Table 1.25: (Change of the distribution of u_i° to $gamma(2, 1/\sqrt{2})$)
$\alpha = \beta = \delta = 0$, $\sigma_v^2 = \lambda = 1$, $\rho = 0.5$, $N = 1000$ [E(exp($-u$)) = 0.5232]

ESTIMATION METHODS	Estimates	Mean	s.d.	MSE	Corr
MLE	$\hat{\delta}$	0.0007	0.0488	0.0024	
	$\hat{\alpha}$	-0.5069	0.0880	0.2647	
	$\hat{\beta}$	0.0009	0.0452	0.0020	
	$\hat{\sigma}_v^2$	1.1766	0.1170	0.0449	
	$\hat{\lambda}^2$	0.8278	0.1491	0.0519	
	$\widehat{TE}$	0.5253	0.1637	0.0684	0.6072
Restricted MLE ($\delta = 0$)	$\tilde{\alpha}$	-0.5013	0.0901	0.2594	
	$\tilde{\beta}$	0.0007	0.0422	0.0018	
	$\tilde{\sigma}_v^2$	1.1716	0.1179	0.0433	
	$\tilde{\lambda}^2$	0.8403	0.1540	0.0492	
	$\widetilde{TE}$	0.5235	0.1641	0.0677	0.6084
Restricted NLLS	$\tilde{\eta}$	-1.4140	0.0452	0.1735	
(OLS on $y_i = \eta + \beta x_i + w_i$:	$\tilde{\beta}$	0.0005	0.0445	0.0020	
$\eta = -1, \beta = 0, \sigma_w^2 = 2$)	$\tilde{\sigma}_w^2$	2.0009	0.1020	0.0104	

STATISTICS	Size4	Mean	s.d.
WALD-OPG	0.0395	0.0161	0.9434
WALD-HES	0.0415	0.0153	0.9745
LM-OPG	0.0610	1.0617	1.4787
LM-HES	0.0475	0.9753	1.3569
GDV	0.0535	1.0122	1.3750
BADGDV	0.0335	0.8640	1.2086
OLS	0.0465	-0.0230	1.0065
OLS-H	0.0500	-0.0225	1.0130
BADOLS	0.0235	-0.0202	0.8711
BADOLS-H	0.0240	-0.0197	0.8754
Rep. dropped	0		

The number of replication is 2000.

Chapter 2

On the Accuracy of Bootstrap Confidence Intervals for Efficiency Levels in Stochastic Frontier Models with Panel Data

2.1 Introduction

This chapter is concerned with the construction of confidence intervals for efficiency levels of individual firms in stochastic frontier models with panel data. A number of different techniques have been proposed in this literature to address this problem. Given a distributional assumption for technical inefficiency, maximum likelihood estimation was proposed by Pitt and Lee (1981). Battese and Coelli (1988) showed how to construct point estimates of technical efficiency for each firm, and Horrace and Schmidt (1996) showed how to construct confidence intervals for these efficiency levels. Without a distributional assumption for technical efficiency, Schmidt and Sickles (1984) proposed fixed effects estimation, and the point estimation problem for efficiency levels was discussed by Schmidt and Sickles (1984) and Park and Simar (1994). Simar (1992) and Hall, Härdle, and Simar (1993) suggested using bootstrapping to conduct inference on the efficiency levels. Horrace and Schmidt (1996) and Horrace and Schmidt (2000) constructed confidence intervals using the theory of multiple comparisons with the best, and Kim and Schmidt (1999) suggested a univariate version of comparisons with the best. Bayesian methods have been suggested by Koop, Osiewalski, and Steel (1997) and Osiewalski and Steel (1998).

In this chapter we will focus on bootstrapping and some related procedures. We provide a survey of various versions of the bootstrap for construction of confidence intervals for efficiency levels. We also propose a simple alternative to the bootstrap that uses standard parametric methods, acting as if the identity of the best firm is known with certainty, and we propose some new resampling methods that correspond to this parametric procedure. We present Monte Carlo simulation evidence on the accuracy of the bootstrap and our simple alternative. Finally, we present some empirical results to indicate how these methods work in practice.

2.2 Fixed-Effects Estimation of the Model

Consider the basic panel data stochastic frontier model of Pitt and Lee (1981) and Schmidt and Sickles (1984),

$$y_{it} = \alpha + x_{it}'\beta + v_{it} - u_i, \quad i = 1, \cdots, N, \quad t = 1, \cdots, T, \tag{2.1}$$

where i indexes firms or productive units and t indexes time periods. y_{it} is the scalar dependent variable representing the logarithm of output for the i^{th} firm in period t, α is a scalar intercept, x_{it} is a $K \times 1$ column vector of inputs (e.g., in logarithms for the Cobb-Douglas specification), β is a $K \times 1$ vector of coefficients, and v_{it} is an i.i.d. error term with zero mean and finite variance. The time-invariant u_i satisfy $u_i \geq 0$, and

$u_i > 0$ is an indication of technical inefficiency. For a logarithmic specification such as Cobb-Douglas, the technical efficiency of the i^{th} firm is defined as $r_i = \exp(-u_i)$, so technical inefficiency is $1 - r_i$. For small values of u_i, u_i is approximately equal to $1 - \exp(-u_i) = 1 - r_i$, so that u_i itself is sometimes used as a measure of technical inefficiency.

Now define $\alpha_i = \alpha - u_i$. With this definition, (2.1) becomes the standard panel data model with time-invariant individual effects:

$$y_{it} = \alpha_i + x'_{it}\beta + v_{it}. \tag{2.2}$$

Obviously we have $u_i = \alpha - \alpha_i$ and $\alpha_i \leq \alpha$ since $u_i \geq 0$. The previous discussion regards zero as the minimal possible value of u_i and α as the maximal possible value of α_i over any possible sample; that is, essentially, as $N \to \infty$. It is also useful to consider the following representation in a given sample size of N. We write the intercepts α_i in ranked order, as:

$$\alpha_{(1)} \leq \alpha_{(2)} \leq \cdots \leq \alpha_{(N)} \tag{2.3}$$

so that in particular (N) is the index of the firm with largest value of α_i among N firms. It is convenient to write the values of u_i in the opposite ranked order, as $u_{(N)} \leq \cdots \leq u_{(2)} \leq u_{(1)}$, so that $\alpha_{(i)} = \alpha - u_{(i)}$. Then obviously $\alpha_{(N)} = \alpha - u_{(N)}$, and firm (N) has the largest value of α_i or equivalently the smallest value of u_i among N firms. We will call this firm the *best* firm in the sample. In some methods we measure inefficiency relative to the best firm in the sample, and this corresponds to considering the relative efficiency measures:

$$u_i^* = u_i - u_{(N)} = \alpha_{(N)} - \alpha_i, \quad r_i^* = \exp(-u_i^*). \tag{2.4}$$

Fixed effects estimation refers to the estimation of the panel data regression model (2.2), treating α_i as fixed parameters. Because the α_i are treated as parameters, we do not need to make any distributional assumption about the inefficiencies; nor do we need to assume that they are uncorrelated with the x_{it} or the v_{it}. We assume strict exogeneity of the regressors x_{it}, in the sense that $(x_{i1}, x_{i2}, \cdots, x_{iT})$ are independent of $(v_{i1}, v_{i2}, \cdots, v_{iT})$. We also assume that the v_{it} are i.i.d. with zero mean and constant variance σ_v^2. We do not need to assume a distribution for the v_{it}.

The fixed effects estimates $\hat{\beta}$, also called the *within* estimates, may be calculated by regressing $(y_{it} - \bar{y}_i)$ on $(x_{it} - \bar{x}_i)$, or equivalently by regressing y_{it} on x_{it} and a set of N dummy variables for firms. We then obtain $\hat{\alpha}_i = \bar{y}_i - \bar{x}_i'\hat{\beta}$, or equivalently the $\hat{\alpha}$ are the estimated coefficients of the dummy variables. This leads to the following expression for $\hat{\alpha}_i$:

$$\hat{\alpha}_i = \alpha_i + \bar{v}_i - \bar{x}_i'(\hat{\beta} - \beta). \tag{2.5}$$

The fixed effects estimate $\hat{\beta}$ is consistent as $NT \to \infty$, and its variance is of order $(N(T-1))^{-1}$. For a given firm i, the estimated intercept $\hat{\alpha}_i$ is a consistent estimate of α_i as $T \to \infty$. Large T is needed for the term $\bar{v}_i$ in (2.5) to become negligible.

Schmidt and Sickles (1984) suggested the following estimates of technical inefficiency, based on the fixed effects estimates:

$$\hat{\alpha} = \max_j \hat{\alpha}_j, \quad \hat{u}_i^* = \hat{\alpha} - \hat{\alpha}_i. \tag{2.6}$$

Since these estimates clearly measure inefficiency relative to the firm estimated to be the best in the sample, they are naturally viewed as estimates of $\alpha_{(N)}$ and u_i^*, that is, of relative rather than absolute inefficiency.

We define some further notation. Suppose we write the estimates $\hat{\alpha}_i$ in ranked order, as follows:

$$\hat{\alpha}_1 \leq \hat{\alpha}_2 \leq \cdots \leq \hat{\alpha}_{[N]}. \tag{2.7}$$

So $[N]$ is the index of the firm with the largest $\hat{\alpha}_i$, whereas (N) was the index of the firm with the largest α_i. These may not be the same; for example, firm 129 could be the true best firm (that is, the one with the biggest α_i), so that $(N) = 129$, but firm 71 could be the estimated best firm (that is, the one with the biggest $\hat{\alpha}_i$), so that $[N] = 71$. Note also that $\hat{\alpha}$ as defined in (2.6) above is the same as $\hat{\alpha}_{[N]}$, but it may not be the same as $\hat{\alpha}_{(N)}$, the estimated α for the *unknown* best firm.

As $T \to \infty$ with N fixed, $\hat{\alpha}$ is a consistent estimate of $\alpha_{(N)}$ and $\hat{u}_i^*$ is a consistent estimate of u_i^*. However, it is important to note that in finite samples (for small T) $\hat{\alpha}$ is likely to be biased upward, since $\hat{\alpha} \geq \hat{\alpha}_{(N)}$ and $\mathrm{E}(\hat{\alpha}_{(N)}) = \alpha_{(N)}$. That is, the "max" operator in (2.6) induces upward bias, since the largest $\hat{\alpha}_i$ is more likely to contain positive estimation error than negative error. This bias is larger when N is larger and when the $\hat{\alpha}_i$ are estimated less precisely. The upward bias in $\hat{\alpha}$ induces an upward bias in the $\hat{u}_i^*$ and a downward bias in $\hat{r}_i^* = \exp(-\hat{u}_i^*)$; we underestimate efficiency because we overestimate the level of the frontier.

Schmidt and Sickles (1984) argued that $\hat{\alpha}$ and $\hat{u}_i^*$ are consistent estimates of α and u_i if both N and T approach ∞; that is, if both N and T are large, we can regard the $\hat{u}_i^*$ as estimates of absolute and not just relative inefficiency. The argument is simple. As $T \to \infty$, $\hat{\alpha}$ and $\hat{u}_i^*$ are consistent estimates of $\alpha_{(N)}$ and u_i^*, as noted above. As $N \to \infty$, $u_{(N)}$ should converge to 0 so that $\alpha_{(N)}$ converges to α and the u_i^* should converge to the corresponding u_i. A more rigorous treatment of the asymptotics for this model is given by Park and Simar (1994), who show that, in addition to $N \to \infty$ and $T \to \infty$, we need to require $T^{-1/2} \ln N \to 0$ in order to ensure the consistency of $\hat{\alpha}$ as an estimate of α. This latter requirement limits the rate at which N can grow relative to T in order to ensure that the upward bias induced by the max operation disappears asymptotically.

2.3 Construction of Confidence Intervals by Bootstrapping

We can use bootstrapping to construct confidence intervals for functions of the fixed effects estimates. The inefficiency measures $\hat{u}_i^*$ and the efficiency measures $r^* = \exp(-\hat{u}_i^*)$ are functions of the fixed effects estimates and so bootstrapping can be used for inference on these measures.

We begin with a very brief discussion of bootstrapping in the general setting in which we have a parameter θ, and there is an estimate $\hat{\theta}$ based on a sample $z_1, \cdots, z_n$ of i.i.d. random variables. The estimator $\hat{\theta}$ is assumed to be regular enough so that $n^{1/2}(\hat{\theta} - \theta)$ is asymptotically normal. The following bootstrap procedure will be repeated many times, say for $b = 1, \cdots, B$ where B is large. For iteration b, construct pseudo data $z_1^{(b)}, \cdots, z_n^{(b)}$ by sampling randomly with replacement from the original data $z_1, \cdots, z_n$. From the pseudo data, construct the estimate $\hat{\theta}^{(b)}$. The basic result of the bootstrap is that, under fairly general circumstances, the asymptotic (large n) distribution of $n^{1/2}(\hat{\theta}^{(b)} - \hat{\theta})$ conditional on the sample is the same as the (unconditional) asymptotic distribution of $n^{1/2}(\hat{\theta} - \theta)$. Thus for large n the distribution of $\hat{\theta}$ around the unknown θ is the same as the bootstrap distribution of $\hat{\theta}^{(b)}$ around $\hat{\theta}$, which is revealed by a large number (B) of draws.

We now consider the application of the bootstrap to the specific case of the fixed effects estimates. Our discussion follows Simar (1992). Let the fixed effects estimates be $\hat{\beta}$ and $\hat{\alpha}_i$, from which we calculate $\hat{u}_i^*$ and $\hat{r}_i^*$ ($i = 1, \cdots, N$). Let the residuals be $\hat{v}_{it} = y_{it} - \hat{\alpha}_i - x_{it}'\hat{\beta}$ ($i = 1, \cdots, N$, $t = 1, \cdots, T$). The bootstrap samples will be drawn by resampling these residuals, because the v_{it} are the quantities analogous to the z's in the previous paragraph, in the sense that they are assumed to be i.i.d., and the $\hat{v}_{it}$ are the observable versions of the v_{it}. (The sample size n above corresponds to NT). So, for bootstrap iteration b $(= 1, \cdots, B)$ we calculate the bootstrap sample $\hat{v}_{it}^{(b)}$ and the pseudo data $y_{it}^{(b)} = \hat{\alpha}_i + x_{it}'\hat{\beta} + \hat{v}_{it}^{(b)}$. From these data we get the bootstrap estimates $\hat{\beta}^{(b)}$, $\hat{\alpha}_i^{(b)}$, $\hat{u}_i^{*(b)}$, and $\hat{r}_i^{*(b)}$, and the bootstrap distribution of these estimates is used to make inferences about the parameters.

We note that the estimates $\hat{u}_i^*$ and $\hat{r}_i^*$ depend on the quantity $\max_j \hat{\alpha}_j$. Since "max" is not a smooth function, it is not immediately apparent that this quantity is asymptotically normal, and if it were not the validity of the bootstrap would be in doubt. A rigorous proof of the validity of the bootstrap for this problem is given by Hall, Härdle, and Simar (1995). They prove the equivalence of the following three statements: (i) $\max_j \hat{\alpha}_j$ is asymptotically normal. (ii) The bootstrap is valid as $T \to \infty$ with N fixed. (iii) There are no ties for $\max_j \alpha_j$: that is, there are a unique index (N) such that $\alpha_{(N)} = \max_j \alpha_j$. There are two important implications of this result. First, the bootstrap will not be reliable unless T is large. Second, this is especially true if there are near ties for $\max_j \alpha_j$, in other words, when there is substantial uncertainty about which firm is best.

We now turn to specific bootstrapping procedures, which differ in the way they draw inferences based on the bootstrap estimates. In each case, suppose that we are trying to construct a confidence interval for $u_i^* = \max_j \alpha_j - \alpha_i$. That is, for a given confidence level c, we seek lower and upper bounds L_i, U_i such that $\mathrm{P}(L_i \leq u_i^* \leq U_i) = 1 - c$.

The simplest version of the bootstrap is the percentile bootstrap. Here we simply take L_i and U_i to be the upper and lower $c/2$ fractiles of the bootstrap distribution of the $\hat{u}_i^{*(b)}$. More formally, let $\hat{F}$ be the cumulative distribution function (cdf) for $\hat{u}_i^*$ so that $\hat{F}(s) = \mathrm{P}(\hat{u}_i^{*(b)} \leq s) =$ the fraction of the B bootstrap replications in which $\hat{u}_i^{*(b)} \leq s$. Then, we take $L_i = \hat{F}^{-1}(c/2)$ and $U_i = \hat{F}^{-1}(1 - c/2)$.

The percentile bootstrap intervals are accurate for large T but may be inaccurate for small to moderate T. This is a general statement, but in the present context there is a specific reason to be worried, which is the finite sample upward bias in $\max_j \hat{\alpha}_j$ as an estimate of $\max_j \alpha_j$. This will be reflected in improper centering of the intervals and therefore inaccurate coverage probabilities. Simulation evidence on the severity of this problem is given by Hall, Härdle, and Simar (1993) and in Section 2.5 of this chapter.

Several more sophisticated versions of the bootstrap have been suggested to construct confidence intervals with higher coverage probabilities. Hall, Härdle, and Simar (1993) and Hall, Härdle, and Simar (1995) suggested the *iterated bootstrap*, also called the double bootstrap, which consists of two stages. The first stage is the usual percentile bootstrap which constructs, for any given c, a confidence interval that is intended to hold with probability of $1-c$. We will call these "nominal" $1-c$ confidence intervals. The second stage of the bootstrap is used to estimate the true coverage probability of the nominal $1-c$ confidence intervals, as a function of c. That is, if we define the function $\pi(c)$ = true coverage probability level of the nominal $1-c$ level confidence interval from the percentile bootstrap, then we attempt to evaluate the function $\pi(c)$. When we have done so, we find c^*, say, such that $\pi(c^*) = 1-c$, and then we use as our confidence interval from the first stage percentile bootstrap, which we "expect" to have a true coverage probability of $1-c$.

The mechanics of the iterated bootstrap are uncomplicated but time-consuming. For each of the original (first stage) bootstrap iterations B, the second stage involves a set of B_2 draws from the bootstrap residuals, construction of pseudo data, and construction of percentile confidence intervals, which then either do or do not cover the original estimate $\hat{\theta}$. The coverage probability function $\pi(c)$, which is the actual rate at which a nominal c-level interval based on the bootstrap estimates covers the true parameter θ, is estimated by the rate at which a nominal c-level interval based on the iterated bootstrap estimates covers the original estimate $\hat{\theta}$. To understand this, note that data generated from the true θ yield $\hat{\theta}$; bootstrap data generated based on $\hat{\theta}$ yield the bootstrap estimates $\hat{\theta}^{(b)}$; and data based on $\hat{\theta}^{(b)}$ yield the iterated bootstrap estimates, say $\hat{\theta}^{(b,b^1)}$. So the iterated bootstrap estimates $\hat{\theta}^{(b,b^1)}$ have the same relationship to $\hat{\theta}$ as the bootstrap estimates $\hat{\theta}^{(b)}$ have to θ.

Generally we take $B_2 = B$, so that the total number of draws has increased from B to B^2. by going to the iterated bootstrap. Theoretically, the error in the percentile bootstrap is of order $n^{-1/2}$ while the error in the iterated bootstrap is of order n^{-1}. There is no clear connection between this statement and the question of how well finite sample bias is handled.

An objection to the iterated bootstrap is that it does not explicitly handle bias. For example, if the nominal 90% confidence intervals only cover 75% of the bootstrap estimate in the first stage, it simply insists on a higher nominal confidence level, like 98%, so as to get 90% coverage. That is, it just makes the intervals wider when bias might more reasonably be handled by recentering the intervals. A technique that does recenter the intervals is the bias-adjusted bootstrap of Efron (1982) and Efron (1985). As above, let θ be the parameter of interest, $\hat{\theta}$ the sample estimate and $\hat{\theta}^{(b)}$ the bootstrap estimate (for $b = 1, \cdots, B$), and $\hat{F}$ the bootstrap cdf. For n large enough that the bootstrap is accurate, we should expect $\hat{F}(\hat{\theta}) = 0.5$, and failure of this to occur is a suggestion of bias. Now define $z_0 = \Phi^{-1}(\hat{F}(\hat{\theta}))$ where Φ is a standard normal cdf, and where $\hat{F}(\hat{\theta}) = 0.5$ would imply $z_0 = 0$. Let $z_{c/2}$ be the usual normal critical value; e.g. for $c = 0.1$, $z_{c/2} = z_{0.05} = 1.645$. Then, the bias-adjusted bootstrap confidence interval is $[L_i, U_i]$ with:

$$L_i = \hat{F}^{-1}(\Phi(2z_0 - z_{c/2})), \quad U_i = \hat{F}^{-1}(\Phi(2z_0 + z_{c/2})) \tag{2.8}$$

For example, suppose that there is an upward bias, reflected by the fact that 60%

of the bootstrap draws are larger than $\hat{\theta}$, so that $\hat{F}(\hat{\theta}) = 0.4$. Then $z_0 = -0.253$, and for $c = 0.1$ we have $\Phi(2z_0 - z_{c/2}) = \Phi(-2.152) = 0.016$ and $\Phi(2z_0 + z_{c/2}) = 0.873$. Thus our confidence interval comes from the lower tail 0.016 fractile and the upper tail 0.127 fractile, and we have compensated for upward bias by moving the interval left. This seems intuitively reasonable.

The assumption that justifies the bias-adjusted bootstrap is that, for some monotone increasing function g, $(g(\hat{\theta}) - g(\theta))$ is distributed as $N(-z_0\sigma, \sigma^2)$ and $(g(\hat{\theta}^{(b)}) - g(\hat{\theta}))$ is also distributed as $N(-z_0\sigma, \sigma^2)$ for some z_0, σ^2. (The first distribution is from the probability law of the sample, and the second is the bootstrap distribution induced by resampling from the given sample.) Thus we have normality, and also equal biases and variances, for some transformation of θ. The transformation function g need not be known. This is an advantage in implementation, but a disadvantage in trying to decide whether the assumption holds. It is not known whether the bias-adjusted bootstrap is valid for our specific problem, but it performs relatively well in the simulations reported in Section 2.5.

The final version of the bootstrap that we will consider is the *bias-adjusted and accelerated bootstrap* of Efron and Tibshirani (1993). This is intended to allow for a possibility that the variances of $\hat{\theta}$ depends on θ, so that a bias-adjustment also requires a change in variance. This correction depends on some quantities defined in terms of the so-called jackknife values of $\hat{\theta}$. For $i = 1, \cdots, n$, let $\hat{\theta}_{(i)}$ be the value of the estimate based on all observations other than observation i; and let $\hat{\theta}_{(\bullet)} = n^{-1}\sum_{i=1}^{n}\hat{\theta}_{(i)}$ be the average of these values. Then the "acceleration" factor a is defined by:

$$a = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{(\bullet)} - \hat{\theta}_{(i)}\right)^3}{6\left(\sum_{i=1}^{n}\left(\hat{\theta}_{(\bullet)} - \hat{\theta}_{(i)}\right)^2\right)^{1.5}} \tag{2.9}$$

With z_0 and $z_{c/2}$ defined as above, define

$$b_{i1} = z_0 + \frac{\left(z_0 + z_{c/2}\right)}{\left(1 - a_i\left(z_0 + z_{c/2}\right)\right)}, \quad b_{i2} = z_0 + \frac{\left(z_0 - z_{c/2}\right)}{\left(1 - a_i\left(z_0 - z_{c/2}\right)\right)}. \tag{2.10}$$

Then the confidence interval is $[L_i, U_i]$ with $L_i = \hat{F}^{-1}(\Phi(b_{i1}))$ and $U_i = \hat{F}^{-1}(\Phi(b_{i2}))$. More discussion can be found in Efron and Tibshirani (1993, chapter 14).

It is important to note that there are cases in which the acceleration factor fails to be defined. This happens when all the jackknifed estimates are the same, which yields zero both for the numerator and for the denominator of the acceleration factor. For example, one firm could be so dominantly efficient in the industry that jackknifing the best firm (in our case, dropping one time dimensional observation) would not change the efficiency rank for the best firm. Also, with large T, the firms' efficiency ranking would not be affected by taking out one time period observation, so that it is more likely for the acceleration factor not to be defined. However, as N gets large, it is less likely for the acceleration factor not to be defined since it would be harder to have one specific firm uniformly as the best estimated firm with more firms in sample. In the following sections, when the acceleration factor is not defined, we do not accelerate the bias-adjusted bootstrap. After all, the bias-adjusted bootstrap is a special case of the bias-adjusted and accelerated bootstrap with the acceleration factor of zero.

2.4 A Simple Alternative to the Bootstrap

In this section we propose a simple parametric alternative to the bootstrap, and some related resampling procedures. We begin with the following simple observation. We wish to construct a confidence interval for $u_i^* = \alpha_{(N)} - \alpha_i$, or $r_i^* = \exp(-u_i^*)$. If we knew which firm was best - that is, if we knew the index (N) - we could construct a parametric confidence interval of the form:

$$(\hat{\alpha}_{(N)} - \hat{\alpha}_i) \pm (\text{critical value}) * (\text{standard error}), \tag{2.11}$$

where "critical value" would be the appropriate $c/2$ level critical value of the standard normal distribution, and "standard error" would be the square root of the quantity: estimated variance of $\hat{\alpha}_{(N)}$ + estimated variance of $\hat{\alpha}_i$ - 2*estimated covariance of $(\hat{\alpha}_{(N)}, \hat{\alpha}_i)$. This interval would be valid asymptotically as $T \to \infty$ with N fixed. In fact, if the v_{it} are i.i.d. normal and we use the critical value from the student-t distribution, this interval would be valid in finite samples as well.

The confidence interval (2.11) is infeasible because the identity of the best firm is unknown. However, we can construct the confidence interval:

$$(\hat{\alpha}_{[N]} - \hat{\alpha}_i) \pm (\text{critical value}) * (\text{standard error}), \tag{2.12}$$

where as before $\max_j \hat{\alpha}_j = \hat{\alpha}_{[N]}$. That is, we use a confidence interval that would be appropriate if (N) were known, and we simply pretend that $[N] = (N)$. That is, we pretend that we do know the identity of the best firm. This is our "simple parametric" confidence interval.

Two details should be noted. First, in calculating the standard error in (2.12), we evaluate $\text{Var}(\hat{\alpha}_{[N]})$ and $\text{Cov}(\hat{\alpha}_{[N]}, \hat{\alpha}_i)$ using the standard formulas that ignore the fact that the index $[N]$ is data-determined. That is, again we pretend that $[N] = (N)$ is known. Second, although $\alpha_{(N)} - \alpha_i \geq 0$, the lower bound of the confidence interval in (2.12) can be negative. If it is, set it to zero. This corresponds to setting the upper bound of the relative efficiency measure r_i^* to one.

The asymptotic ($T \to \infty$ with N fixed) validity of this procedure follows from the same argument that Hall, Härdle, and Simar (1995) used to show that $\max_j \hat{\alpha}_j$ is asymptotically normal. If there are no ties for $\max_j \alpha_j$, then as $T \to \infty$, $\text{P}([N] =$

$(N)) \to 1$. That is, with no ties, in the limit there is no uncertainty about the identity of the best firm.

An obvious implication of this argument is the following. For data sets in which there is substantial uncertainty about the identity of the best firm, the accuracy of either bootstrap intervals or our simple parametric intervals is doubtful.

The simple parametric intervals differ from bootstrap intervals in an important way that goes beyond parametric versus resampling methods. Consider the following resampling scheme, which could also be used to create a confidence interval for $u_i^* = \alpha_{(N)} - \alpha_i$, treating $(N) = [N]$ as known. Create bootstrap samples $b = 1, \cdots, B$ as above. For sample b, calculate

$$\hat{u}^{*(b)}_{i,max-best} = \hat{\alpha}^{(b)}_{[N]} - \hat{\alpha}^{(b)}_i \tag{2.13}$$

where $[N]$ is still the index such that $\hat{\alpha}_{[N]} = \max_j \hat{\alpha}_j$ *in the original sample.* Then create a percentile-interval from these quantities.

Note that the quantities $\hat{u}^{*(b)}_{i,max-best}$ differ from the bootstrap quantities

$$\hat{u}^{*(b)}_i = \max_j \hat{\alpha}^{(b)}_j - \hat{\alpha}^{(b)}_i, \tag{2.14}$$

as defined in Section 2.3. For the bootstrap quantities, there is a "max" in the original data to get $\hat{\alpha}_{[N]}$ and then there is another "max" in each bootstrap sample. That is, the bootstrap samples are deliberately analyzed in exactly the same way as the original sample was. In (2.13), there is still a "max" in the original sample, but in the bootstrap samples we maintain the identity of the "best" firm in the original samples. We will call this the "max-best bootstrap," although actually it is not really a bootstrap procedure at all. It is just a resampling scheme. Semantic issues aside, it is the "max-best" bootstrap that should be similar to our simple parametric procedure. Our motivation for discussing the "max-best" procedure is mostly to make clear why our simple parametric intervals may be expected to be rather different from percentile bootstrap intervals, when the identity of the best firms is in doubt.

As noted above, the "max" operator causes $\hat{\alpha}_{(N)}$ to be biased upward as an estimate of $\alpha_{(N)}$, and this causes an upward bias in $\hat{u}_i^*$ and a downward bias in $\hat{r}_i^* = \exp(-\hat{u}_i^*)$. The second "max" in (2.14) in the bootstrap samples causes additional bias. For this reason the percentile bootstrap intervals will tend to be seriously miscentered. Our simple parametric intervals, or "max-best" bootstrap intervals, do not contain the second source of bias and may be expected to be more accurate than percentile bootstrap intervals. Of course, precisely because they do not contain the second source of bias, the parametric or "max-best" intervals cannot be bias-adjusted. The bias-adjusted (or bias-adjusted and accelerated) bootstrap intervals described in the previous section use the bias at the bootstrap stage to correct the bias in the original estimates. The ability to do this is a potentially significant advantage of bootstrap methods.

2.5 Simulations

In this section we conduct Monte Carlo simulations to investigate the reliability of confidence intervals based on bootstrapping and on the alternative procedures described in the last section. We are interested in the coverage rates of the confidence intervals and the way that they are related to bias in estimation of efficiency levels. Results for other methods including the MLE can be found in Kim (1999).

The model is the basic panel data stochastic frontier model given in (2.1) above. However, we consider the model with no regressors so that we can concentrate our interest on the estimation of efficiencies without having to be concerned about the nature of the regressors. In practical cases, the regression parameters β are likely to be estimated so much more efficiently than the other parameters that treating them as known is not likely to make much difference.

Our data generating process is:

$$y_{it} = \alpha + v_{it} - u_i = \alpha_i + v_{it}, \quad i = 1, \cdots, N, \quad t = 1, \cdots, T, \tag{2.15}$$

in which the v_{it} are i.i.d. $N(0, \sigma_v^2)$ and the u_i are i.i.d. half-normal: that is, let $u_i = |\mathrm{u}_i|$ where $\mathrm{u}_i \sim N(0, \sigma_u^2)$. Since our point estimates and confidence intervals are based on the fixed effects estimates of $\alpha_1, \cdots, \alpha_N$, the distributional assumptions on v_{it} and u_i do not enter into the estimation procedure. They just define the data generation mechanism.

The parameter space is $(\alpha, \sigma_v^2, \sigma_u^2, N, T)$, but this can be reduced. Without loss of generality, we can fix α to any number, since a change in the constant term only shifts the estimated constant term by the same amount, without any effect on the bias and variance of any of the estimates. For simplicity, we fix the constant term equal to one.

We need two parameters to characterize the variance structure of model. It is natural to think in terms of σ_v^2 and σ_u^2. Alternatively, recognizing that σ_u^2 is the variance of the untruncated normal from which u is derived, not the variance of u, we can think instead in terms of σ_v^2 and $\mathrm{Var}(u)$, where $\mathrm{Var}(u) = \sigma_u^2(\pi - 2)/\pi$. However, we obtain more readily interpretable results if we think instead in terms of the size of total variance and the relative allocation of total variance between v and u. The total variance is defined as $\sigma_\epsilon^2 = \sigma_v^2 + \mathrm{Var}(u)$. Olson, Schmidt, and Waldman (1980) used $\lambda = \sigma_u/\sigma_v$ to represent the relative variance structure, so that their parametrization was in terms of σ_ϵ^2 and λ. Coelli (1995) used σ_ϵ^2 and either $\gamma = \sigma_u^2/(\sigma_v^2 + \sigma_u^2)$ or $\gamma^* = \mathrm{Var}(u)/(\sigma_v^2 + \mathrm{Var}(u))$. The choice between these two parameters is a matter of convenience. We decided to use γ^* due to its ease of interpretation, so that we use the parameters σ_ϵ^2 and γ^*. The reason this is a convenient parametrization (compared to the "obvious" choice of σ_v^2 and σ_u^2) is that, following Olson, Schmidt, and Waldman (1980), one can show that comparisons among the various estimators are not affected by σ_ϵ^2. The effect of multiplying σ_ϵ^2 by a factor of k holding γ^* constant, is as follows.

1. constant term: bias change by a factor of $\sqrt{k}$ and variance changes by a factor of k,

2. σ_v^2 and σ_u^2: bias changes by a factor of k and variance changes by a factor of k^2,

3. γ^* (or γ or λ): bias and variance are unaffected.

We set σ_ϵ^2 at 0.25 arbitrarily, so that the only parameters left to consider are (γ^*, N, T). We consider three values for γ^*, to include a case in which the variance of v dominates, a case in which the variance of u dominates, and an intermediate case. We take $\gamma^* = 0.1$, 0.5, and 0.9 to represent the above three cases. With $\sigma_\epsilon^2 = 0.25$, σ_v^2, $\text{Var}(u)$, and σ_u^2 are determined as follows for each value of γ^*.

1. $\gamma^* = 0.1$: $\sigma_v^2 = 0.225$, $\text{Var}(u) = 0.025$, $\sigma_u^2 = 0.069$,

2. $\gamma^* = 0.5$: $\sigma_v^2 = 0.125$, $\text{Var}(u) = 0.125$, $\sigma_u^2 = 0.344$,

3. $\gamma^* = 0.9$: $\sigma_v^2 = 0.025$, $\text{Var}(u) = 0.225$, $\sigma_u^2 = 0.619$.

Four values of N and T are considered. In order to investigate the effect of changing N, we fix $T = 10$ and consider N =10, 20, 50, and 100. Similarly, T is assigned the values of 10, 20, 50, and 100 while fixing $N = 10$. This is done for each different value of γ^*.

For each parameter configuration (γ^*, N, T), we perform $R = 300$ replications of the experiment. For each replication, we calculate the following:

1. The estimate of α, $\hat{\alpha} = \max_j \hat{\alpha}_j = \hat{\alpha}_{[N]}$.

2. The infeasible estimate of α, $\hat{\alpha}_{(N)}$.

3. The relative efficiency estimate, $\hat{u}_i^* = \hat{\alpha} - \hat{\alpha}_i$, for each $i = 1, 2, \cdots, N$.

4. The percentile bootstrap confidence interval for u_i^*, for each i.

5. The BC_a bootstrap confidence interval for u_i^*, for each i.

6. The simple parametric confidence interval (of Section 2.4) for u_i^*, for each i.

7. The "max-best" bootstrap confidence interval for u_i^*, for each i.

8. The infeasible parametric confidence interval (of Section 2.4) for u_i^*, for each i.

The bootstrap results were based on $B = 1000$ replications. Note that we did not consider the iterated bootstrap due to its computational demands.

We are primarily interested in the biases of the point estimates and the coverage rates of the confidence intervals. These biases and coverage rates are reported as averages over both the N firms (where relevant) and the R replications. In particular, the coverage rate of the confidence intervals is just the fraction of times that coverage occurs.

We begin the discussion of our results with Table 2.1. Three measures of biases are considered. $bias1 = \mathrm{E}(\hat{\alpha} - \alpha)$ is the bias in the overall constant, $bias2 = \mathrm{E}(\hat{u}_i^* - u_i)$ is the bias of the estimated relative inefficiency compared to true inefficiency, and $bias3 = \mathrm{E}(\hat{u}_i^* - u_i^*)$ is the bias of the estimated relative inefficiency compared to true relative inefficiency.

There are two different sources of $bias1$. These are easily understood in terms of the identity:

$$\hat{\alpha} - \alpha = (\hat{\alpha} - \alpha_{(N)}) - (\alpha - \alpha_{(N)}). \tag{2.16}$$

$bias1$ is $\mathrm{E}(\hat{\alpha} - \alpha)$. The first (and generally most important) source of this bias is $\mathrm{E}(\hat{\alpha} - \alpha_{(N)})$, which is positive. That is, $\hat{\alpha}$ is biased upward as an estimate of $\alpha_{(N)}$, because of the "max" operation that defines $\hat{\alpha} = \max_j \hat{\alpha}_j$. This bias increases with N, but decreases when T and/or γ^* increase. It disappears as $T \to \infty$ or $\gamma^* \to 1$. The second source of bias is that $\mathrm{E}(\alpha_{(N)}) < \alpha$, resulting in downward bias for $\hat{\alpha}$. This reflects the fact that $\alpha - \alpha_{(N)} = \min_j u_j \geq 0$. This bias disappears as $N \to \infty$. More generally, it decreases as N increases, and increases with γ^*, but does not depend on T. We see examples of both positive and negative bias in column (1) of Table 2.1. As expected, the largest positive bias occurs for large N and small T and γ^*, whereas negative bias (absolute value) increases for larger γ^* and T and smaller N.

The bias of $\hat{u}_i^*$ as an estimate of u_i is given in column (2) of Table 2.1. It is essentially the same as the bias of the overall constant term:

$$\begin{aligned} bias2 &= \mathrm{E}(\hat{u}_i^* - u_i) = \mathrm{E}\left((\hat{\alpha} - \hat{\alpha}_i) - (\alpha - \alpha_i)\right) = \mathrm{E}(\hat{\alpha} - \alpha) - \mathrm{E}(\hat{\alpha}_i - \alpha_i) \\ &= bias1 - \mathrm{E}(\hat{\alpha}_i - \alpha_i) \end{aligned} \tag{2.17}$$

and $\mathrm{E}(\hat{\alpha}_i - \alpha_i) = 0$.

The estimate $\hat{u}_i^*$ is perhaps more naturally viewed as an estimate of u_i^*. Column (3) gives the bias of $\hat{u}_i^*$ as an estimate of u_i^*:

$$\begin{aligned} bias3 &= \mathrm{E}(\hat{u}_i^* - u_i^*) = \mathrm{E}\Big((\hat{\alpha} - \hat{\alpha}_i) - (\alpha_{(N)} - \alpha_i)\Big) = \mathrm{E}(\hat{\alpha} - \alpha_{(N)}) - \mathrm{E}(\hat{\alpha}_i - \alpha_i) \\ &= \mathrm{E}(\hat{\alpha} - \alpha_{(N)}) > 0 \end{aligned} \tag{2.18}$$

since $\mathrm{E}(\hat{\alpha}_i - \alpha) = 0$. Note that *bias*3 is the first source of *bias*1, as described above and is always positive. In other words, $\hat{u}_i^*$ can overestimate or underestimate the absolute efficiency u_i, but (on average) it overestimates the relative efficiency u_i^*.

We now turn our attention to question of the accuracy of the various types of confidence intervals we have discussed. We present results for 90% confidence intervals for $r_i^* = \exp(-u_i^*)$, but the coverage rates would be exactly the same for the corresponding confidence intervals for u_i^*. We are primarily interested in the coverage rates of the intervals, and the proportions of observations that fall below the lower bound and above the upper bound. The reason we present intervals for r_i^* (rather than u_i^*) is that it is bounded between zero and one, and so the average width of the intervals is easier to interpret.

Table 2.2 gives the results for the infeasible parametric intervals based on equation (2.11) of Section 2.4. The coverage rates of these intervals are very close to 0.90, as they should be. These intervals are infeasible in practice, since they depend on knowledge of the identity of the best firm, but they illustrate two points. First, for obvious reasons, the intervals are narrower when T is large and when γ^* is large (that is, when the variance of inefficiency is large relative to the variance of noise). The number of firms, N, is not really relevant if we know which one is best. Second, and more fundamentally, there is no difficulty in constructing accurate confidence intervals for technical efficiency if we know which firm is best. All of the problems that we will see with the accuracy of feasible intervals are due to not knowing with certainty which firm is best.

Table 2.3 gives the results for the percentile bootstrap and BC_a bootstrap confidence intervals. Consider first the percentile bootstrap. Its coverage rate is virtually always less than the nominal level of 90%. The problem is that the intervals are not centered on the true values, due to the bias problem discussed above. (The upward bias of $\hat{\alpha}$ as an estimate of $\alpha(N)$ corresponds to an upward bias in $\hat{u}_i^*$ and a downward bias in $\hat{r}_i^*$. Thus too many r_i^* lie above the upper bound of the confidence intervals.) Theoretically, the intervals should be accurate in the limit (as $T \to \infty$ with N fixed), if there are no ties for $\max_j \alpha_j$, and so the validity of the percentile bootstrap depends on large T. The bias problem is small when we have large T and γ^* and small N, and the coverage probability reaches almost 0.9 for these cases, but it falls in the opposite cases where the bias is big. The width of the intervals decreases as T or γ^* increases.

However, the intervals get narrower with larger N, while the bias increases as N increases. This explains why the coverage probabilities of the percentile intervals fall rapidly as N increases.

The results in Table 2.3 indicate that the BC_a intervals provide better coverage rates than the uncorrected percentile intervals, but with the same pattern. They are more accurate when T and γ^* are large and when N is small. When T and γ^* is small or N is large, there are very considerable improvements over the uncorrected percentile intervals, even though the BC_a intervals do not succeed entirely in yielding correct coverage rates.

The bias corrected confidence intervals are obtained by shifting the bootstrap distribution by approximately twice the estimated bias in the bootstrapping stage. If on average $(\max_j \hat{\alpha}_j^{(b)} - \max_j \hat{\alpha}_j)$ were the same as $(\max_j \hat{\alpha}_j - \max_j \alpha_j)$, we would expect a properly centered interval with a coverage rate of approximately 0.9 after the bias is corrected. In our simulations, however, only some part of the bias gets corrected. Some evidence on this point is given in Table 2.4, which shows the average of $\max_j \alpha_j$, $\max_j \hat{\alpha}_j$, and $\max_j \hat{\alpha}_j^{(b)}$ over different values of N, T, and γ^*. The fourth column in the table shows the average bias in the fixed effects estimates of $\max_j \alpha_j$, and the last column shows the average bias in the bootstrap estimates. We see that $(\max_j \hat{\alpha}_j^{(b)} - \max_j \hat{\alpha}_j)$ is always smaller than $(\max_j \hat{\alpha}_j - \max_j \alpha_j)$ and the difference is substantial when γ^* is small and N is large. As a result, the bias correction is incomplete especially when γ^* is small and N is large. However, the bias correction is always in the right direction, and this explains why BC_a intervals are better than the percentile intervals.

Table 2.5 gives the results for the feasible parametric intervals based on equation (2.12) of Section 2.4, and for the "max-best" bootstrap. We expect the feasible parametric intervals and those from the "max-best" bootstrap to give similar results, and they do. The parametric intervals have slightly better coverage rates, because they are wider, but the differences are quite small. As a result we will limit our further discussion to the feasible parametric intervals.

The feasible parametric intervals are clearly more accurate than the percentile bootstrap intervals. This is especially true in the worst cases. For example, for $N = 100$, $T = 10$ and $\gamma^* = 0.1$, compare coverage rates of 0.195 for the percentile bootstrap and 0.663 for the parametric intervals. The parametric intervals are wider and they are better centered, both of which imply higher coverage rates. To understand the point about better centering, recall the discussion of bias in Section 2.4. The parametric intervals have one level of bias ($\hat{\alpha}$ is a biased estimate of $\alpha_{(N)}$) whereas the percentile bootstrap has two ($\hat{\alpha}$ is a biased estimate of $\alpha_{(N)}$, *and* $\max_j \hat{\alpha}_j^{(b)}$ is a biased "estimator" of $\hat{\alpha}$).

A more interesting comparison is the feasible parametric intervals versus the BC_a intervals. The feasible parametric intervals generally but not always have better coverage rates than the BC_a intervals. This is because they are wider. The cases in which the BC_a intervals have better coverage rates than the parametric intervals are cases in which T, N and γ^* are all small. These are cases of considerable bias but not the cases with the most bias (see Table 2.4), which would be cases in which T and γ^* are small but N is big. Overall, it is hard to say whether the parametric or BC_a intervals are better, because there is a conflict between our desire for confidence intervals to cover with correct probability and our desire for them not to be wide.

Our last set of simulations is designed to consider cases in which the identity of the best firm is clear. Here we set out one u_i at the 0.05 quantile of the half normal distribution, while the other $(N-1)$ are set at equally spaced points between the 0.75 and 0.95 quantiles, inclusive. These u_i are then held fixed across replications of the experiment. The only randomness therefore comes from the stochastic error v. Since the identity of the best firm should be clear, the bias caused by the max operator should be minimal. Table 2.6 gives the bias of the fixed effects estimate, and is of the same format as Table 2.1. Recall that *bias*3 is the component of the bias caused by the max operator (see equation (2.18) above) and should be small when the identity of the best firm is clear. We can see that *bias*3 in Table 2.6 is indeed much smaller than in Table 2.1.

Correspondingly, we expect the various bootstrap and parametric intervals to be more accurate in the current cases than in the previous ones. Table 2.7 gives the results for the percentile bootstrap, the BC_a bootstrap, and the feasible parametric intervals. Clearly the intervals are much more reliable now than they were in the previous cases for which results were reported in Tables 2.3 and 2.5. Note in particular that the percentile bootstrap now does pretty well in all cases except the least favorable (small T and γ^*, and large N). The BC_a bootstrap is now usually worse than the percentile bootstrap. It is counterproductive to try to correct for bias when there is little or no bias. The parametric intervals often cover too often, rather than too seldom, and again this is a reflection of the intervals being wider than the bootstrap intervals.

The overall conclusions we draw from our simulations are straightforward. If it is clear from the data which firm is best, all of the methods of constructing confidence intervals work fairly well. There is no need to consider more complicated procedures than the percentile bootstrap. The parametric intervals are also reliable, but they may be wider than necessary. Conversely, if it is not clear from the data which firm is best, none of the methods of constructing confidence intervals are very reliable. The percentile bootstrap is particularly bad. The BC_a bootstrap intervals or the parametric intervals are probably preferred.

2.6 Empirical Results

We now apply the procedures described above to two well-known data sets. These data sets were chosen to have rather different characteristics. The first data set consists of $N = 171$ Indonesian rice farms observed for $T = 6$ growing seasons. For this data set, the variance of stochastic noise (v) is large relative to the variability in u (Var(u)): that is, $\hat{\gamma}^* = 0.222$ with $\hat{\sigma}_\epsilon^2 = 0.138$. Inference on inefficiencies will be very imprecise because T is small, $\hat{\gamma}^*$ is small and N is large. The second data set consists of $N = 10$ Texas utilities observed for $T = 18$ years. For this data set, σ_v^2 is small relative to Var(u): $\hat{\gamma}^* = 0.700$ with $\hat{\sigma}_\epsilon^2 = 0.010$. In this case we can estimate inefficiencies much more precisely because T and γ^* are larger, and N is smaller. We will see that the precision of the estimates will differ across these data sets, and that choice of technique matters more where precision is low. A more detailed analysis of these data, including Bayesian results and results for multiple and marginal comparisons with the best, can be found in Kim and Schmidt (1999).

2.6.1 Indonesian Rice Farms

These data are due to Erwidodo (1990) and have been analyzed subsequently by Lee (1991), Lee and Schmidt (1993), Horrace and Schmidt (1996), Horrace and Schmidt (2000) and others. There are $N = 171$ rice farms and $T = 6$ six-month growing seasons. Output is rice in kilograms and inputs are land in hectares, labor in hours, seed in kilograms and two types of fertilizer (urea in kilograms and phosphate in kilograms). The functional form is Cobb-Douglas with some dummy variables added for region, seasonality for dry or wet season, the use of pesticide and seed types for high yield or traditional or mixed. For a complete discussion of the data, see Erwidodo (1990).

The estimated regression parameters are given in Horrace and Schmidt (1996) and we will not repeat them here. Instead we will give point estimates of efficiencies and 90% confidence intervals for these efficiencies. There are 171 firms and so we report results for the three firms (164, 118, and 163) that are most efficient; for the firms at the $75^{(th)}$ percentile (31), $50^{(th)}$ percentile (15) and $25^{(th)}$ percentile (16) of the efficiency distribution; and for the two worst firms (117, 45). All of these rankings are according to fixed effects estimates.

We begin with Table 2.8. It gives the fixed effects point estimates and the lower and upper bounds of the 90% parametric confidence intervals. For the purpose of comparison we also give the point estimates and the lower and upper bound of the 90% confidence intervals for the MLE based on the assumption that inefficiency has a half-normal distribution. See Horrace and Schmidt (1996) for the details of calculations for the MLE.

The estimated efficiency levels based on the fixed effects estimates are rather low. They are certainly much smaller than the MLE estimates. This is presumably due to bias in the fixed effects estimates, as discussed previously. This data set has characteristics that should make the bias problem severe: N is large; the α_i are estimated imprecisely because σ_v^2 is large and T is small; and there are near ties for $\max_j \alpha_j$ because σ_u^2 is small.

Table 2.9 gives 90% confidence intervals based on the percentile bootstrap, the BC_a bootstrap, and the iterated bootstrap, as well as the (feasible) parametric intervals and the "max-best" bootstrap intervals. The bootstrap results are based on 1000 replications, and in the case of the iterated bootstrap each second-level bootstrap is also based on 1000 replications.

There is some similarity between the intervals from different methods, but there are also some interesting comparisons to make. The percentile bootstrap intervals are clearly closest to zero (i.e. they would indicate the lowest levels of efficiency). This is presumably a reflection of bias. Note, for example, that the midpoints of these intervals are clearly less than the fixed effects estimate (which is itself biased toward zero). For the reasons given above, we do not regard these intervals as trustworthy for this data set. The iterated bootstrap intervals are centered similarly to the percentile bootstrap but are wider. The BC_a intervals are an upward shift (in the direction of higher efficiency) of the percentile intervals and might be a good choice for this data set. The parametric intervals are also an upward shift of the percentile intervals, though not by as much as the BC_a intervals. They are wider than the BC_a intervals, and in fact they are about as wide as the iterated bootstrap intervals. They are another possible good choice for this data set; in a sense they are conservative choice. The "max-best" bootstrap intervals are similar to the parametric intervals and are therefore another possible good choice.

2.6.2 Texas Utilities

In this section, we consider the Texas utility data of Kumbhakar (1996), which was also analyzed by Horrace and Schmidt (1996) and Horrace and Schmidt (2000). As in the previous section, we will estimate a Cobb-Douglas production function, whereas Kumbhakar (1996) estimated a cost function. The data contain information on output and inputs of 10 privately owned Texas electric utilities for 18 years from 1966 to 1983. Output is electric power generated, and input measures on labor, capital and fuel are derived from dividing expenditures on each input by its price. For more details on the data see Kumbhakar (1996).

Table 2.10 gives the fixed effects point estimates, the 90% parametric intervals, and the MLE point estimates and 90% confidence intervals. The format is the same as that of Table 2.8, except that now we can report the results for all of the firms. Table 2.11 gives the 90% confidence intervals for the same set of procedures as before, and it is of the same format as Table 2.9, except that results are given for all firms.

Compared to the previous data set, we estimate the intercepts α_i much more precisely, because T is larger and σ_v^2 is smaller. For this reason, and also because N is smaller, we expect there not to be a severe finite sample bias problem in the fixed effects estimates, and we expect that the choice of technique will not matter as much.

The MLE estimated efficiencies are larger than those based on fixed effects (except for the "best" firm), but the difference is not nearly as large as for the previous data set. Similarly, the MLE confidence intervals are narrower than the parametric intervals, but not by nearly as much as in Table 2.8. A distributional assumption is much less valuable in the present case. In fact, the accuracy of the MLE intervals is now suspect, because we have only 10 firms, and the asymptotic justification for the MLE requires large N.

In Table 2.11, we can see that the parametric intervals and all of the bootstrapping intervals are quite similar. The bias problem is apparently negligible for this data set, and correspondingly our faith in the accuracy of these intervals is relatively strong.

We can compare the features of this data set with the setup of our simulation. One of the parametric configurations in our simulation had $N = 10$, $T = 20$, and $\gamma^* = 0.5$, which matches these data quite well. In that case the coverage rates of the various confidence intervals were in the range of 0.87 to 0.88, which are obviously close to 0.90.

A technical detail worth noting is that the acceleration factor in the BC_a bootstrap was undefined and was therefore set equal to zero. This is further evidence that there was very little bias in estimation.

2.7 Conclusions

In this chapter we have provided a survey of the use of bootstrapping to construct confidence intervals for efficiency measures. We discussed several versions of the bootstrap, including the percentile bootstrap, the iterated bootstrap, and the bias-adjusted and accelerated bootstrap. In stochastic frontier models, these methods can be applied to the fixed effects estimates, yielding inferences that are correct asymptotically as $T \to \infty$ with N fixed.

We have proposed a simple parametric method of constructing confidence intervals. It uses standard methods and simply acts as if the identity of the best firm is known. We also proposed a resampling scheme, the "max-best" bootstrap, which ought to yield confidence intervals similar to the parametric intervals. These procedures are valid under the same conditions that the bootstrap methods are valid, namely, as $T \to \infty$ with N fixed, and provided that there is a unique best firm.

The main problem that we encounter is the upward bias in the fixed effects estimate of the frontier, which translates into a downward bias for the estimated efficiencies. The bias is large when T is small, N is large, and/or statistical noise is large relative to the variation in the frontier. These are exactly the same circumstances in which the identity of the best firm is uncertain, and so it is fair to say that bias is a problem when the identity of the best firm is in question.

Our simulation results show that the percentile bootstrap is seriously inaccurate when the bias problem exists, that is, when the identity of the best firm is not clear. The percentile bootstrap intervals are miscentered because the bias in the original estimates is compounded by similar "bias" in the bootstrap estimates. Our parametric intervals, or our "max-best" bootstrap intervals, avoid the second source of bias, are more reliable than the percentile bootstrap intervals. The bias corrected and accelerated (BC_a) bootstrap makes a bias correction based on the "bias" in the second round, and these intervals are also more reliable than the percentile bootstrap intervals. Comparing the parametric intervals and the BC_a intervals, neither clearly dominates the other. The parametric intervals are more conservative.

A negative conclusion of the simulations is that none of the methods of constructing confidence intervals based on the fixed effects estimates is very reliable if the identity of the best firm is in serious doubt. In such cases it may be worthwhile to consider assuming a distribution for technical inefficiency and using MLE.

We performed an empirical analysis of two data sets, one of which had characteristics very unfavorable to the bootstrap (large N, small T, and large variance of noise). In this case there was evidence of bias, and the bootstrap intervals were both unreliable and too wide to be informative. Our other data set had more favorable characteristics, and the empirical analysis yielded results that were quite precise and seemingly sensible. Hence, as in the simulations, a major lesson is that the reliability of inference on efficiencies can be judged based on observable features of the data.

2.8 Output Tables

Table 2.1: Biases of Fixed Effects Estimates

T	γ^*	N	*bias*1 $E(\hat{\alpha} - \alpha)$ (1)	*bias*2 $E(\hat{u}_i^* - u_i)$ (2)	*bias*3 $E(\hat{u}_i^* - u_i^*)$ (3)
10	0.1	10	0.103	0.105	0.133
10	0.1	20	0.153	0.155	0.169
10	0.1	50	0.234	0.235	0.241
10	0.1	100	0.276	0.274	0.277
10	0.5	10	-0.009	-0.008	0.055
10	0.5	20	0.045	0.046	0.078
10	0.5	50	0.119	0.119	0.132
10	0.5	100	0.153	0.152	0.159
10	0.9	10	-0.076	-0.075	0.010
10	0.9	20	-0.028	-0.028	0.016
10	0.9	50	0.018	0.018	0.035
10	0.9	100	0.039	0.039	0.049
10	0.1	10	0.103	0.105	0.133
20	0.1	10	0.049	0.046	0.078
50	0.1	10	0.006	0.005	0.035
100	0.1	10	-0.007	-0.007	0.021
10	0.5	10	-0.009	-0.008	0.055
20	0.5	10	-0.038	-0.041	0.030
50	0.5	10	-0.054	-0.054	0.013
100	0.5	10	-0.058	-0.058	0.004
10	0.9	10	-0.076	-0.075	0.010
20	0.9	10	-0.090	-0.091	0.004
50	0.9	10	-0.087	-0.088	0.002
100	0.9	10	-0.084	-0.085	0.000

Table 2.2: 90% Confidence Intervals for Relative Efficiency (r_i^*)

T	γ^*	N	Infeasible Parametric width	$P_{(<lb)}$	$P_{(>ub)}$	cover
10	0.1	10	0.551	0.057	0.037	0.905
10	0.1	20	0.564	0.038	0.052	0.910
10	0.1	50	0.599	0.059	0.043	0.898
10	0.1	100	0.594	0.048	0.049	0.903
10	0.5	10	0.326	0.057	0.037	0.905
10	0.5	20	0.335	0.038	0.052	0.910
10	0.5	50	0.352	0.059	0.043	0.898
10	0.5	100	0.351	0.048	0.049	0.903
10	0.9	10	0.127	0.057	0.037	0.905
10	0.9	20	0.131	0.038	0.052	0.910
10	0.9	50	0.136	0.059	0.043	0.898
10	0.9	100	0.137	0.048	0.049	0.903
10	0.1	10	0.551	0.057	0.037	0.905
20	0.1	10	0.379	0.044	0.045	0.910
50	0.1	10	0.236	0.038	0.043	0.919
100	0.1	10	0.167	0.050	0.038	0.912
10	0.5	10	0.326	0.057	0.037	0.905
20	0.5	10	0.228	0.044	0.045	0.910
50	0.5	10	0.143	0.038	0.043	0.919
100	0.5	10	0.101	0.050	0.038	0.912
10	0.9	10	0.127	0.057	0.037	0.905
20	0.9	10	0.090	0.044	0.045	0.910
50	0.9	10	0.057	0.038	0.043	0.919
100	0.9	10	0.040	0.050	0.038	0.912

Table 2.3: 90% Confidence Intervals for Relative Efficiency (r_i^*)

T	γ^*	N	Percentile Bootstrap width	$P_{(<lb)}$	$P_{(>ub)}$	cover	BC_a Bootstrap width	$P_{(<lb)}$	$P_{(>ub)}$	cover
10	0.1	10	0.354	0.001	0.289	0.709	0.336	0.015	0.130	0.855
10	0.1	20	0.346	0.000	0.447	0.553	0.328	0.015	0.164	0.821
10	0.1	50	0.323	0.000	0.676	0.324	0.320	0.008	0.275	0.717
10	0.1	100	0.305	0.000	0.805	0.195	0.306	0.007	0.341	0.652
10	0.5	10	0.248	0.015	0.157	0.829	0.252	0.044	0.092	0.864
10	0.5	20	0.245	0.003	0.235	0.762	0.243	0.041	0.108	0.851
10	0.5	50	0.230	0.001	0.448	0.552	0.232	0.023	0.184	0.794
10	0.5	100	0.219	0.000	0.603	0.397	0.221	0.018	0.229	0.753
10	0.9	10	0.111	0.040	0.084	0.876	0.115	0.057	0.081	0.861
10	0.9	20	0.112	0.018	0.116	0.867	0.113	0.061	0.084	0.855
10	0.9	50	0.108	0.005	0.234	0.761	0.108	0.048	0.115	0.837
10	0.9	100	0.105	0.002	0.363	0.636	0.104	0.037	0.150	0.813
10	0.1	10	0.354	0.001	0.289	0.709	0.336	0.015	0.130	0.855
20	0.1	10	0.282	0.002	0.225	0.773	0.267	0.027	0.099	0.874
50	0.1	10	0.197	0.005	0.152	0.843	0.190	0.036	0.079	0.885
100	0.1	10	0.145	0.008	0.131	0.861	0.144	0.034	0.072	0.895
10	0.5	10	0.248	0.015	0.157	0.829	0.252	0.044	0.092	0.864
20	0.5	10	0.192	0.014	0.113	0.872	0.196	0.044	0.088	0.868
50	0.5	10	0.131	0.018	0.085	0.897	0.136	0.044	0.074	0.882
100	0.5	10	0.094	0.028	0.070	0.902	0.097	0.061	0.074	0.866
10	0.9	10	0.111	0.040	0.084	0.876	0.115	0.057	0.081	0.861
20	0.9	10	0.083	0.031	0.068	0.901	0.085	0.059	0.083	0.858
50	0.9	10	0.055	0.031	0.063	0.906	0.056	0.044	0.076	0.880
100	0.9	10	0.039	0.045	0.047	0.908	0.040	0.053	0.069	0.878

Table 2.4: Bias Correction in the BC_a Bootstrap Intervals

T	γ^*	N	$\max_j \alpha_j$ (1)	$\max_j \hat{\alpha}_j$ (2)	$\max_j \hat{\alpha}_j^{(b)}$ (3)	(2)-(1)	(3)-(2)
10	0.1	10	0.972	1.103	1.175	0.132	0.072
50	0.1	10	0.970	1.006	1.034	0.036	0.029
10	0.1	50	0.994	1.234	1.342	0.240	0.108
10	0.5	10	0.937	0.991	1.027	0.054	0.037
50	0.5	10	0.933	0.946	0.957	0.013	0.011
10	0.5	50	0.988	1.119	1.183	0.131	0.064
10	0.9	10	0.915	0.924	0.933	0.009	0.008
50	0.9	10	0.910	0.913	0.915	0.003	0.002
10	0.9	50	0.983	1.018	1.039	0.035	0.021

Table 2.5: 90% Confidence Intervals for Relative Efficiency (r_i^*)

T	γ^*	N	Feasible Parametric width	$P_{(<lb)}$	$P_{(>ub)}$	cover	"Max-best" Bootstrap width	$P_{(<lb)}$	$P_{(>ub)}$	cover
10	0.1	10	0.463	0.071	0.113	0.816	0.433	0.072	0.138	0.790
10	0.1	20	0.471	0.039	0.144	0.817	0.444	0.039	0.173	0.787
10	0.1	50	0.455	0.018	0.258	0.724	0.429	0.018	0.295	0.687
10	0.1	100	0.445	0.010	0.327	0.663	0.420	0.010	0.374	0.617
10	0.5	10	0.301	0.058	0.070	0.872	0.282	0.060	0.089	0.852
10	0.5	20	0.308	0.033	0.085	0.881	0.290	0.035	0.107	0.858
10	0.5	50	0.301	0.017	0.163	0.820	0.285	0.017	0.190	0.793
10	0.5	100	0.298	0.009	0.215	0.776	0.281	0.009	0.248	0.743
10	0.9	10	0.124	0.055	0.049	0.896	0.117	0.059	0.061	0.880
10	0.9	20	0.129	0.032	0.061	0.907	0.122	0.039	0.075	0.886
10	0.9	50	0.130	0.019	0.096	0.885	0.123	0.021	0.116	0.864
10	0.9	100	0.130	0.010	0.132	0.857	0.123	0.011	0.156	0.833
10	0.1	10	0.463	0.071	0.113	0.816	0.433	0.072	0.138	0.790
20	0.1	10	0.344	0.067	0.090	0.844	0.333	0.067	0.099	0.834
50	0.1	10	0.227	0.053	0.073	0.874	0.224	0.053	0.078	0.869
100	0.1	10	0.162	0.053	0.067	0.880	0.161	0.053	0.068	0.879
10	0.5	10	0.301	0.058	0.070	0.872	0.282	0.060	0.089	0.852
20	0.5	10	0.219	0.051	0.065	0.884	0.212	0.053	0.070	0.877
50	0.5	10	0.141	0.042	0.055	0.904	0.139	0.042	0.058	0.900
100	0.5	10	0.100	0.050	0.049	0.901	0.100	0.052	0.051	0.897
10	0.9	10	0.124	0.055	0.049	0.896	0.117	0.059	0.061	0.880
20	0.9	10	0.089	0.048	0.051	0.901	0.087	0.052	0.056	0.893
50	0.9	10	0.057	0.038	0.048	0.914	0.056	0.041	0.052	0.907
100	0.9	10	0.040	0.052	0.041	0.908	0.040	0.055	0.043	0.901

Table 2.6: Biases of Fixed Effects Estimates (Case that u_i are fixed over replications)

T	γ^*	N	*bias*1 $E(\hat{\alpha} - \alpha)$ (1)	*bias*2 $E(\hat{u}_i^* - u_i)$ (2)	*bias*3 $E(\hat{u}_i^* - u_i^*)$ (3)
10	0.1	10	0.010	0.013	0.029
10	0.1	20	0.006	0.006	0.023
10	0.1	50	0.045	0.046	0.062
10	0.1	100	0.061	0.061	0.078
10	0.5	10	-0.035	-0.032	0.004
10	0.5	20	-0.049	-0.049	-0.012
10	0.5	50	-0.029	-0.028	0.008
10	0.5	100	-0.042	-0.041	-0.005
10	0.9	10	-0.048	-0.047	0.002
10	0.9	20	-0.055	-0.055	-0.006
10	0.9	50	-0.046	-0.046	0.004
10	0.9	100	-0.052	-0.051	-0.002
10	0.1	10	0.010	0.013	0.029
20	0.1	10	-0.021	-0.021	-0.004
50	0.1	10	-0.019	-0.018	-0.001
100	0.1	10	-0.019	-0.019	-0.002
10	0.5	10	-0.035	-0.032	0.004
20	0.5	10	-0.042	-0.042	-0.005
50	0.5	10	-0.039	-0.038	-0.001
100	0.5	10	-0.039	-0.039	-0.002
10	0.9	10	-0.048	-0.047	0.002
20	0.9	10	-0.052	-0.052	-0.002
50	0.9	10	-0.050	-0.050	0.000
100	0.9	10	-0.050	-0.050	-0.001

Table 2.7: 90% Confidence Intervals for Relative Efficiency (r_i^*) (Case that u_i are fixed across replications)

			Percentile Bootstrap				BC_a Bootstrap				Feasible Parametric			
T	γ^*	N	width	$P_{(<lb)}$	$P_{(>ub)}$	cover	width	$P_{(<lb)}$	$P_{(>ub)}$	cover	width	$P_{(<lb)}$	$P_{(>ub)}$	cover
10	0.1	10	0.354	0.020	0.095	0.885	0.354	0.100	0.071	0.829	0.437	0.031	0.046	0.923
10	0.1	20	0.354	0.013	0.122	0.865	0.343	0.131	0.058	0.811	0.466	0.026	0.040	0.934
10	0.1	50	0.334	0.003	0.224	0.772	0.333	0.120	0.078	0.803	0.460	0.012	0.062	0.926
10	0.1	100	0.319	0.001	0.305	0.693	0.322	0.119	0.079	0.802	0.458	0.007	0.063	0.930
10	0.5	10	0.196	0.053	0.057	0.890	0.197	0.055	0.081	0.864	0.210	0.043	0.043	0.913
10	0.5	20	0.213	0.065	0.045	0.890	0.213	0.068	0.068	0.865	0.226	0.052	0.034	0.913
10	0.5	50	0.215	0.051	0.065	0.884	0.215	0.053	0.076	0.871	0.228	0.042	0.055	0.902
10	0.5	100	0.219	0.069	0.057	0.875	0.219	0.075	0.064	0.861	0.234	0.060	0.046	0.894
10	0.9	10	0.066	0.054	0.057	0.890	0.066	0.055	0.054	0.892	0.070	0.043	0.043	0.913
10	0.9	20	0.071	0.065	0.045	0.890	0.071	0.066	0.045	0.889	0.075	0.052	0.034	0.913
10	0.9	50	0.072	0.051	0.065	0.884	0.072	0.052	0.066	0.883	0.076	0.042	0.055	0.902
10	0.9	100	0.074	0.069	0.057	0.874	0.074	0.070	0.058	0.872	0.078	0.060	0.046	0.894
10	0.1	10	0.354	0.020	0.095	0.885	0.354	0.100	0.071	0.829	0.437	0.031	0.046	0.923
20	0.1	10	0.282	0.035	0.052	0.913	0.291	0.101	0.070	0.829	0.315	0.040	0.038	0.922
50	0.1	10	0.191	0.059	0.054	0.887	0.196	0.066	0.113	0.820	0.196	0.057	0.053	0.890
100	0.1	10	0.137	0.050	0.038	0.911	0.137	0.050	0.066	0.884	0.138	0.048	0.039	0.913
10	0.5	10	0.196	0.053	0.057	0.890	0.197	0.055	0.081	0.864	0.210	0.043	0.043	0.913
20	0.5	10	0.144	0.051	0.041	0.908	0.144	0.051	0.043	0.905	0.149	0.045	0.038	0.917
50	0.5	10	0.091	0.060	0.054	0.886	0.091	0.062	0.056	0.883	0.093	0.057	0.053	0.890
100	0.5	10	0.065	0.050	0.038	0.911	0.065	0.049	0.039	0.912	0.065	0.048	0.039	0.913
10	0.9	10	0.066	0.054	0.057	0.890	0.066	0.055	0.054	0.892	0.070	0.043	0.043	0.913
20	0.9	10	0.048	0.051	0.041	0.908	0.048	0.051	0.043	0.906	0.050	0.045	0.038	0.917
50	0.9	10	0.031	0.060	0.054	0.886	0.031	0.062	0.056	0.883	0.031	0.057	0.053	0.890
100	0.9	10	0.022	0.050	0.038	0.911	0.022	0.049	0.039	0.912	0.022	0.048	0.039	0.913

Table 2.8: Estimated Efficiencies and 90% Confidence Intervals: Indonesian Rice Farms

	Fixed Effects			MLE		
Firm No.	Point Estimate	LB	UB	Point Estimate	LB	UB
164	1.000	1.000	1.000	0.964	0.903	0.998
118	0.933	0.682	1.000	0.964	0.902	0.998
⋮	⋮	⋮	⋮	⋮	⋮	⋮
31	0.620	0.447	0.859	0.924	0.823	0.994
⋮	⋮	⋮	⋮	⋮	⋮	⋮
15	0.554	0.403	0.762	0.923	0.792	0.990
⋮	⋮	⋮	⋮	⋮	⋮	⋮
16	0.501	0.362	0.694	0.845	0.725	0.969
⋮	⋮	⋮	⋮	⋮	⋮	⋮
117	0.380	0.275	0.524	0.773	0.658	0.907
45	0.366	0.266	0.504	0.774	0.659	0.908

Table 2.9: 90% Confidence Intervals: Indonesian Rice Farms

Firm	FE	Percentile Bootstrap		BC_a Bootstrap		Iterated Bootstrap		Parametric Interval		"Max-best" Bootstrap	
No.	Est.	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
164	1.000	0.742	1.000	0.807	1.000	0.663	1.000	1.000	1.000	1.000	1.000
118	0.933	0.671	1.000	0.775	1.000	0.579	1.000	0.682	1.000	0.687	1.000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
31	0.620	0.446	0.750	0.512	0.820	0.393	0.807	0.447	0.859	0.465	0.844
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
15	0.554	0.400	0.638	0.477	0.720	0.361	0.696	0.403	0.762	0.417	0.734
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
16	0.501	0.358	0.583	0.421	0.649	0.316	0.639	0.362	0.694	0.370	0.674
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
117	0.380	0.274	0.447	0.320	0.508	0.252	0.499	0.275	0.524	0.286	0.515
45	0.366	0.267	0.424	0.309	0.497	0.237	0.467	0.266	0.504	0.279	0.499

Table 2.10: Estimated Efficiencies and 90% Confidence Intervals: Texas Utilities

Firm No.	Fixed Effects Point Estimate	LB	UB	MLE Point Estimate	LB	UB
5	1.000	1.000	1.000	0.987	0.971	0.999
3	0.916	0.823	1.000	0.978	0.959	0.996
10	0.861	0.786	0.943	0.908	0.889	0.927
1	0.835	0.784	0.889	0.864	0.846	0.882
8	0.820	0.773	0.869	0.846	0.828	0.864
9	0.806	0.766	0.848	0.826	0.809	0.843
2	0.801	0.749	0.855	0.831	0.814	0.848
7	0.786	0.732	0.844	0.817	0.800	0.834
6	0.785	0.730	0.845	0.820	0.803	0.837
4	0.762	0.719	0.808	0.786	0.770	0.801

Table 2.11: 90% Confidence Intervals: Texas Utilities

Firm	FE	Percentile Bootstrap		BC_a Bootstrap		Iterated Bootstrap		Parametric Interval		"Max-best" Bootstrap	
No.	Est.	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
5	1.000	0.988	1.000	0.988	1.000	0.997	1.000	1.000	1.000	1.000	1.000
3	0.916	0.827	1.000	0.823	1.000	0.835	1.000	0.823	1.000	0.827	1.000
10	0.861	0.793	0.924	0.789	0.922	0.796	0.920	0.786	0.943	0.793	0.934
1	0.835	0.788	0.877	0.783	0.874	0.792	0.874	0.784	0.889	0.788	0.887
8	0.820	0.777	0.859	0.776	0.859	0.781	0.857	0.773	0.869	0.777	0.867
9	0.806	0.769	0.841	0.771	0.843	0.775	0.838	0.766	0.848	0.769	0.846
2	0.801	0.753	0.842	0.753	0.842	0.758	0.839	0.749	0.855	0.753	0.851
7	0.786	0.736	0.830	0.736	0.830	0.739	0.827	0.732	0.844	0.736	0.837
6	0.785	0.732	0.832	0.730	0.831	0.737	0.828	0.730	0.845	0.732	0.841
4	0.762	0.720	0.798	0.718	0.797	0.726	0.795	0.719	0.808	0.720	0.804

Chapter 3

Indicator KPSS with a Time Trend

3.1 Introduction

In this chapter, we propose a statistic to test whether a time series is stationary, and we allow for a time trend. A standard test for stationarity is the KPSS test by Kwiatkowski, Phillips, Schmidt, and Shin (1992). The KPSS test, $\hat{\eta}_\mu$ uses the scaled sum of squares of cumulations of demeaned data with a long-run variance estimate in the denominator. A deterministic trend can be allowed in the test of trend-stationarity in which the demeaned data in $\hat{\eta}_\mu$ are replaced by the residuals from the regression of the series on intercept and trend.

In the construction of the KPSS tests, conditions enough to imply Functional Central Limit Theorems (FCLT) are assumed. One of these conditions is the finite variance assumption. However, when the data have fat-tailed errors such as those from the Cauchy distribution in which the moments do not exist, the limiting distributions of the KPSS statistics are functionals of the Lévy process (Amsler and Schmidt 2000), not a Wiener process. In the paper by de Jong, Amsler, and Schmidt (2002), the authors relax the moment assumption and propose a modified version of KPSS test, $\hat{\eta}_\mu$. They call their test the "indicator KPSS" test which we will label $\hat{\iota}_\mu$. The sample data are transformed using an indicator which gives the value of 1, 0, or -1 depending on whether or not the sample observation is above, on, or below the sample median.

Under the null of level-stationarity, $\hat{\iota}_\mu$ is shown to have the same limiting distribution as the KPSS test, $\hat{\eta}_\mu$.

In this chapter, we use a similar indicator to transform the data, but allow for a deterministic trend as well as a non-zero level for the data. Let the indicator KPSS statistic with a time trend be denoted as $\hat{\iota}_\tau$. We show that the asymptotic distribution of $\hat{\iota}_\tau$ under the null of trend-stationarity is a function of the second-level Brownian bridge, which is also the limiting distribution of the KPSS statistic with a time trend, $\hat{\eta}_\tau$.

3.2 Asymptotic Theory

3.2.1 Assumptions

Let $\{\{x_{Tj}\}_{j=1}^{T}\}_{T=1}^{\infty}$ be a triangular array of random variables such that

$$x_{Tj} = \alpha_0 + \beta_0 \frac{j}{T} + \epsilon_j. \tag{3.1}$$

Assumption 1. *There exist unique* α_0, β_0 *such that* $\text{med}(x_{Tj}) = \alpha_0 + \beta_0 j/T$ *for all* T *and* $j = 1, \cdots, T$.

Note that this implies that the unique median of $\epsilon_j = x_{Tj} - \alpha_0 - \beta_0 j/T$ is zero. The next assumption is a convergence condition on the average variance of the sum of transformed ϵ_j with the finiteness of the long run variance, σ^2.

Assumption 2. *Define* $\sigma^2 = \lim_{T\to\infty} \mathrm{E}\left(T^{-1/2}\sum_{j=1}^{T} \text{sgn}(\epsilon_j)\right)^2$, *where the* sgn *function takes three different values of 1, 0, or -1 depending the sign of an argument:* $\text{sgn}(x) = 1$ *if* $x > 0$, $\text{sgn}(x) = 0$ *if* $x = 0$, *and* $\text{sgn}(x) = -1$ *if* $x < 0$. *Then,* $0 < \sigma^2 < \infty$.

The next assumption is about the kernel function, $k(\cdot)$.

Assumption 3. *$k(\cdot)$ is continuous at 0 and at all but a finite number of points. $k(x) = k(-x)$ for all $x \in \mathbb{R}$. $k(0) = 1$. $|k(x)| \leq l(x)$ where $l(x)$ is nonincreasing and $\int_0^\infty |l(x)|\, dx < \infty$. Also, $k(\cdot)$ satisfies $\int_{-\infty}^{\infty} |\psi(\xi)|\, d\xi < \infty$, where*

$$\psi(\xi) = (2\pi)^{-1} \int_{-\infty}^{\infty} k(x) \exp(-i\xi x) dx. \tag{3.2}$$

The Bartlett, Parzen, Quadratic Spectral, and Tukey-Hanning kernel functions are some possible choices (de Jong and Davidson 2000). These kernel functions are designed to lessen the effects of the longer lags smoothly to zero so that the kernel function such as the uniform or the truncated kernel is excluded. The next set of assumptions is about the ϵ_j, and will be used in deriving the asymptotic distribution of the indicator KPSS statistic under the null of trend-stationarity.

Assumption 4. *The ϵ_j are stationary random variables and strong ($\alpha-$) mixing with mixing coefficients $\alpha(m)$ which satisfies $\alpha(m) \leq Cm^{-\frac{r}{r-2}-\eta}$ for some finite $r > 2$, some $\eta > 0$ and a constant C. And ϵ_j has a continuous density $f(\epsilon)$ in a neighborhood $[-\eta, \eta]$ of 0 for some $\eta > 0$, and $\inf_{\epsilon \in [-\eta,\eta]} f(\epsilon) > 0$.*

Assumption 4[1]is different from general conditions on the stationary errors used in the derivation of the asymptotic distributions of the KPSS statistics (Phillips (1987) or Phillips and Perron (1988)). The important difference is moment conditions on ϵ_j. For example, in Phillips (1987), the moment condition like $\sup_j \mathrm{E}\,|\epsilon_j|^{\vartheta} < \infty$ for some $\vartheta > 2$ is assumed. However, in this chapter, we do not assume the existence of moments of ϵ_j under the null. This is made possible by the use of the indicators.

The next assumption is for the alternative of unit root.

Assumption 5. *The ϵ_j satisfy $T^{-1/2}\epsilon_{[\xi T]} \Rightarrow \lambda W(\xi)$ for some $\lambda \in (0, \infty)$ and $\xi \in$*

[1]In this chapter, Assumption 4 is stated in terms of ϵ_j, not x_{Tj} as in de Jong, Amsler, and Schmidt (2002). This is to emphasize that we are interested in the test of trend-stationarity. That is, the assumptions for ϵ_j in this chapter (or the detrended series, $x_{Tj} - \alpha_0 - \beta_0 j/T$) are the same as the Assumption 2 in de Jong, Amsler, and Schmidt (2002).

$[0,1]$, *where* $W(\cdot)$ *is a Wiener process or Brownian motion.*

Note that Assumption 5 also implies that $T^{-1/2}x_{T,[\xi T]} \Rightarrow \lambda W(\xi)$ since $T^{-1/2}$ $x_{T,[\xi T]} = T^{-1/2}(\alpha_0 + \beta_0[\xi T]/T) + T^{-1/2}\epsilon_{[\xi T]} = o_p(1) + T^{-1/2}\epsilon_{[\xi T]} \Rightarrow \lambda W(\xi)$.

3.2.2 Indicator KPSS statistic

Using the least absolute deviations (LAD) estimators $\hat{\alpha}$, $\hat{\beta}$ which are solutions to

$$\min_{\alpha,\beta} \sum_{j=1}^{T} \left| x_{Tj} - \alpha - \beta \frac{j}{T} \right|, \tag{3.3}$$

we define the cumulation of the indicator data

$$S_{Tt} = \sum_{j=1}^{t} \operatorname{sgn}(x_{Tj} - \hat{\alpha} - \hat{\beta}\frac{j}{T}). \tag{3.4}$$

Then, the indicator KPSS statistic with a time trend, $\hat{\iota}_T$ is defined as

$$\hat{\sigma}^{-2}T^{-2}\sum_{t=1}^{T} S_{Tt}^2. \tag{3.5}$$

A consistent estimator of σ^2, $\hat{\sigma}^2$ can be constructed from the "indicator" residuals, $\operatorname{sgn}(x_{Tj} - \hat{\alpha} - \hat{\beta}j/T)$. Using a weighting function, the heteroskedasticity-autocorrelation consistent (HAC) estimator, $\hat{\sigma}^2$ is obtained by

$$\hat{\sigma}^2 = T^{-1}\sum_{i=1}^{T}\sum_{j=1}^{T} k(\frac{i-j}{\gamma_T}) \operatorname{sgn}(x_{Ti} - \hat{\alpha} - \hat{\beta}\frac{i}{T}) \operatorname{sgn}(x_{Tj} - \hat{\alpha} - \hat{\beta}\frac{j}{T}), \tag{3.6}$$

where $k(\cdot)$ is the kernel function. γ_T is the lag truncation parameter which goes to ∞ as $T \to \infty$ and satisfies the condition of $\gamma_T/T \to 0$.

Note that $\hat{\iota}_T$ is defined in a similar way as the KPSS statistic, $\hat{\eta}_T$. The difference is that we use the deviations from the median while $\hat{\eta}_T$ is based on the deviations from the mean of the series. The indicator KPSS is based on the sample median which is the generalization of the fit from a LAD regression.

As noted in de Jong, Amsler, and Schmidt (2002), the purpose of trimming the data is to remove the effects of fat tails or make the variance finite. We use the sgn function to bypass the problem of how to scale the data so that only the location of the data is used to transform the data. This is because $\text{sgn}(x) = (\text{I}(x \geq 0) - \text{I}(x \leq 0))$ and $|x| = x \cdot (\text{I}(x \geq 0) - \text{I}(x \leq 0)) = x \cdot \text{sgn}(x)$, where $I(\cdot)$ takes the value of one if the argument is true and zero otherwise.

3.2.3 Conjectures

Before stating theorems for the asymptotic distributions of $\hat{\iota}_T$, let us make conjectures on $\hat{\alpha}$ and $\hat{\beta}$ as the proofs for the following claims are only partially done.

Conjecture 1. *Under Assumptions 1 and 4,* $T^{1/2}(\hat{\alpha} - \alpha_0) = O_p(1)$ *and* $T^{1/2}(\hat{\beta} - \beta_0) = O_p(1)$.

What we want to assert in this conjecture is, for an arbitrarily large $K > 0$,

$$\begin{aligned} &\limsup_{T\to\infty} P(\sup_{\phi_1 > K} \sup_{\phi_2 > K} Y_{1T}(\phi_1, \phi_2) \geq 0) \\ &= \limsup_{T\to\infty} P(\sup_{\phi_1 > K} \sup_{\phi_2 > K} Y_{2T}(\phi_1, \phi_2) \geq 0) = 0 \end{aligned} \tag{3.7}$$

so that the probability of having solutions ϕ_1, ϕ_2 outside $\Phi = \{(\phi_1, \phi_2) \in \mathbb{R}^2 : -K \leq \phi_1 \leq K, -K \leq \phi_2 \leq K\}$ goes to zero as $T \to \infty$ where

$$\begin{aligned} Y_{1T}(\phi_1, \phi_2) &= T^{-1/2} \sum_{j=1}^{T} \text{sgn}(x_{Tj} - \alpha_0 - \beta_0 \frac{j}{T} - T^{-1/2}(\phi_1 + \phi_2 \frac{j}{T})), \\ Y_{2T}(\phi_1, \phi_2) &= T^{-1/2} \sum_{j=1}^{T} \text{sgn}(x_{Tj} - \alpha_0 - \beta_0 \frac{j}{T} - T^{-1/2}(\phi_1 + \phi_2 \frac{j}{T})) \frac{j}{T}. \end{aligned} \tag{3.8}$$

However, there are the four possibilities for obtaining large values for ϕ_1 and/or ϕ_2:

- case [1]: $\phi_1 > K$ and $\phi_2 > K$,
- case [2]: $\phi_1 < -K$ and $\phi_2 < -K$,
- case [3]: $\phi_1 < -K$ and $\phi_2 > K$,
- case [4]: $\phi_1 > K$ and $\phi_2 < -K$.

The proof of (3.7) corresponds to case [1]. The proof for case [1] and case [2] is shown in an Appendix. However, the proof for case [3] and case [4] remains to be done. Also, note that the case in which only one of $|\phi_1|$ and $|\phi_2|$ is larger than K is a special case of either case [1] or case [2] and can be proved in a similar way as in the proof for the first two cases.

The following conjecture makes a similar claim as Conjecture 1, but the difference is that we assume ϵ_j is an I(1) process.

Conjecture 2. *Under Assumptions 1 and 5,* $T^{-1/2}(\hat{\alpha} - \alpha_0) = O_p(1)$ *and* $T^{-1/2}(\hat{\beta} - \beta_0) = O_p(1)$.

Here we also have to consider the four possibilities in which $|T^{-1/2}\hat{\alpha}|$ and/or $|T^{-1/2}\hat{\beta}|$ are greater than K. In the Appendix, we prove two cases when both $|T^{-1/2}\hat{\alpha}|$ and $|T^{-1/2}\hat{\beta}|$ are either greater than K or less than K. The two other cases would be proved in a similar way as in the unsolved cases of Conjecture 1.

3.2.4 The Asymptotic Distributions of the Indicator KPSS Statistic

Theorem 1. *Under Assumptions 1, 2, 3, and 4 and Conjecture 1,*

$$T^{-2}\sum_{t=1}^{T} S_{Tt}^2 \xrightarrow{d} \sigma^2 \int_0^1 V_2(r)^2 dr, \tag{3.9}$$

where $V_2(r)$ is the second-level Brownian bridge,

$$V_2(r) = W(r) + (2r - 3r^2)W(1) + (-6r + 6r^2)\int_0^1 W(r)dr. \tag{3.10}$$

And

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2. \tag{3.11}$$

The limiting distribution of $\hat{\iota}_T$ is $\int_0^1 V_2(r)^2 dr$, which is also the limiting distribution of the KPSS test with time trend, $\hat{\eta}_T$ so that the same critical values in the paper by Kwiatkowski, Phillips, Schmidt, and Shin (1992, p.166) can be used. Under the alternative in which x_{Tj} is an I(1) process, we have the following result.

Theorem 2. *Under Assumptions 1 and 5 and Conjecture 2,*

$$T^{-3}\sum_{t=1}^{T} S_{Tt}^2 \xrightarrow{d} \lambda^2 \int_0^1 \left(\int_0^\zeta \operatorname{sgn}\left(W(\xi) - \frac{A}{\lambda} - \frac{B}{\lambda}\xi\right) d\xi\right)^2 d\zeta, \tag{3.12}$$

where $(T^{-1/2}\hat{\alpha}, T^{-1/2}\hat{\beta})' \xrightarrow{d} (A, B)'$ for random variables A and B, and $\hat{\sigma}^2/\gamma_T \xrightarrow{d}$ $2\int_0^\infty k(\zeta)d\zeta$.

Other than whether the underlying series is stationary or not, the important difference between the assumptions used in deriving the results of Theorem 1 and Theorem 2 is the moment condition on ϵ_j. In Theorem 1, we do not impose a condition for the existence of the moments. However, in Theorem 2, we need a finite second moment of ϵ_j in order to apply FCLT.

Also, note that the limiting distribution under the alternative of unit root in Theorem 2 is different from that of the KPSS statistic with a time trend which is

$$\int_0^1 \left(\int_0^a W^*(s)ds\right)^2 da/K\int_0^1 W^*(s)^2 ds \tag{3.13}$$

where $W^*(s) = W(s) + (6s-4)\int_0^1 W(r)dr + (-12s+6)\int_0^1 rW(r)dr$. The differences in the asymptotic distributions will turn into power differences as in de Jong, Amsler, and Schmidt (2002). Under the alternative of unit root with the fat-tailed errors, the indicator KPSS test with a time trend would be more powerful than the KPSS test with trend, and less powerful when the errors are normally distributed. This is because the indicator is only concerned with the location of the data.

3.3 Concluding remarks

In this chapter, we have extended the indicator KPSS test proposed by de Jong, Amsler, and Schmidt (2002) to the case in which a time trend as well as non-zero level is allowed. The indicator KPSS test with a time trend also does not require the existence of the moments of the series, yet produce the same asymptotic results as the KPSS test with a time trend, $\hat{\eta}_T$. However, this result depends on our conjectures on the estimators.

The indicator can be extended to unit root tests such as Dickey-Fuller, Phillips-Perron, or Schmidt-Phillips tests. We expect that the use of the indicator would produce more powerful results when the errors have sufficiently fat tails, which is commonly associated with financial time series. However, the asymptotic results of unit root tests with the indicator under the null of unit root might be different from those without the indicator. As in our chapter and de Jong, Amsler, and Schmidt (2002), the unit root tests with the indicator might produce the same asymptotic results as the tests without the indicator under the alternative of stationarity.

3.4 Appendix: Mathematical Proof

Here is the outline of proofs. Lemma 1 shows an inequality involving the L_p-norm which will be used in Lemma 2. Lemma 2 states that $G_T(1,\boldsymbol{\phi}) - \mathrm{E}\, G_T(1,\boldsymbol{\phi}) = T^{-1/2}\sum_{j=1}^{T}(y_{Tj}(\boldsymbol{\phi}) - \mathrm{E}\, y_{Tj}(\boldsymbol{\phi}))$ is stochastically equicontinuous. In Lemma 3, the uniform convergences of $G_T(r,\boldsymbol{\phi})$ and $H_T(1,\boldsymbol{\gamma})$ over corresponding compact sets of parameter values will be established. A partial proof of Conjecture 1 follows. In Lemma 4, the asymptotic distributions of the estimators of the regression coefficients are derived. Then, Theorem 1 proves the asymptotic distribution of the indicator KPSS statistic along with the consistency of the long-run variance estimator. Conjecture 2 is partially proved. Finally, in Theorem 2, we show the limiting distribution of the statistic when the x_{Tj} have a unit root.

Lemma 4. *For strong (α-) mixing random variables $y_{Tj} \in \mathbb{R}$ whose α-mixing coefficients satisfy $\alpha(m) \le Cm^{-\frac{r}{r-2}-\eta}$ for some $\eta > 0$,*

$$\mathrm{E}\left(\sum_{j=1}^{T}(y_{Tj} - \mathrm{E}\, y_{Tj})\right)^2 \le \mathrm{E}\left(\max_{1\le t\le T}\left(\sum_{j=1}^{t}(y_{Tj} - \mathrm{E}\, y_{Tj})\right)^2\right) \le C'\sum_{j=1}^{T} \| y_{Tj} \|_r^2 \tag{3.14}$$

for constants C, C'.

Proof of Lemma 1. By Theorem 17.5 and Corollary 16.10 of Davidson (1994). □

Lemma 5. *Let* $\mathbf{z} = (1, j/T)'$, $\boldsymbol{\phi} = (\phi_1, \phi_2)'$ *and* $\boldsymbol{\psi} = (\psi_1, \psi_2)'$. *Let* $y_{Tj}(\boldsymbol{\phi}) = \mathrm{sgn}(x_{Tj} - \alpha_0 - \beta_0 j/T - T^{-1/2}\mathbf{z}'\boldsymbol{\phi}) - \mathrm{sgn}(x_{Tj} - \alpha_0 - \beta_0 j/T)$. *Then, under Assumptions*

1 and 4, for all $K, \varepsilon > 0$,

$$\lim_{\delta \to 0} \limsup_{T \to \infty} P\left(\sup_{\substack{|\phi_1| \leq K, |\phi_2| \leq K, \\ |\psi_1| \leq K, |\psi_2| \leq K, \\ (|\phi_1 - \psi_1| + |\phi_2 - \psi_2|) < \delta}} T^{-1/2} \sum_{j=1}^{T} \left| (y_{Tj}(\boldsymbol{\phi}) - \mathrm{E}\, y_{Tj}(\boldsymbol{\phi})) - (y_{Tj}(\boldsymbol{\psi}) - \mathrm{E}\, y_{Tj}(\boldsymbol{\psi})) \right| < \varepsilon \right) = 1. \tag{3.15}$$

Proof of Lemma 2. For T large enough such that $2KT^{-1/2} \leq \eta$

$$\begin{aligned}
&\sup_{(|\phi_1 - \psi_1| + |\phi_2 - \psi_2|) < \delta} T^{-1/2} \sum_{j=1}^{T} \left| \mathrm{E}\, y_{Tj}(\boldsymbol{\phi}) - \mathrm{E}\, y_{Tj}(\boldsymbol{\psi}) \right| \\
&= 2 \sup_{(|\phi_1 - \psi_1| + |\phi_2 - \psi_2|) < \delta} T^{-1/2} \sum_{j=1}^{T} \left| F(T^{-1/2} \mathbf{z}' \boldsymbol{\phi}) - F(T^{-1/2} \mathbf{z}' \boldsymbol{\psi}) \right| \\
&= 2 \sup_{(|\phi_1 - \psi_1| + |\phi_2 - \psi_2|) < \delta} T^{-1/2} \sum_{j=1}^{T} \Big| f(T^{-1/2} \mathbf{z}' \boldsymbol{\xi}) \cdot T^{-1/2} (\phi_1 - \psi_1) \\
&\qquad + f(T^{-1/2} \mathbf{z}' \boldsymbol{\xi}) \cdot T^{-1/2} (\phi_2 - \psi_2) \frac{j}{T} \Big| \\
&= 2 \sup_{(|\phi_1 - \psi_1| + |\phi_2 - \psi_2|) < \delta} T^{-1} \sum_{j=1}^{T} f(T^{-1/2} \mathbf{z}' \boldsymbol{\xi}) \cdot \left| (\phi_1 - \psi_1) + (\phi_2 - \psi_2) \frac{j}{T} \right| \\
&\leq 2 \sup_{(|\phi_1 - \psi_1| + |\phi_2 - \psi_2|) < \delta} T^{-1} \sum_{j=1}^{T} f(T^{-1/2} \mathbf{z}' \boldsymbol{\xi}) \cdot \left| |\phi_1 - \psi_1| + |\phi_2 - \psi_2| \right| \\
&\leq 2\delta T^{-1} \sum_{j=1}^{T} f(T^{-1/2} (\xi_1 + \xi_2 \frac{j}{T})) \\
&\leq 2\delta \sup_{\tilde{\xi} \in [-\eta, \eta]} f(\tilde{\xi}),
\end{aligned}$$

where $F(\cdot)$ is the cdf of ϵ and $\boldsymbol{\xi} \in L(\boldsymbol{\phi}, \boldsymbol{\psi})$, a line segment from $\boldsymbol{\phi}$ to $\boldsymbol{\psi}$. This establishes the equicontinuity of $T^{-1/2} \sum_{j=1}^{T} \mathrm{E}\, y_{Tj}(\boldsymbol{\phi})$ on $\Phi = \{(\phi_1, \phi_2) \in \mathbb{R}^2 : -K \leq \phi_1 \leq K, -K \leq \phi_2 \leq K\}$ since δ can be made arbitrarily small. Then, the stochastic equicontinuity of $T^{-1/2}\sum_{j=1}^{T} y_{Tj}(\boldsymbol{\phi})$ can be shown as follows. Let $\mathbf{i}_{i\delta} = (i\delta, i\delta)'$ and $\mathbf{i}_{(i+2)\delta} = ((i+2)\delta, (i+2)\delta)'$.

$$\sup_{\substack{|\phi_1|\le K, |\phi_2|\le K, \\ |\psi_1|\le K, |\psi_2|\le K, \\ (|\phi_1-\psi_1|+|\phi_2-\psi_2|)<\delta}} T^{-1/2}\sum_{j=1}^{T} |y_{Tj}(\boldsymbol{\phi}) - y_{Tj}(\boldsymbol{\psi})|$$

$$= \sup_{-\left[\frac{K}{\delta}\right]-1\le i\le \left[\frac{K}{\delta}\right]} \sup_{\phi_1,\psi_1\in[i\delta,(i+2)\delta]\cap[-K,K]}$$

$$\sup_{\phi_2,\psi_2\in[i\delta,(i+2)\delta]\cap[-K,K]} T^{-1/2}\sum_{j=1}^{T} |y_{Tj}(\boldsymbol{\phi}) - y_{Tj}(\boldsymbol{\psi})|$$

$$\le \sup_{-\left[\frac{K}{\delta}\right]-1\le i\le \left[\frac{K}{\delta}\right]} T^{-1/2}\sum_{j=1}^{T} \left(y_{Tj}(\mathbf{i}_{i\delta}) - y_{Tj}(\mathbf{i}_{(i+2)\delta})\right)$$

$$\xrightarrow{p} \sup_{-\left[\frac{K}{\delta}\right]-1\le i\le \left[\frac{K}{\delta}\right]} T^{-1/2}\sum_{j=1}^{T} \left(\mathrm{E}\, y_{Tj}\left(\mathbf{i}_{i\delta}\right) - \mathrm{E}\, y_{Tj}\left(\mathbf{i}_{(i+2)\delta}\right)\right)$$

$$\le \sup_{\substack{|\phi_1|\le K, |\phi_2|\le K, \\ |\psi_1|\le K, |\psi_2|\le K, \\ (|\phi_1-\psi_1|+|\phi_2-\psi_2|)<\delta}} T^{-1/2}\sum_{j=1}^{T} \left|\mathrm{E}\, y_{Tj}(\boldsymbol{\phi}) - \mathrm{E}\, y_{Tj}(\boldsymbol{\psi})\right|.$$

For the first inequality above, note that y_{Tj} is nonincreasing so that the maximum distance between $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ in sub-intervals of $[-K, K]$ will give rise to the supremum of $(y_{Tj}(\mathbf{i}_{i\delta}) - y_{Tj}(\mathbf{i}_{(i+2)\delta}))$ for each sub-interval. For the last inequality, $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ are now chosen all over the interval. The pointwise convergence for every i holds because, for T large enough such that $2KT^{-1/2} \le \eta$, by Lemma 1,

$$\mathrm{E}\left(T^{-1/2}\sum_{j=1}^{T}\left[\left(y_{Tj}\left(\mathbf{i}_{i\delta}\right) - y_{Tj}\left(\mathbf{i}_{(i+2)\delta}\right)\right) - \mathrm{E}\left(y_{Tj}\left(\mathbf{i}_{i\delta}\right) - y_{Tj}\left(\mathbf{i}_{(i+2)\delta}\right)\right)\right]\right)^2$$

$$\le C\sum_{j=1}^{T} \|T^{-1/2}\left(y_{Tj}\left(\mathbf{i}_{i\delta}\right) - y_{Tj}\left(\mathbf{i}_{(i+2)\delta}\right)\right)\|_r^2$$

$$
\begin{aligned}
&= CT^{-1}\sum_{j=1}^{T}\left\|\left(y_{Tj}\left(\mathbf{i}_{i\delta}\right) - y_{Tj}\left(\mathbf{i}_{(i+2)\delta}\right)\right)\right\|_r^2 \\
&\leq CT^{-1}\sum_{j=1}^{T} \sup_{\substack{|\phi_1|\leq K,\\ |\phi_2|\leq K}} \left\|y_{Tj}(\boldsymbol{\phi})\right\|_r^2 \\
&= CT^{-1}\sum_{j=1}^{T} \sup_{\substack{|\phi_1|\leq K,\\ |\phi_2|\leq K}} \Big\| \operatorname{sgn}(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T} - T^{-1/2}\phi_1 - T^{-1/2}\phi_2\frac{j}{T}) \\
&\qquad - \operatorname{sgn}(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T})\Big\|_r^2 \\
&\leq C'\left|1 - 2F(-2T^{-1/2}K)\right|^{\frac{2}{r}} \\
&= C'\left|2T^{-1/2}K \sup_{|\xi|\leq\eta} f(\xi)\right|^{\frac{2}{r}} \longrightarrow 0,
\end{aligned}
$$

as $T \to \infty$ and for some positive constants C, C'. Note that $F(\cdot)$ is the cdf of ϵ and there are two cases to consider for the last inequality since $(T^{-1/2}\phi_1 + T^{-1/2}\phi_2 j/T)$ will be either nonpositive or nonnegative. In the below, we prove that the inequality holds in either case;

- case [1]: $T^{-1/2}\phi_1 + T^{-1/2}\phi_2 j/T \leq 0$. This implies $-2T^{-1/2}K \leq -T^{-1/2}K - T^{-1/2}Kj/T \leq T^{-1/2}\phi_1 + T^{-1/2}\phi_2 j/T$ so that

$$
\begin{aligned}
&\sup_{\substack{|\phi_1|\leq K,\\ |\phi_2|\leq K}} \Big\| \operatorname{sgn}(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T} - T^{-1/2}\phi_1 - T^{-1/2}\phi_2\frac{j}{T}) \\
&\qquad - \operatorname{sgn}(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T})\Big\|_r^2 \\
&= \sup_{\substack{|\phi_1|\leq K,\\ |\phi_2|\leq K}} \left(\mathrm{E}\left|2\,\mathrm{I}\left(T^{-1/2}\phi_1 + T^{-1/2}\phi_2\frac{j}{T} \leq x_{Tj} - \alpha_0 - \beta_0\frac{j}{T} \leq 0\right)\right|^r\right)^{\frac{2}{r}} \\
&\leq \left(\mathrm{E}\left|2\,\mathrm{I}\left(-2T^{-1/2}K \leq x_{Tj} - \alpha_0 - \beta_0\frac{j}{T} \leq 0\right)\right|^r\right)^{\frac{2}{r}} \\
&= 4\left(\mathrm{E}\,\mathrm{I}\left(-2T^{-1/2}K \leq x_{Tj} - \alpha_0 - \beta_0\frac{j}{T} \leq 0\right)\right)^{\frac{2}{r}}
\end{aligned}
$$

$$= 4\left(P\left(-2T^{-1/2}K \le x_{Tj} - \alpha_0 - \beta_0\frac{j}{T} \le 0\right)\right)^{\frac{2}{r}}$$
$$= 2^{2-\frac{2}{r}}\left[1 - 2F\left(-2T^{-1/2}K\right)\right]^{\frac{2}{r}}.$$

- case [2]: $0 \le T^{-1/2}\phi_1 + T^{-1/2}\phi_2 j/T$. Then,

$$\sup_{\substack{|\phi_1|\le K,\\ |\phi_2|\le K}} \Big\| \operatorname{sgn}(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T} - T^{-1/2}\phi_1 - T^{-1/2}\phi_2\frac{j}{T})$$
$$- \operatorname{sgn}(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T})\Big\|_r^2$$
$$= \sup_{\substack{|\phi_1|\le K,\\ |\phi_2|\le K}} \left(\mathrm{E}\left|-2\,\mathrm{I}\left(0 \le x_{Tj} - \alpha_0 - \beta_0\frac{j}{T} \le T^{-1/2}\phi_1 + T^{-1/2}\phi_2\frac{j}{T}\right)\right|^r\right)^{\frac{2}{r}}$$
$$\le \left(\mathrm{E}\,|-2|^r\left|\mathrm{I}\left(0 \le x_{Tj} - \alpha_0 - \beta_0\frac{j}{T} \le 2T^{-1/2}K\right)\right|^r\right)^{\frac{2}{r}}$$
$$= 2^{2-\frac{2}{r}}\left(2F\left(2T^{-1/2}K\right) - 1\right)^{\frac{2}{r}}.$$

□

Lemma 6. *Let $y_{Tj}(\boldsymbol{\phi})$ be as in Lemma 2, and let*

$$G_T(r,\boldsymbol{\phi}) = T^{-1/2}\sum_{j=1}^{[rT]} y_{Tj}(\boldsymbol{\phi}), \qquad H_T(1,\boldsymbol{\gamma}) = \begin{pmatrix} T^{-1/2}\sum_{j=1}^{T} y_{Tj}(\boldsymbol{\gamma}) \\ T^{-1/2}\sum_{j=1}^{T}\frac{j}{T}y_{Tj}(\boldsymbol{\gamma}) \end{pmatrix}. \tag{3.16}$$

Then, under Assumptions 1 and 4, for any $K > 0$,

$$\sup_{\substack{|\phi_1|\le K,\\ |\phi_2|\le K}} \sup_{r\in[0,1]} |G_T(r,\boldsymbol{\phi}) - \mathrm{E}\,G_T(r,\boldsymbol{\phi})| \xrightarrow{p} 0, \tag{3.17}$$

$$\sup_{\substack{|\gamma_1|\le K,\\ |\gamma_2|\le K}} |H_T(1,\boldsymbol{\gamma}) - \mathrm{E}\,H_T(1,\boldsymbol{\gamma})| \xrightarrow{p} 0. \tag{3.18}$$

Proof of Lemma 3. Let

$$J_T(\boldsymbol{\phi}) = \sup_{r\in[0,1]} |G_T(r,\boldsymbol{\phi}) - \mathrm{E}\, G_T(r,\boldsymbol{\phi})| . \tag{3.19}$$

For each $\boldsymbol{\phi}$ with its elements in $[-K, K]$ and T large enough such that $2KT^{-1/2} \leq \eta$,

$$\begin{aligned}
\mathrm{E}\,(J_T(\boldsymbol{\phi}))^2 &= \mathrm{E}\left(\sup_{r\in[0,1]} |G_T(r,\boldsymbol{\phi}) - \mathrm{E}\, G_T(r,\boldsymbol{\phi})|\right)^2 \\
&= \mathrm{E}\left(\sup_{r\in[0,1]} \left|T^{-1/2}\sum_{j=1}^{[rT]} \left(y_{Tj}(\boldsymbol{\phi}) - \mathrm{E}\, y_{Tj}(\boldsymbol{\phi})\right)\right|\right)^2 \\
&\leq CT^{-1}\sum_{j=1}^{T} \|y_{Tj}(\boldsymbol{\phi})\|_r^2 \qquad (3.20)\\
&\leq C'\left|1 - 2F(-2T^{-1/2}K)\right|^{\frac{2}{r}} \\
&= C'\left|2T^{-1/2}K \sup_{|\xi|\leq\eta} f(\xi)\right|^{\frac{2}{r}} \to 0,
\end{aligned}$$

as $T \to \infty$ and for some positive constants C, C'. This implies $J_T(\phi) = o_p(1)$. $J_T(\boldsymbol{\phi})$

is also stochastically equicontinuous because, for $\boldsymbol{\phi} \in \Phi$ and $\boldsymbol{\phi}' \in \Phi$,

$$
\begin{aligned}
&\left| J_T(\boldsymbol{\phi}) - J_T(\boldsymbol{\phi}') \right| \\
&= \left| \sup_{r\in[0,1]} \left| G_T(r,\boldsymbol{\phi}) - \mathrm{E}\, G_T(r,\boldsymbol{\phi}) \right| - \sup_{r\in[0,1]} \left| G_T(r,\boldsymbol{\phi}') - \mathrm{E}\, G_T(r,\boldsymbol{\phi}') \right| \right| \\
&\leq \sup_{r\in[0,1]} \left| G_T(r,\boldsymbol{\phi}) - \mathrm{E}\, G_T(r,\boldsymbol{\phi}) - G_T(r,\boldsymbol{\phi}') + \mathrm{E}\, G_T(r,\boldsymbol{\phi}') \right| \\
&= \sup_{r\in[0,1]} \left| T^{-1/2} \sum_{j=1}^{[rT]} \left(y_{Tj}(\boldsymbol{\phi}) - \mathrm{E}\, y_{Tj}(\boldsymbol{\phi}) - y_{Tj}(\boldsymbol{\phi}') + \mathrm{E}\, y_{Tj}(\boldsymbol{\phi}') \right) \right| \qquad (3.21) \\
&\leq \sup_{r\in[0,1]} T^{-1/2} \sum_{j=1}^{[rT]} \left| y_{Tj}(\boldsymbol{\phi}) - \mathrm{E}\, y_{Tj}(\boldsymbol{\phi}) - y_{Tj}(\boldsymbol{\phi}') + \mathrm{E}\, y_{Tj}(\boldsymbol{\phi}') \right| \\
&\leq T^{-1/2} \sum_{j=1}^{T} \left| y_{Tj}(\boldsymbol{\phi}) - \mathrm{E}\, y_{Tj}(\boldsymbol{\phi}) - y_{Tj}(\boldsymbol{\phi}') + \mathrm{E}\, y_{Tj}(\boldsymbol{\phi}') \right|.
\end{aligned}
$$

Then, by applying Lemma 2 to (3.21) and together with pointwise convergence in (3.20), uniform convergence of $J_T(\boldsymbol{\phi})$ follows, which proves (3.17). Pointwise convergence in probability of $G_T(r,\boldsymbol{\phi})$ to $\mathrm{E}\, G_T(r,\boldsymbol{\phi})$ in $\boldsymbol{\phi}$ for every possible r has been proved. In other words, $J_T(\boldsymbol{\phi})$ is pointwise convergent in probability to zero and stochastically equicontinuous. The uniform convergence in probability of $J_T(\boldsymbol{\phi})$ to zero or that of $G_T(r,\boldsymbol{\phi})$ to $\mathrm{E}\, G_T(r,\boldsymbol{\phi})$ follows because, with compact sets and the equicontinuity of $\mathrm{E}\, G_T(r,\boldsymbol{\phi})$, pointwise convergence with stochastic equicontinuity is equivalent to uniform convergence. See Newey (1991).

For the uniform convergence of $(H_T(1,\boldsymbol{\gamma}) - \mathrm{E}\, H_T(1,\boldsymbol{\gamma}))$ to zero, note that

$$
H_T(1,\boldsymbol{\gamma}) - \mathrm{E}\, H_T(1,\boldsymbol{\gamma}) = \begin{pmatrix} G_T(1,\boldsymbol{\gamma}) - \mathrm{E}\, G_T(1,\boldsymbol{\gamma}) \\ T^{-1/2} \sum_{j=1}^{T} \frac{j}{T} \left(y_{Tj}(\boldsymbol{\gamma}) - \mathrm{E}\, y_{Tj}(\boldsymbol{\gamma}) \right) \end{pmatrix},
$$

where $G_T(1,\boldsymbol{\gamma}) - \mathrm{E}\, G_T(1,\boldsymbol{\gamma})$ is uniformly convergent from (3.17), and $T^{-1/2} \sum_{j=1}^{T} (y_{Tj}(\boldsymbol{\gamma}) - \mathrm{E}\, y_{Tj}(\boldsymbol{\gamma}))\, j/T$ is also uniformly convergent which can be seen by comparing the following expressions to (3.20) and (3.21), respectively:

$$
\begin{aligned}
&\mathrm{E}\left(T^{-1/2}\sum_{j=1}^{T}\frac{j}{T}\left(y_{Tj}(\gamma)-\mathrm{E}\,y_{Tj}(\gamma)\right)\right)^2\\
&\leq \mathrm{E}\left(T^{-1/2}\sum_{j=1}^{T}\left(y_{Tj}(\gamma)-\mathrm{E}\,y_{Tj}(\gamma)\right)\right)^2\\
&\leq C\sum_{j=1}^{T}\|y_{Tj}(\gamma)\|_r^2,
\end{aligned}
$$

and

$$
\begin{aligned}
&\left|T^{-1/2}\sum_{j=1}^{T}\frac{j}{T}\left(y_{Tj}(\gamma)-\mathrm{E}\,y_{Tj}(\gamma)\right)-T^{-1/2}\sum_{j=1}^{T}\frac{j}{T}\left(y_{Tj}(\gamma')-\mathrm{E}\,y_{Tj}(\gamma')\right)\right|\\
&\leq T^{-1/2}\sum_{j=1}^{T}\frac{j}{T}\left|\left(y_{Tj}(\gamma)-\mathrm{E}\,y_{Tj}(\gamma)-y_{Tj}(\gamma')+\mathrm{E}\,y_{Tj}(\gamma')\right)\right|\\
&\leq T^{-1/2}\sum_{j=1}^{T}\left|\left(y_{Tj}(\gamma)-\mathrm{E}\,y_{Tj}(\gamma)-y_{Tj}(\gamma')+\mathrm{E}\,y_{Tj}(\gamma')\right)\right|,
\end{aligned}
$$

thereby proving (3.18). □

Partial Proof of Conjecture 1. Here, we provide a partial proof of Conjecture 1 under Assumptions 1 and 2: that is, $T^{1/2}(\hat{\alpha}-\alpha_0)$ and $T^{1/2}(\hat{\beta}-\beta_0)$ are $O_p(1)$. Let

$$
\begin{aligned}
Y_{1T}(\phi_1,\phi_2) &= T^{-1/2}\sum_{j=1}^{T}\operatorname{sgn}\left(x_{Tj}-\alpha_0-\beta_0\frac{j}{T}-T^{-1/2}\left(\phi_1+\phi_2\frac{j}{T}\right)\right),\\
Y_{2T}(\phi_1,\phi_2) &= T^{-1/2}\sum_{j=1}^{T}\operatorname{sgn}\left(x_{Tj}-\alpha_0-\beta_0\frac{j}{T}-T^{-1/2}\left(\phi_1+\phi_2\frac{j}{T}\right)\right)\frac{j}{T}.
\end{aligned}
\tag{3.22}
$$

In order to show $T^{1/2}(\hat{\alpha}-\alpha_0)=O_p(1)$ and $T^{1/2}(\hat{\beta}-\beta_0)=O_p(1)$, four possibilities for obtaining large values for ϕ_1 and/or ϕ_2 have to be considered;

- case [1]: $\phi_1 > K$ and $\phi_2 > K$,

- case [2]: $\phi_1 < -K$ and $\phi_2 < -K$,
- case [3]: $\phi_1 < -K$ and $\phi_2 > K$,
- case [4]: $\phi_1 > K$ and $\phi_2 < -K$.

However, we prove only first two cases in which both ϕ_1 and ϕ_2 are either greater than K or less than $-K$, which are done by showing that the probability of having such solutions outside the compact set $\Phi = \{(\phi_1, \phi_2) \in \mathbb{R}^2 : -K \leq \phi_1 \leq K, -K \leq \phi_2 \leq K\}$ is arbitrarily small. Also, cases when only one of ϕ_1 and ϕ_2 is outside Φ can be proved in the same way as in de Jong, Amsler, and Schmidt (2002).

Suppose that $\phi_1 > K$ and $\phi_2 > K$. Let's start with $Y_{1T}(\phi_1, \phi_2)$. For $T \geq 4K^2\eta^{-2}$,

$$\sup_{\phi_1 > K}\sup_{\phi_2 > K} Y_{1T}(\phi_1, \phi_2) = T^{-1/2}\sum_{j=1}^{T} \text{sgn}(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T} - T^{-1/2}(K + K\frac{j}{T}))$$

$$\xrightarrow{p} T^{-1/2}\sum_{j=1}^{T}(1 - 2F(T^{-1/2}(K + K\frac{j}{T})))$$

$$= T^{-1/2}\sum_{j=1}^{T} 2(F(0) - F(T^{-1/2}(K + K\frac{j}{T})))$$

$$= T^{-1/2}\sum_{j=1}^{T} 2f(x)(-T^{-1/2}(K + K\frac{j}{T}))$$

for some $x \in [0, T^{-1/2}K(1 + \frac{j}{T})] \subseteq [0, T^{-1/2}K(1+1)] \subseteq [0, \eta]$

$$\leq -2\inf_{|x|\leq\eta} f(x)T^{-1}K\sum_{j=1}^{T}(1 + \frac{j}{T}) = -2K(1 + \frac{T+1}{2T})\inf_{|x|\leq\eta} f(x)$$

$$\leq -2K(1 + \frac{T}{2T})\inf_{|x|\leq\eta} f(x) = -3K\inf_{|x|\leq\eta} f(x)$$

so that

$$\limsup_{T\to\infty} P(\sup_{\phi_1 > K}\sup_{\phi_2 > K} Y_{1T}(\phi_1, \phi_2) \geq 0) \leq \limsup_{T\to\infty} P(-3K\inf_{|x|\leq\eta} f(x) \geq 0) = 0, \quad (3.23)$$

since $K > 0$ and $\inf_{|x|\le\eta} f(x) > 0$. Next, in case of $Y_{2T}(\phi_1, \phi_2)$,

$$
\begin{aligned}
&\sup_{\phi_1>K}\sup_{\phi_2>K} T^{-1/2}\sum_{j=1}^{T}\operatorname{sgn}(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T} - T^{-1/2}(\phi_1 + \phi_2\frac{j}{T}))\frac{j}{T} \\
&= T^{-1/2}\sum_{j=1}^{T}\operatorname{sgn}(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T} - T^{-1/2}(K + K\frac{j}{T}))\frac{j}{T} \\
&= T^{-1/2}\sum_{j=1}^{T}\mathrm{I}(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T} - T^{-1/2}(K + K\frac{j}{T}) \ge 0)\frac{j}{T} \\
&\quad - T^{-1/2}\sum_{j=1}^{T}\mathrm{I}(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T} - T^{-1/2}(K + K\frac{j}{T}) < 0)\frac{j}{T} \\
&\xrightarrow{p} T^{-1/2}\sum_{j=1}^{T}\left(\int_{T^{-1/2}(K+K\frac{j}{T})}^{\infty} f(x)dx - \int_{-\infty}^{T^{-1/2}(K+K\frac{j}{T})} f(x)dx\right)\frac{j}{T} \\
&= T^{-1/2}\sum_{j=1}^{T}(F(\infty) - 2F(T^{-1/2}(K + K\frac{j}{T})) + F(-\infty))\frac{j}{T} \\
&= T^{-1/2}\sum_{j=1}^{T}(1 - 2F(T^{-1/2}(K + K\frac{j}{T}))\frac{j}{T} \\
&= T^{-1/2}\sum_{j=1}^{T}2f(x)(-T^{-1/2}(K + K\frac{j}{T}))\frac{j}{T} \quad \text{for some } x \in [0, \eta] \\
&\le -2T^{-1}K\inf_{|x|\le\eta} f(x)\sum_{j=1}^{T}\left(\frac{j}{T} + \frac{j^2}{T^2}\right) \\
&= -2K\inf_{|x|\le\eta} f(x)\left(\frac{T(T+1)}{2T^2} + \frac{T(T+1)(2T+1)}{6T^3}\right) \\
&\le -2K\inf_{|x|\le\eta} f(x)\left(\frac{1}{2} + \frac{1}{3}\right) = -\frac{5}{3}K\inf_{|x|\le\eta} f(x).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\limsup_{T\to\infty} P\Bigg(&\sup_{\phi_1>K}\sup_{\phi_2>K} \\
&T^{-1/2}\sum_{j=1}^{T}\operatorname{sgn}\left(x_t - \alpha_0 - \beta_0\frac{j}{T} - T^{-1/2}(\phi_1 + \phi_2\frac{j}{T})\right)\frac{j}{T} \ge 0\Bigg) = 0.
\end{aligned}
\tag{3.24}
$$

The second case to consider is when $\phi_1 < -K$ and $\phi_2 < -K$. The proofs in this case are similar to the ones just done. Let's start with $Y_{1T}(\phi_1, \phi_2)$.

$$\begin{aligned}
&\inf_{\phi_1<-K}\inf_{\phi_2<-K} Y_{1T}(\phi_1,\phi_2)\\
&= T^{-1/2}\sum_{j=1}^{T}\text{sgn}(x_{Tj}-\alpha_0-\beta_0\frac{j}{T}-T^{-1/2}(-K-K\frac{j}{T}))\\
&\xrightarrow{p} T^{-1/2}\sum_{j=1}^{T}(1-2F(T^{-1/2}(-K-K\frac{j}{T})))\\
&= T^{-1/2}\sum_{j=1}^{T}2(F(0)-F(T^{-1/2}(-K-K\frac{j}{T})))\\
&= T^{-1/2}\sum_{j=1}^{T}2f(x)(T^{-1/2}(K+K\frac{j}{T})) \quad \text{for some } x\in[-\eta,0]\\
&\le 2\inf_{|x|\le\eta} f(x)T^{-1}K\sum_{j=1}^{T}(1+\frac{j}{T}) = 2K(1+\frac{T+1}{2T})\inf_{|x|\le\eta} f(x)\\
&\le 2K(1+\frac{T}{2T})\inf_{|x|\le\eta} f(x) = 3K\inf_{|x|\le\eta} f(x)
\end{aligned}$$

so that

$$\limsup_{T\to\infty} P(\inf_{\phi_1<-K}\inf_{\phi_2<-K} Y_{1T}(\phi_1,\phi_2)\le 0)\le \limsup_{T\to\infty} P(3K\inf_{|x|\le\eta} f(x)\le 0)=0. \tag{3.25}$$

In case of $Y_{2T}(\phi_1, \phi_2)$,

$$\begin{aligned}
&\inf_{\phi_1<-K}\inf_{\phi_2<-K} T^{-1/2}\sum_{j=1}^{T}\text{sgn}(x_t-\alpha_0-\beta_0\frac{j}{T}-T^{-1/2}(\phi_1+\phi_2\frac{j}{T}))\frac{j}{T}\\
&= T^{-1/2}\sum_{j=1}^{T}\text{sgn}(x_t-\alpha_0-\beta_0\frac{j}{T}+T^{-1/2}(K+K\frac{j}{T}))\frac{j}{T}
\end{aligned}$$

$$\xrightarrow{p} T^{-1/2}\sum_{j=1}^{T}(1 - 2F(-T^{-1/2}(K + K\frac{j}{T})))\frac{j}{T}$$

$$= T^{-1/2}\sum_{j=1}^{T} 2f(x)(T^{-1/2}(K + K\frac{j}{T}))\frac{j}{T}$$

$$\geq 2T^{-1}K \inf_{|x|\leq\eta} f(x)\sum_{j=1}^{T}\left(\frac{j}{T} + \frac{j^2}{T^2}\right) \quad \text{for some } x \in [-\eta, 0]$$

$$\geq \frac{5}{3}K \inf_{|x|\leq\eta} f(x).$$

Therefore,

$$\limsup_{T\to\infty} P\Bigg(\inf_{\phi_1<-K}\inf_{\phi_2<-K} T^{-1/2}\sum_{j=1}^{T}\operatorname{sgn}\left(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T} - T^{-1/2}(\phi_1 + \phi_2\frac{j}{T})\right)\frac{j}{T} \leq 0\Bigg) = 0. \tag{3.26}$$

The above two cases imply that $\limsup_{T\to\infty} P(T^{1/2}(|\hat{\alpha} - \alpha_0| + |\hat{\beta} - \beta_0|) > K)$ can be made arbitrarily small by choosing K large enough. □

Lemma 7. *Under Assumptions 1 and 4,*

$$\begin{pmatrix} T^{1/2}(\hat{\alpha} - \alpha_0) \\ T^{1/2}(\hat{\beta} - \beta_0) \end{pmatrix} = \frac{T}{2f(0)}\begin{pmatrix} T & \frac{T+1}{2} \\ \frac{T+1}{2} & \frac{(T+1)(2T+1)}{6T} \end{pmatrix}^{-1} \times \begin{pmatrix} \sigma W_T(1) \\ \sigma W_T(1) - \sigma T^{-1}\sum_{t=1}^{T} W_T(\frac{t}{T}) \end{pmatrix} + \begin{pmatrix} o_p(1) \\ o_p(1) \end{pmatrix}. \tag{3.27}$$

Proof of Lemma 4. Let[2]

$$\hat{\gamma} = \begin{pmatrix} T^{1/2}(\hat{\alpha} - \alpha_0) \\ T^{1/2}(\hat{\beta} - \beta_0) \end{pmatrix}.$$

Note that

$$\begin{aligned}
\mathrm{E}\, H_T(1, \hat{\gamma}) &= \begin{pmatrix} T^{-1/2} \sum_{j=1}^{T} \mathrm{E}\, \mathrm{sgn}(x_{Tj} - \hat{\alpha} - \hat{\beta}\frac{j}{T}) \\ T^{-1/2} \sum_{j=1}^{T} \frac{j}{T}\, \mathrm{E}\, \mathrm{sgn}(x_{Tj} - \hat{\alpha} - \hat{\beta}\frac{j}{T}) \end{pmatrix} \\
&= \begin{pmatrix} 2T^{-1/2} \sum_{j=1}^{T} (F(0) - F(\hat{\alpha} + \hat{\beta}\frac{j}{T} - \alpha_0 - \beta_0\frac{j}{T})) \\ 2T^{-1/2} \sum_{j=1}^{T} \frac{j}{T}(F(0) - F(\hat{\alpha} + \hat{\beta}\frac{j}{T} - \alpha_0 - \beta_0\frac{j}{T})) \end{pmatrix} \\
&= \begin{pmatrix} -2T^{-1/2} \sum_{j=1}^{T} (f(\xi_j) - f(0) + f(0))(\hat{\alpha} + \hat{\beta}\frac{j}{T} - \alpha_0 - \beta_0\frac{j}{T}) \\ -2T^{-1/2} \sum_{j=1}^{T} \frac{j}{T}(f(\xi_j) - f(0) + f(0))(\hat{\alpha} + \hat{\beta}\frac{j}{T} - \alpha_0 - \beta_0\frac{j}{T}) \end{pmatrix} \\
&= \begin{pmatrix} -2f(0)T^{-1/2} \sum_{j=1}^{T} (\hat{\alpha} + \hat{\beta}\frac{j}{T} - \alpha_0 - \beta_0\frac{j}{T}) \\ -2f(0)T^{-1/2} \sum_{j=1}^{T} \frac{j}{T}(\hat{\alpha} + \hat{\beta}\frac{j}{T} - \alpha_0 - \beta_0\frac{j}{T}) \end{pmatrix} \\
&\quad + \begin{pmatrix} -2T^{-1/2} \sum_{j=1}^{T} (f(\xi_j) - f(0))(\hat{\alpha} + \hat{\beta}\frac{j}{T} - \alpha_0 - \beta_0\frac{j}{T}) \\ -2T^{-1/2} \sum_{j=1}^{T} \frac{j}{T}(f(\xi_j) - f(0))(\hat{\alpha} + \hat{\beta}\frac{j}{T} - \alpha_0 - \beta_0\frac{j}{T}) \end{pmatrix} \\
&= -\frac{2f(0)}{T} \begin{pmatrix} T & \frac{T+1}{2} \\ \frac{T+1}{T} & \frac{(T+1)(2T+1)}{6T} \end{pmatrix} \begin{pmatrix} T^{1/2}(\hat{\alpha} - \alpha_0) \\ T^{1/2}(\hat{\beta} - \beta_0) \end{pmatrix} + 0,
\end{aligned}$$

since $\max_{1 \le j \le T} |\xi_j| \le \max_{1 \le j \le T} |(\hat{\alpha} - \alpha_0) + (\hat{\beta} - \beta_0)j/T| \le |\hat{\alpha} - \alpha_0| + |\hat{\beta} - \beta_0|$

[2] so that

$$G_T(r, \hat{\gamma}) = T^{-1/2} \sum_{j=1}^{[rT]} (\mathrm{sgn}(x_{Tj} - \hat{\alpha} - \hat{\beta}\frac{j}{T}) - \mathrm{sgn}(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T})),$$

$$\mathrm{E}\, G_T(r, \hat{\gamma}) = T^{-1/2} \sum_{j=1}^{[rT]} \mathrm{E}\, \mathrm{sgn}(x_{Tj} - \hat{\alpha} - \hat{\beta}\frac{j}{T}).$$

$= O_p(T^{-1/2})$ by Conjecture 1, and $\max_{1 \leq j \leq T} |f(\xi_j) - f(0)| \longrightarrow 0$. Then,

$$
\begin{aligned}
O_p(T^{-1/2}) &= \begin{pmatrix} T^{-1/2} \sum_{j=1}^{T} \operatorname{sgn}(x_{Tj} - \hat{\alpha} - \hat{\beta}\frac{j}{T}) \\ T^{-1/2} \sum_{j=1}^{T} \frac{j}{T} \operatorname{sgn}(x_{Tj} - \hat{\alpha} - \hat{\beta}\frac{j}{T}) \end{pmatrix} \\
&= (H_T(1, \hat{\gamma}) - \mathrm{E}\, H_T(1, \hat{\gamma})) \\
&\qquad + \begin{pmatrix} T^{-1/2} \sum_{j=1}^{T} \operatorname{sgn}(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T}) \\ T^{-1/2} \sum_{j=1}^{T} \frac{j}{T} \operatorname{sgn}(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T}) \end{pmatrix} + \mathrm{E}\, H_T(1, \hat{\gamma}) \\
&= o_p(1) + \begin{pmatrix} \sigma W_T(1) \\ \sigma W_T(1) - \sigma T^{-1} \sum_{j=1}^{T} W_T(\frac{j}{T}) \end{pmatrix} \\
&\qquad - \frac{2f(0)}{T} \begin{pmatrix} T & \frac{T+1}{2} \\ \frac{T+1}{T} & \frac{(T+1)(2T+1)}{6T} \end{pmatrix} \begin{pmatrix} T^{1/2}(\hat{\alpha} - \alpha_0) \\ T^{1/2}(\hat{\beta} - \beta_0) \end{pmatrix}.
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
&\begin{pmatrix} T^{1/2}(\hat{\alpha} - \alpha_0) \\ T^{1/2}(\hat{\beta} - \beta_0) \end{pmatrix} \\
&= \frac{T}{2f(0)} \begin{pmatrix} T & \frac{T+1}{2} \\ \frac{T+1}{2} & \frac{(T+1)(2T+1)}{6T} \end{pmatrix}^{-1} \begin{pmatrix} \sigma W_T(1) \\ \sigma W_T(1) - \sigma T^{-1} \sum_{j=1}^{T} W_T(\frac{j}{T}) \end{pmatrix} + o_p(1) \\
&\longrightarrow \frac{1}{2f(0)} \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix}^{-1} \begin{pmatrix} \sigma W(1) \\ \sigma W(1) - \sigma \int_0^1 W(r)dr \end{pmatrix}.
\end{aligned}
$$

□

Proof of Theorem 1. First, we derive the asymptotic distribution of $T^{1/2} S_{Tt}$. Second, we show the consistency of the long run variance estimator. Put together, we have the limiting distribution of the indicator KPSS statistic with a time trend, $\hat{\iota}_T$.

$$
\begin{aligned}
&T^{-1/2}\sum_{j=1}^{[rT]} \operatorname{sgn}(x_{Tj} - \hat{\alpha} - \hat{\beta}\frac{j}{T}) \\
&= (G_T(r,\boldsymbol{\phi}) - \mathrm{E}\, G_T(r,\boldsymbol{\phi})) + T^{-1/2}\sum_{j=1}^{[rT]} \operatorname{sgn}(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T}) \\
&\quad + \begin{pmatrix} T^{-1/2}\sum_{j=1}^{[rT]} \frac{\partial}{\partial\alpha}\left(E\operatorname{sgn}(x_{Tj} - \alpha - \beta\frac{j}{T})\right)|_{\alpha=\alpha_0,\beta=\beta_0} \\ T^{-1/2}\sum_{j=1}^{[rT]} \frac{\partial}{\partial\beta}\left(E\operatorname{sgn}(x_{Tj} - \alpha - \beta\frac{j}{T})\right)|_{\alpha=\alpha_0,\beta=\beta_0} \end{pmatrix}' \begin{pmatrix} (\hat{\alpha} - \alpha_0) \\ (\hat{\beta} - \beta_0) \end{pmatrix} \\
&= (G_T(r,\boldsymbol{\phi}) - EG_T(r,\boldsymbol{\phi})) + T^{-1/2}\sum_{j=1}^{[rT]} \operatorname{sgn}(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T}) \\
&\quad - 2f(0)T^{-1/2}\sum_{j=1}^{[rT]}(\hat{\alpha} - \alpha_0) - 2f(0)T^{-1/2}\sum_{j=1}^{[rT]}\frac{j}{T}(\hat{\beta} - \beta_0) \\
&= (G_T(r,\boldsymbol{\phi}) - EG_T(r,\boldsymbol{\phi})) + T^{-1/2}\sum_{j=1}^{[rT]} \operatorname{sgn}(x_{Tj} - \alpha_0 - \beta_0\frac{j}{T}) \\
&\quad - 2f(0)\frac{[rT]}{T}T^{1/2}(\hat{\alpha} - \alpha_0) - 2f(0)\frac{[rT]([rT]+1)}{2T^2}T^{1/2}(\hat{\beta} - \beta_0) \\
&= o_p(1) + \sigma W_T(r) + (\frac{-4[rT]}{T} - \frac{3[rT]([rT]+1)}{T^2})T^{-1/2}\sum_{j=1}^{T}\operatorname{sgn}(\epsilon_j) \\
&\quad + (\frac{6[rT]}{T} - \frac{6[rT]([rT]+1)}{T^2})T^{-1/2}\sum_{j=1}^{T}\frac{j}{T}\operatorname{sgn}(\epsilon_j) \qquad \text{by (3.27)} \\
&= o_p(1) + \sigma W_T(r) + (\frac{-4[rT]}{T} - \frac{3[rT]([rT]+1)}{T^2})\sigma W_T(1) \\
&\quad + (\frac{6[rT]}{T} - \frac{6[rT]([rT]+1)}{T^2})\sigma(W_T(1) - T^{-1}\sum_{j=1}^{T} W_T(\frac{j}{T}) + o_p(1)) \\
&= o_p(1) + \sigma W_T(r) + (\frac{2[rT]}{T} - \frac{3[rT]([rT]+1)}{T^2})\sigma W_T(1) \\
&\quad + (-\frac{6[rT]}{T} + \frac{6[rT]([rT]+1)}{T^2})T^{-1}\sum_{j=1}^{T}\sigma W_T(\frac{j}{T}) \\
&\xrightarrow{d} \sigma\left(W(r) + (2r - 3r^2)W(1) + (-6r + 6r^2)\int_0^1 W(r)dr\right) = \sigma V_2(r),
\end{aligned}
$$

where the $o_p(1)$ term is uniform in r. Note that for each $j = 1, ..., T$,

$$\begin{aligned}
&\mathrm{E}\,\mathrm{sgn}(x_{Tj} - \alpha - \beta\frac{j}{T})\Big|_{\alpha=\alpha_0,\beta=\beta_0} = 1 - 2F(\hat{\alpha} + \hat{\beta}\frac{j}{T} - \alpha_0 - \beta_0\frac{j}{T}) \\
&= 1 - 2F(0) - 2(\hat{\alpha} + \hat{\beta}\frac{j}{T} - \alpha_0 - \beta_0\frac{j}{T})f(\tilde{\alpha} + \tilde{\beta}\frac{j}{T} - \alpha_0 - \beta_0\frac{j}{T}).
\end{aligned} \tag{3.28}$$

where F is the cdf of ϵ_j. When $\hat{\alpha}$ and $\hat{\beta}$ are consistent, $f(\tilde{\alpha} + \tilde{\beta}\frac{j}{T} - \alpha_0 - \beta_0\frac{j}{T})$ would be asymptotically equal to $f(0)$.

$$\begin{aligned}
&\frac{\partial}{\partial\alpha}\mathrm{E}\,\mathrm{sgn}(x_{Tj} - \alpha - \beta\frac{j}{T})\Big|_{\alpha=\alpha_0,\beta=\beta_0} \\
&= -2f(\alpha + \beta\frac{j}{T} - \alpha_0 - \beta_0\frac{j}{T})\Big|_{\alpha=\alpha_0,\beta=\beta_0} = -2f(0), \\
&\frac{\partial}{\partial\beta}\mathrm{E}\,\mathrm{sgn}(x_{Tj} - \alpha - \beta\frac{j}{T})\Big|_{\alpha=\alpha_0,\beta=\beta_0} \\
&= -2\frac{j}{T}f(\alpha + \beta\frac{j}{T} - \alpha_0 - \beta_0\frac{j}{T})\Big|_{\alpha=\alpha_0,\beta=\beta_0} = -2\frac{j}{T}f(0).
\end{aligned} \tag{3.29}$$

Also, note that

$$\begin{aligned}
&T^{-1/2}\sum_{j=1}^{T}\frac{j}{T}\,\mathrm{sgn}(\epsilon_j) \\
&= T^{-1/2}\sum_{j=1}^{T}\mathrm{sgn}(\epsilon_j) - T^{-1}\sum_{j=1}^{T}\left(T^{-1/2}\sum_{i=1}^{t}\mathrm{sgn}(\epsilon_i)\right) + o_p(1) \\
&= W_T(1) - T^{-1}\sum_{j=1}^{T}W_T(\frac{j}{T}) + o_p(1)
\end{aligned} \tag{3.30}$$

since

$$\begin{aligned}
&\sum_{j=1}^{T}\sum_{i=1}^{t}\operatorname{sgn}(\epsilon_i)\\
&=\operatorname{sgn}(\epsilon_1)+(\operatorname{sgn}(\epsilon_1)+\operatorname{sgn}(\epsilon_2))+\cdots+(\operatorname{sgn}(\epsilon_1)+\operatorname{sgn}(\epsilon_2)+\cdots+\operatorname{sgn}(\epsilon_T))\\
&=\sum_{j=1}^{T}(T-j+1)\operatorname{sgn}(\epsilon_j)\\
&=T\sum_{j=1}^{T}\operatorname{sgn}(\epsilon_j)-\sum_{j=1}^{T}(j\cdot\operatorname{sgn}(\epsilon_j))+\sum_{j=1}^{T}\operatorname{sgn}(\epsilon_j),
\end{aligned}$$

and

$$\begin{aligned}
&T^{-\frac{3}{2}}\sum_{j=1}^{T}\sum_{i=1}^{t}\operatorname{sgn}(\epsilon_i)\\
&=T^{-1/2}\sum_{j=1}^{T}\operatorname{sgn}(\epsilon_j)-T^{-1/2}\sum_{j=1}^{T}\frac{j}{T}\operatorname{sgn}(\epsilon_j)+T^{-1}T^{-1/2}\sum_{j=1}^{T}\operatorname{sgn}(\epsilon_j)\\
&=T^{-1/2}\sum_{j=1}^{T}\operatorname{sgn}(\epsilon_j)-T^{-1/2}\sum_{j=1}^{T}\frac{j}{T}\operatorname{sgn}(\epsilon_j)+o_p(1).
\end{aligned}$$

Since the sgn function is regular (Park and Phillips 1999, p. 272), we apply Theorem 3.2 in Park and Phillips (1999) to derive the limiting distribution (3.5):

$$\begin{aligned}
&\sigma^2T^{-2}\sum_{t=1}^{T}S_{Tt}^2\\
&\xrightarrow{d}\int_0^1\left(W(r)+(2r-3r^2)W(1)+(-6r+6r^2)\int_0^1 W(r)\right)^2 dr.
\end{aligned}\tag{3.31}$$

Next, we will prove the consistency of $\hat{\sigma}^2$. First, note that for $t=1,\cdots,T$,

$$\begin{aligned}
&\operatorname{sgn}(x_t-\hat{\alpha}-\hat{\beta}\frac{t}{T})\\
&=(y_t(\gamma)-\operatorname{E}y_t(\gamma))+\operatorname{E}\operatorname{sgn}(x_t-\hat{\alpha}-\hat{\beta}\frac{t}{T})+\operatorname{sgn}(x_t-\alpha_0-\beta_0\frac{t}{T})
\end{aligned}$$

$$= (y_t(\gamma) - \mathrm{E}\, y_t(\gamma)) + \left(1 - 2F(\hat{\alpha} + \hat{\beta}\frac{t}{T} - \alpha_0 - \beta_0\frac{t}{T})\right) + \mathrm{sgn}(x_t - \alpha_0 - \beta_0\frac{t}{T})$$

$$= (y_t(\gamma) - \mathrm{E}\, y_t(\gamma)) + \left(1 - 2F(\hat{\alpha} + \hat{\beta}\frac{t}{T} - \alpha_0 - \beta_0\frac{t}{T})\right) + \mathrm{sgn}(\epsilon_t)$$

$$= a_{Tt} + b_{Tt} + c_t.$$

Then,

$$\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} k(\frac{t-s}{\gamma_T})(a_{Tt} + b_{Tt} + c_t)(a_{Ts} + b_{Ts} + c_s), \tag{3.32}$$

which will be shown to be asymptotically equivalent to

$$T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} k(\frac{t-s}{\gamma_T})c_t c_s$$

so that

$$\hat{\sigma}^2 - T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} k(\frac{t-s}{\gamma_T})c_t c_s \xrightarrow{p} 0.$$

Then, by Theorem 2.1 of de Jong and Davidson (2000),

$$T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} k(\frac{t-s}{\gamma_T})c_t c_s \xrightarrow{p} \sigma^2.$$

What is shown below is that all the cross products in the right hand side of (3.32) except the term involving $c_t c_s$ are $o_p(1)$ as $T \to \infty$. First, note that, for T large enough,

$$\begin{aligned}\sup_{1\le t\le T} |b_{Tt}| &= \sup_{1\le t\le T}\left|1 - 2F(\hat{\alpha} + \hat{\beta}\frac{t}{T} - \alpha_0 - \beta_0\frac{t}{T})\right| \\ &= \sup_{1\le t\le T}\left|2(\hat{\alpha} + \hat{\beta}\frac{t}{T} - \alpha_0 - \beta_0\frac{t}{T}) \cdot f(\tilde{\alpha} + \tilde{\beta}\frac{t}{T} - \alpha_0 - \beta_0\frac{t}{T})\right|\end{aligned}$$

$$\leq 2 \sup_{|x|\leq\eta} f(x) \left(|\hat{\alpha} - \alpha_0| + |\hat{\beta} - \beta_0| \right) = O_p(T^{-1/2}),$$

since for large T and consistent $\hat{\alpha}$ and $\hat{\beta}$, $f(\tilde{\alpha} + \tilde{\beta}t/T - \alpha_0 - \beta_0 t/T)$ converges to $f(0)$ uniformly in t so that the above inequality holds for $|x| \leq \eta$. Second, $T^{-1/2}\sum_{t=1}^{T} |a_{Tt}| = o_p(1)$ by Lemma 3. Then,

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} k(\frac{t-s}{\gamma_T}) a_{Tt}a_{Ts} \\
&= T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\int_{-\infty}^{\infty} \exp\left(i\xi\frac{t-s}{\gamma_T}\right)\psi(\xi)d\xi \cdot a_{Tt}a_{Ts} \\
&= T^{-1}\int_{-\infty}^{\infty}\sum_{t=1}^{T} a_{Tt}\sum_{s=1}^{T} a_{Ts}\exp\left(i\xi\frac{t-s}{\gamma_T}\right)\psi(\xi)d\xi \\
&\leq T^{-1}\int_{-\infty}^{\infty}\sum_{t=1}^{T} |a_{Tt}| \cdot \left|\sum_{s=1}^{T} a_{Ts}\exp\left(i\xi\frac{t-s}{\gamma_T}\right)\psi(\xi)\right| d\xi \\
&\leq T^{-1}\int_{-\infty}^{\infty}\sum_{t=1}^{T} |a_{Tt}|\sum_{s=1}^{T} |a_{Ts}| \cdot |\psi(\xi)|\, d\xi \\
&= \int_{-\infty}^{\infty} |\psi(\xi)|\, d\xi \left(T^{-1/2}\sum_{t=1}^{T} |a_{Tt}|\right)^2 \\
&= o_p(1),
\end{aligned}$$

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} k(\frac{t-s}{\gamma_T}) a_{Tt}b_{Ts} \\
&\leq T^{-1}\sum_{t=1}^{T}\left(|a_{Tt}|\sum_{s=1}^{T}\left|k(\frac{t-s}{\gamma_T})b_{Ts}\right|\right) \\
&\leq T^{-\frac{3}{2}}\sum_{t=1}^{T}\left(|a_{Tt}|\sum_{s=1}^{T}\left|k(\frac{t-s}{\gamma_T})T^{1/2}\sup_{1\leq s\leq T}|b_{Ts}|\right|\right) \\
&\leq T^{-\frac{3}{2}}\sum_{t=1}^{T}|a_{Tt}|\sum_{j=-T}^{T}\left|k(\frac{j}{\gamma_T})\right| \cdot T^{1/2} \cdot O_p(T^{-1/2})
\end{aligned}$$

$$= T^{-1} \cdot T^{-1/2} \sum_{t=1}^{T} |a_{Tt}| \cdot \sum_{j=-T}^{T} \left| k(\frac{j}{\gamma_T}) \right| \cdot O_p(1)$$

$$= o_p(1) \cdot \frac{\gamma_T}{T} \cdot O_p(1) = o_p(1),$$

$$T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} k(\frac{t-s}{\gamma_T}) a_{Tt} c_s$$

$$= T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \int_{-\infty}^{\infty} \exp\left(i\xi \left(\frac{t-s}{\gamma_T} \right) \right) \psi(\xi) d\xi \cdot a_{Tt} c_s$$

$$= T^{-1} \int_{-\infty}^{\infty} \sum_{t=1}^{T} a_{Tt} \sum_{s=1}^{T} c_s \exp\left(i\xi \left(\frac{t-s}{\gamma_T} \right) \right) \psi(\xi) d\xi$$

$$\leq T^{-1} \int_{-\infty}^{\infty} \sum_{t=1}^{T} |a_{Tt}| \cdot \left| \sum_{s=1}^{T} c_s \exp\left(i\xi \left(\frac{t-s}{\gamma_T} \right) \right) \psi(\xi) \right| d\xi$$

$$= T^{-1/2} \sum_{t=1}^{T} |a_{Tt}| \int_{-\infty}^{\infty} T^{-1/2} \left| \sum_{s=1}^{T} c_s \exp\left(i\xi \left(\frac{t-s}{\gamma_T} \right) \right) \psi(\xi) \right| d\xi$$

$$= o_p(1) \cdot O_p(1) = o_p(1),$$

$$T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} k(\frac{t-s}{\gamma_T}) b_{Tt} b_{Ts}$$

$$\leq T^{-1} \sum_{t=1}^{T} \left(|b_{Tt}| \sum_{s=1}^{T} \left| k(\frac{t-s}{\gamma_T}) b_{Ts} \right| \right)$$

$$\leq T^{-1} \sup_{1 \leq t \leq T} |b_{Tt}| \sup_{1 \leq s \leq T} |b_{Ts}| \sum_{t=1}^{T} \sum_{s=1}^{T} \left| k(\frac{t-s}{\gamma_T}) \right|$$

$$\leq \sup_{1 \leq t \leq T} |b_{Tt}| \sup_{1 \leq s \leq T} |b_{Ts}| T^{-1} \sum_{t=1}^{T} \sum_{j=-T}^{T} \left| k(\frac{j}{\gamma_T}) \right|$$

$$\leq \sup_{1 \leq t \leq T} |b_{Tt}| \sup_{1 \leq s \leq T} |b_{Ts}| \sum_{j=-T}^{T} \left| k(\frac{j}{\gamma_T}) \right|$$

$$\leq O_p(T^{-1/2}) \cdot O_p(T^{-1/2}) \cdot \gamma_T = O_p(\frac{\gamma_T}{T}) = o_p(1),$$

$$T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} k(\frac{t-s}{\gamma_T}) b_{Tt} c_s$$

$$
\begin{aligned}
&\le T^{-1}\sum_{t=1}^{T}\left(|b_{Tt}|\sum_{s=1}^{T}\left|k(\frac{t-s}{\gamma_T})c_s\right|\right)\\
&\le \sup_{1\le t\le T}|b_{Tt}|\cdot T^{-1}\left(\sum_{s=1}^{T}|c_s|\sum_{t=1}^{T}\left|k(\frac{t-s}{\gamma_T})\right|\right)\\
&= T^{1/2}\sup_{1\le t\le T}|b_{Tt}|\cdot T^{-\frac{3}{2}}\left(\sum_{s=1}^{T}|c_s|\sum_{t=1}^{T}\left|k(\frac{t-s}{\gamma_T})\right|\right)\\
&\le O_p(1)\cdot O_p(\frac{\gamma_T}{T}) = O_p(\frac{\gamma_T}{T}) = o_p(1).
\end{aligned}
$$

□

Partial Proof of Conjecture 2. We provide a partial proof for $(T^{-1/2}\hat{\alpha}, T^{-1/2}\hat{\beta})' \xrightarrow{d} (A, B)'$ for some random variables A and B. In order to show that $T^{-1/2}\hat{\alpha}$ and $T^{-1/2}\hat{\beta}$ are $O_p(1)$, we need to consider four cases as in the proof of Conjecture 1. However, we only provide two cases in which both $T^{-1/2}\hat{\alpha}$ and $T^{-1/2}\hat{\beta}$ are either greater than K or less than $-K$. First, we show that the probability such that both $T^{-1/2}\hat{\alpha}$ and $T^{-1/2}\hat{\beta}$ are greater than K goes to zero as $K \to \infty$ and $T \to \infty$. Note that

$$
\begin{aligned}
&\sup_{a>K}\sup_{b>K} T^{-1}\sum_{j=1}^{T}\operatorname{sgn}(T^{-1/2}x_{Tj} - a - b\frac{j}{T})\\
&= T^{-1}\sum_{j=1}^{T}\operatorname{sgn}(T^{-1/2}x_{Tj} - K - K\frac{j}{T})\\
&\xrightarrow{d} \int_0^1 \operatorname{sgn}(\lambda W(\xi) - K - K\xi)d\xi\\
&\equiv T_1(K)
\end{aligned}
$$

and

$$
\sup_{a>K}\sup_{b>K} T^{-1}\sum_{j=1}^{T}\operatorname{sgn}(T^{-1/2}x_{Tj} - a - b\frac{j}{T})\frac{j}{T}
$$

$$
\begin{aligned}
&= T^{-1}\sum_{j=1}^{T} \operatorname{sgn}(T^{-1/2}x_{Tj} - K - K\frac{j}{T})\frac{j}{T} \\
&\xrightarrow{d} \int_0^1 \operatorname{sgn}(\lambda W(\xi) - K - K\xi)\xi d\xi \\
&\equiv T_2(K),
\end{aligned}
$$

where the limiting distributions of $T^{-1}\sum_{j=1}^{T} \operatorname{sgn}\ (T^{-1/2}x_{Tj} - K - Kj/T)$ and $T^{-1}\sum_{j=1}^{T} \operatorname{sgn}(T^{-1/2}x_{Tj} - K - Kj/T)j/T$ are obtained by applying Theorem 3.2 of Park and Phillips (1999) because the sgn function is regular (Park and Phillips 1999, p. 272).

Then, the probability with which $T^{-1/2}\hat{\alpha}$ and $T^{-1/2}\hat{\beta}$ are not bounded in the limit is calculated as follows:

$$
\begin{aligned}
&P(T^{-1/2}\hat{\alpha} > K \text{ and } T^{-1/2}\hat{\beta} > K) \\
&= P(\sup_{a>K}\sup_{b>K} T^{-1}\sum_{j=1}^{T} \operatorname{sgn}(T^{-1/2}x_{Tj} - a - b\frac{j}{T}) \geq 0) \\
&\quad + P(\sup_{a>K}\sup_{b>K} T^{-1}\sum_{j=1}^{T} \operatorname{sgn}(T^{-1/2}x_{Tj} - a - b\frac{j}{T})\frac{j}{T} \geq 0) \\
&\longrightarrow P(T_1(K) \geq 0) + P(T_2(K) \geq 0).
\end{aligned}
$$

Note that as $K \to \infty$, $\operatorname{sgn}(\lambda W(\xi) - K - K\xi) \to \operatorname{sgn}(-\infty) = -1$ so that $T_1(K) \xrightarrow{p} \int_0^1 (-1)d\xi = -1$ and $T_2(K) \xrightarrow{p} \int_0^1 (-\xi)d\xi = -0.5$. This implies that $P(T_1(K) \geq 0)$ and $P(T_2(K) \geq 0)$ will go to zero as $K \to \infty$. Therefore,

$$
\limsup_{K\to\infty}\limsup_{T\to\infty} P(T^{-1/2}\hat{\alpha} > K \text{ and } T^{-1/2}\hat{\beta} > K) = 0. \tag{3.33}
$$

Similarly, it can be shown that

$$\limsup_{K\to\infty}\limsup_{T\to\infty} P(T^{-1/2}\hat{\alpha} < -K \text{ and } T^{-1/2}\hat{\beta} < -K) = 0. \tag{3.34}$$

Note that

$$\begin{aligned}
&\inf_{a<-K}\inf_{b<-K} T^{-1}\sum_{j=1}^{T}\mathrm{sgn}(T^{-1/2}x_{Tj} - a - b\frac{j}{T}) \\
&= T^{-1}\sum_{j=1}^{T}\mathrm{sgn}(T^{-1/2}x_{Tj} + K + K\frac{j}{T}) \\
&\xrightarrow{d} \int_0^1 \mathrm{sgn}(\lambda W(\xi) + K + K\xi)d\xi \\
&\equiv T_3(K)
\end{aligned}$$

and

$$\begin{aligned}
&\inf_{a<-K}\inf_{b<-K} T^{-1}\sum_{j=1}^{T}\mathrm{sgn}(T^{-1/2}x_{Tj} - a - b\frac{j}{T})\frac{j}{T} \\
&= T^{-1}\sum_{j=1}^{T}\mathrm{sgn}(T^{-1/2}x_{Tj} + K + K\frac{j}{T})\frac{j}{T} \\
&\xrightarrow{d} \int_0^1 \mathrm{sgn}(\lambda W(\xi) + K + K\xi)\xi d\xi \\
&\equiv T_4(K).
\end{aligned}$$

As $K \to \infty$, $T_3(K) \xrightarrow{p} 1$ since $\mathrm{sgn}(\lambda W(\xi) + K + K\xi) \to \mathrm{sgn}(\infty) = 1$ and $T_4(K) \xrightarrow{p}$ 0.5 since $\int_0^1 \mathrm{sgn}(\lambda W(\xi) + K + K\xi)\xi d\xi \to \int_0^1 \xi d\xi = 0.5$. Then,

$$\begin{aligned}
&P(T^{-1/2}\hat{\alpha} < -K \text{ and } T^{-1/2}\hat{\beta} < -K) \\
&= P(\inf_{a<-K}\inf_{b<-K} T^{-1}\sum_{j=1}^{T}\mathrm{sgn}(T^{-1/2}x_{Tj} - a - b\frac{j}{T}) \leq 0)
\end{aligned}$$

$$+P(\inf_{a<-K}\inf_{b<-K} T^{-1}\sum_{j=1}^{T}\text{sgn}(T^{-1/2}x_{Tj} - a - b\frac{j}{T})\frac{j}{T} \leq 0)$$

$$\longrightarrow P(T_3(K) \leq 0) + P(T_4(K) \leq 0).$$

Since $T_3(K) \xrightarrow{p} 1$ and $T_4(K) \xrightarrow{p} 0.5$ as $K \to \infty$, $P(T_3(K) \leq 0) \to 0$ and $P(T_4(K) \leq 0) \to 0$ which implies (3.34). Therefore, conditional on proof of the other two cases, $T^{-1/2}\hat{\alpha}$ and $T^{-1/2}\hat{\beta}$ are $O_p(1)$. □

Proof of Theorem 2. Let

$$\begin{aligned} Q_T(\hat{\boldsymbol{\theta}}) &= \begin{pmatrix} Q_{1T}(\hat{\boldsymbol{\theta}}) \\ Q_{2T}(\hat{\boldsymbol{\theta}}) \end{pmatrix} \\ &= \begin{pmatrix} T^{-1}\sum_{j=1}^{T}\text{sgn}(T^{-1/2}x_{Tj} - T^{-1/2}\hat{\alpha} - T^{-1/2}\hat{\beta}\frac{j}{T}) \\ T^{-1}\sum_{j=1}^{T}\text{sgn}(T^{-1/2}x_{Tj} - T^{-1/2}\hat{\alpha} - T^{-1/2}\hat{\beta}\frac{j}{T})\frac{j}{T} \end{pmatrix}, \end{aligned} \tag{3.35}$$

where $\hat{\boldsymbol{\theta}}' = (\hat{\alpha}, \hat{\beta})'$. We rely on Theorem 2.7 of Kim and Pollard (1990) to ensure that $(T^{-1/2}\hat{\alpha}, T^{-1/2}\hat{\beta})'$ converge to the solutions to the asymptotic version of (3.35), $Q((A,B)')$ for some random variables A, B where

$$Q((A,B)') = \begin{pmatrix} \int_0^1 \text{sgn}(\lambda W(\xi) - A - B\xi)d\xi \\ \int_0^1 \text{sgn}(\lambda W(\xi) - A - B\xi)\xi d\xi \end{pmatrix}. \tag{3.36}$$

As noted in the proof of de Jong, Amsler, and Schmidt (2002), in order to use Theorem 2.7 in Kim and Pollard (1990), $|Q_T(\boldsymbol{\theta})|$ has to go to ∞ as $|\boldsymbol{\theta}| \to \infty$, which does not hold. But, this can be fixed by considering $\Psi^{-1}(|Q_T(\cdot)|)$ where Ψ is the cdf of normal distribution as $|Q_T(\cdot)|$ is bound between zero and one.

Note that for any $(\alpha, \beta)' \in \mathbb{R}^2$,

$$Q_T((\alpha,\beta)') \xrightarrow{d} Q((A,B)'). \tag{3.37}$$

Although this does not follow directly from the continuous mapping theorem due to the discontinuity of the sgn function, a continuous function arbitrarily close to the sgn function can be used in the place of the sgn function, which is the argument used in Park and Phillips (1999).

Now, we will prove the stochastic equicontinuity of $Q_T(\boldsymbol{\theta})$ on $\Phi = \{(\theta_1, \theta_2)' \in \mathbb{R}^2 : -K \leq \theta_1 \leq K, -K \leq \theta_2 \leq K\}$, thereby establishing that $Q_T(\boldsymbol{\theta}) \Rightarrow Q((A, B)')$ on Φ. As the equations below get longer, we define some notation for substitution:

$$
\begin{aligned}
\Xi_{1,Tj} &= \operatorname{sgn}(T^{-1/2}x_{Tj} - T^{-1/2}\alpha_{1T} - T^{-1/2}\beta_{1T}\frac{j}{T}) \\
&\quad - \operatorname{sgn}(T^{-1/2}x_{Tj} - T^{-1/2}\alpha_{1T} - T^{-1/2}\beta_{2T}\frac{j}{T}), \\
\Xi_{2,Tj} &= \operatorname{sgn}(T^{-1/2}x_{Tj} - T^{-1/2}\alpha_{1T} - T^{-1/2}\beta_{2T}\frac{j}{T}) \\
&\quad - \operatorname{sgn}(T^{-1/2}x_{Tj} - T^{-1/2}\alpha_{2T} - T^{-1/2}\beta_{2T}\frac{j}{T}), \\
\Lambda_{1,Tj} &= \operatorname{sgn}(T^{-1/2}x_{Tj} - a_1 - b_1\frac{j}{T}) - \operatorname{sgn}(T^{-1/2}x_{Tj} - a_1 - b_2\frac{j}{T}), \\
\Lambda_{2,Tj} &= \operatorname{sgn}(T^{-1/2}x_{Tj} - a_1 - b_2\frac{j}{T}) - \operatorname{sgn}(T^{-1/2}x_{Tj} - a_2 - b_2\frac{j}{T}).
\end{aligned}
$$

First, for $\boldsymbol{\theta} \in \Phi$ and $\boldsymbol{\theta}' \in \Phi$, we prove the stochastic equicontinuity of $Q_{1T}(\cdot)$ by showing that

$$
\begin{aligned}
&\lim_{\delta \to 0} \limsup_{n \to \infty} P\left(\sup_{\boldsymbol{\theta}, \boldsymbol{\theta}' : |\theta_1 - \theta_1'| < \delta; |\theta_2 - \theta_2'| < \delta} |Q_{1T}(\boldsymbol{\theta}) - Q_{1T}(\boldsymbol{\theta}')| > \epsilon \right) \\
&\leq \lim_{\delta \to 0} \limsup_{n \to \infty} P\left(\delta(1 + \xi) \sup_{s \in [-K, K]} |L(1, s)| > \epsilon \right) = 0.
\end{aligned}
\tag{3.38}
$$

Note that

$$
\begin{aligned}
&Q_{1T}((\alpha_{1T},\beta_{1T})') - Q_{1T}((\alpha_{2T},\beta_{2T})') \\
&= \left(Q_{1T}((\alpha_{1T},\beta_{1T})') - Q_{1T}((\alpha_{1T},\beta_{2T})')\right) \\
&\qquad + \left(Q_{1T}((\alpha_{1T},\beta_{2T})') - Q_{1T}((\alpha_{2T},\beta_{2T})')\right) \\
&= T^{-1}\sum_{j=1}^{T}\Xi_{1,Tj} + T^{-1}\sum_{j=1}^{T}\Xi_{2,Tj}.
\end{aligned}
\tag{3.39}
$$

By Conjecture 2, we can replace $T^{-1/2}\alpha_{1T}$ with $a_1 \in [-K,K]$: similarly, $T^{-1/2}\alpha_{2T}$ with $a_2 \in [-K,K]$, $T^{-1/2}\beta_{1T}$ with $b_1 \in [-K,K]$, and $T^{-1/2}\beta_{2T}$ with $b_2 \in [-K,K]$. Then, (3.39) becomes

$$
\begin{aligned}
&T^{-1}\sum_{j=1}^{T}\Lambda_{1,Tj} + T^{-1}\sum_{j=1}^{T}\Lambda_{2,Tj} \\
&\le \sup_{b_1\in[-K,K]}\ \sup_{b_2:|b_2-b_1|<\delta}\left|T^{-1}\sum_{j=1}^{T}\Lambda_{1,Tj}\right| \\
&\qquad + \sup_{a_1\in[-K,K]}\ \sup_{a_2:|a_2-a_1|<\delta}\left|T^{-1}\sum_{j=1}^{T}\Lambda_{2,Tj}\right|.
\end{aligned}
\tag{3.40}
$$

Now, by dividing an interval of $[-K,K]$ into sub-intervals of an equal length of δ, (3.40) can be written as

$$
\begin{aligned}
&\sup_{-\left[\frac{K}{\delta}\right]-1\le i\le\left[\frac{K}{\delta}\right]}\ \sup_{b_1\in[i\delta,(i+1)\delta]}\ \sup_{b_2:|b_2-b_1|<\delta}\left|T^{-1}\sum_{j=1}^{T}\Lambda_{1,Tj}\right| \\
&\quad + \sup_{-\left[\frac{K}{\delta}\right]-1\le i\le\left[\frac{K}{\delta}\right]}\ \sup_{a_1\in[i\delta,(i+1)\delta]}\ \sup_{a_2:|a_2-a_1|<\delta}\left|T^{-1}\sum_{j=1}^{T}\Lambda_{2,Tj}\right| \\
&\le \sup_{-\left[\frac{K}{\delta}\right]-1\le i\le\left[\frac{K}{\delta}\right]}\ \sup_{b_1\in[i\delta,(i+1)\delta]}\ \sup_{b_2:|b_2-b_1|<\delta} T^{-1}\sum_{j=1}^{T}|\Lambda_{1,Tj}|
\end{aligned}
$$

$$+ \sup_{-\left[\frac{K}{\delta}\right]-1\le i\le\left[\frac{K}{\delta}\right]} \sup_{a_1\in[i\delta,(i+1)\delta]} \sup_{a_2:|a_2-a_1|<\delta} T^{-1}\sum_{j=1}^{T}|\Lambda_{2,Tj}|$$

$$\le \sup_{-\left[\frac{K}{\delta}\right]-1\le i\le\left[\frac{K}{\delta}\right]} T^{-1}\sum_{j=1}^{T}\mathrm{I}\left(a_1+i\delta\frac{j}{T}\le T^{-1/2}x_{Tj}\le a_1+(i+1)\delta\frac{j}{T}\right)$$

$$+ \sup_{-\left[\frac{K}{\delta}\right]-1\le i\le\left[\frac{K}{\delta}\right]} T^{-1}\sum_{j=1}^{T}\mathrm{I}\left(i\delta+b_2\frac{j}{T}\le T^{-1/2}x_{Tj}\le (i+1)\delta+b_2\frac{j}{T}\right)$$

$$= \sup_{-\left[\frac{K}{\delta}\right]-1\le i\le\left[\frac{K}{\delta}\right]} \int_0^1 \mathrm{I}\left(a_1+i\delta\xi\le W_T\left(\frac{[\xi T]}{T}\right)\le a_1+(i+1)\delta\xi\right)d\xi$$

$$+ \sup_{-\left[\frac{K}{\delta}\right]-1\le i\le\left[\frac{K}{\delta}\right]} \int_0^1 \mathrm{I}\left(i\delta+b_2\xi\le W_T\left(\frac{[\xi T]}{T}\right)\le (i+1)\delta+b_2\xi\right)d\xi$$

$$= \sup_{-\left[\frac{K}{\delta}\right]-1\le i\le\left[\frac{K}{\delta}\right]} \left(\int_{a_1+i\delta\xi}^{a_1+(i+1)\delta\xi} L(1,s)ds+\int_{i\delta+b_2\xi}^{(i+1)\delta+b_2\xi} L(1,s)ds\right)$$

$$\le \delta\xi \sup_{s\in[-K,K]}|L(1,s)|+\delta\sup_{s\in[-K,K]}|L(1,s)|,$$

which can be made arbitrarily small as $\delta\to 0$. This proves (3.38). That is, $Q_{1T}(\cdot)$ is stochastically equicontinuous on a compact set Φ. In the last equality, we use the occupation times formula as in Park and Phillips (1999). $L(1,s)$, a local time is a continuous stochastic process of time spent by the Brownian motion at the spatial point s over the interval $[0,1]$, and $\sup_{s\in[-K,K]}|L(1,s)|$ is a well-defined random variable.

The stochastic equicontinuity of $Q_{2T}(\cdot)$ can be proved along the same lines as in the above.

$$Q_{2T}((\alpha_{1T},\beta_{1T})')-Q_{2T}((\alpha_{2T},\beta_{2T})')$$
$$=(Q_{2T}((\alpha_{1T},\beta_{1T})')-Q_{2T}((\alpha_{1T},\beta_{2T})'))$$
$$+(Q_{2T}((\alpha_{1T},\beta_{2T})')-Q_{2T}((\alpha_{2T},\beta_{2T})'))$$

$$
\begin{aligned}
&= T^{-1}\sum_{j=1}^{T}\Xi_{1,Tj}\frac{j}{T} + T^{-1}\sum_{j=1}^{T}\Xi_{2,Tj}\frac{j}{T} \\
&= T^{-1}\sum_{j=1}^{T}\Lambda_{1,Tj}\frac{j}{T} + T^{-1}\sum_{j=1}^{T}\Lambda_{2,Tj}\frac{j}{T} \\
&\leq \sup_{b_1\in[-K,K]}\ \sup_{b_2:|b_2-b_1|<\delta} T^{-1}\sum_{j=1}^{T}\left|\Lambda_{1,Tj}\right|\frac{j}{T} \\
&\quad + \sup_{a_1\in[-K,K]}\ \sup_{a_2:|a_2-a_1|<\delta} T^{-1}\sum_{j=1}^{T}\left|\Lambda_{2,Tj}\right|\frac{j}{T} \\
&\leq \sup_{b_1\in[-K,K]}\ \sup_{b_2:|b_2-b_1|<\delta} T^{-1}\sum_{j=1}^{T}\left|\Lambda_{1,Tj}\right| \\
&\quad + \sup_{a_1\in[-K,K]}\ \sup_{a_2:|a_2-a_1|<\delta} T^{-1}\sum_{j=1}^{T}\left|\Lambda_{2,Tj}\right|
\end{aligned}
$$

since $(j/T) \leq 1$ for all $t = 1, \cdots, T$. Then, the remaining lines of proof for the stochastic equicontinuity of $Q_{2T}(\cdot)$ follow from those in the proof of the stochastic equicontinuity of $Q_{1T}(\cdot)$.

Next, note that the finite-dimensional convergence of $T^{-1}\sum_{j=1}^{[\xi T]}\operatorname{sgn}(x_{Tj} - \hat{\alpha} - \hat{\beta}j/T)$ for each $\xi \in [0,1]$ holds because of a similar argument in (3.37) so that

$$
T^{-1}\sum_{j=1}^{[\xi T]}\operatorname{sgn}(T^{-1/2}x_{Tj} - T^{-1/2}\hat{\alpha} - T^{-1/2}\hat{\beta}\frac{j}{T}) \xrightarrow{d} \int_0^{\xi}\operatorname{sgn}(\lambda W(\xi) - A - B\xi)d\xi. \quad (3.41)
$$

From (3.41) and the stochastic equicontinuity implied by that of $Q_T(\cdot)$ in the above, the limiting distribution of $T^{-3}\sum_{j=1}^{T} S_{Tt}^2$ is as follows:

$$
\begin{aligned}
&T^{-3}\sum_{j=1}^{T}\left(\sum_{i=1}^{t}\operatorname{sgn}\left(x_{Ti} - \hat{\alpha} - \hat{\beta}\frac{i}{T}\right)\right)^2 \\
&= T^{-1}\sum_{j=1}^{T}\left(T^{-1}\sum_{i=1}^{t}\operatorname{sgn}\left(T^{-1/2}x_{Ti} - T^{-1/2}\hat{\alpha} - T^{-1/2}\hat{\beta}\frac{i}{T}\right)\right)^2
\end{aligned}
$$

$$\xrightarrow{d} \lambda^2 \int_0^1 \left(\int_0^\zeta \text{sgn} \left(W(\xi) - \frac{A}{\lambda} - \frac{B}{\lambda}\xi \right) d\xi \right)^2 d\zeta.$$

Finally, the estimate of long run variance, $\hat{\sigma}^2/\gamma_T$ is equal to

$$\begin{aligned}
&\gamma_T^{-1}T^{-1}\sum_{j=1}^{T} \text{sgn}\,(x_{Tj} - \hat{\alpha} - \hat{\beta}\frac{j}{T})^2 \\
&\quad + 2\gamma_T^{-1}T^{-1}\sum_{j=1}^{T} k(\frac{j}{\gamma_T}) \\
&\quad \times \sum_{i=1}^{T-j} \text{sgn}\,(x_{Ti} - \hat{\alpha} - \hat{\beta}\frac{i}{T})\,\text{sgn}\,(x_{T,(i+j)} - \hat{\alpha} - \hat{\beta}\frac{i+j}{T}) \\
&= o_p(1) + 2\gamma_T^{-1}\int_1^{T+1} k(\frac{[j]}{\gamma_T})dj \\
&\quad \times T^{-1}\int_1^{T-j+1} \Big[\text{sgn}(x_{T,[\xi T]} - \hat{\alpha} - \hat{\beta}\xi) \\
&\quad \times \text{sgn}\,(x_{T,[(\xi+\frac{\zeta\gamma_T}{T})T]} - \hat{\alpha} - \hat{\beta}(\xi + \frac{\zeta\gamma_T}{T}))\Big] dt \\
&= o_p(1) + 2\int_{\frac{1}{\gamma_T}}^{\frac{T}{\gamma_T}+\frac{1}{\gamma_T}} \Big[k(\frac{[\zeta\gamma_T]}{\gamma_T}) \\
&\quad \times \int_{\frac{1}{T}}^{1+\frac{1-\zeta\gamma_T}{T}} \Big(\text{sgn}(x_{T,[\xi T]} - \hat{\alpha} - \hat{\beta}\xi) \\
&\quad \times \text{sgn}\,(x_{T,[(\xi+\frac{\zeta\gamma_T}{T})T]} - \hat{\alpha} - \hat{\beta}(\xi + \frac{\zeta\gamma_T}{T}))\Big) d\xi \Big] d\zeta \\
&\to 2\int_0^\infty k(\zeta)\cdot\int_0^1 \text{sgn}(x_{T,[\xi T]} - \alpha - \beta\xi)^2 d\xi d\zeta \\
&= 2\int_0^\infty k(\zeta)d\zeta,
\end{aligned}$$

where the substitutions $(j/T) = \xi$ and $(j/\gamma_T) = \zeta$ are made. □

BIBLIOGRAPHY

Alvarez, A., C. Amsler, L. Orea, and P. Schmidt, 2005, Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics, *Journal of Productivity Analysis,* forthcoming.

Amsler, C., and P. Schmidt, 2000, Tests of short memory with thick tailed errors, Unpublished Manuscript, Department of Economics, Michigan State University.

Battese, G.E., and T.J. Coelli, 1988, Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data, *Journal of Econometrics* 38, 387–399.

Caudill, S.B., and J.M. Ford, 1993, Biases in frontier estimation due to heteroskedasticity, *Economics Letters* 41, 17–20.

——— , and D.M. Gropper, 1995, Frontier estimation and firm-specific inefficiency measures in the presence of heteroskedasticity, *Journal of Business and Economic Statistics* 13, 105–111.

Coelli, T.J., 1995, Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis, *Journal of Productivity Analysis* 6, 247–268.

Davidson, J., 1994, *Stochastic Limit Theory* (Oxford University Press: Oxford).

de Jong, R.M., C. Amsler, and P. Schmidt, 2002, A robust version of the KPSS test, based on indicators, Unpublished Manuscript, Department of Economics, Michigan State University.

de Jong, R.M., and J. Davidson, 2000, Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices, *Econometrica* 68, 407–423.

Efron, B., 1982, *The Jackknife, the Bootstrap and Other Resampling Plans* (Philadelphia: Society for Industrial and Applied Mathematics).

——— , 1985, Bootstrap confidence intervals for a class of parametric problems, *Biometrika* 72, 45–58.

——— , and R.J. Tibshirani, 1993, *An Introduction to the Bootstrap* (New York: Chapman and Hall).

Erwidodo, 1990, Panel data analysis on farm-level efficiency, input demand and output supply of rice farming in West Java, Indonesia, Ph.D. dissertation, Department of Agricultural Economics, Michigan State University.

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

Hall, P., W. Härdle, and L. Simar, 1993, On the inconsistency of bootstrap distribution estimators, *Computational Statistics and Data Analysis* 16, 11–18.

———, 1995, Iterated bootstrap with applications to frontier models, *Journal of Productivity Analysis* 6, 63–76.

Hansen, B.E., 1996, Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* 64, 413–430.

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

———, 2000, Multiple comparisons with the best, with economic applications, *Journal of Applied Econometrics* 15, 1–26.

Jondrow, J., C.A.K. Lovell, I.S. Materov, and P. Schmidt, 1982, On the estimation of technical efficiency in stochastic frontier production model, *Journal of Econometrics* 19, 233–238.

Kim, J., and D. Pollard, 1990, Cube root asymptotics, *Annals of Statistics* 18, 191–219.

Kim, Y., 1999, A study in estimation and inference on firm efficiency, Ph.D. dissertation, Department of Economics, Michigan State University.

———, and P. Schmidt, 1999, Marginal comparisons with the best and the efficiency measurement problem, Unpublished Manuscript, Department of Economics, Michigan State University.

Koop, G., J. Osiewalski, and M.F. Steel, 1997, Bayesian efficiency analysis through individual effects: Hospital cost frontiers, *Journal of Econometrics* 76, 77–106.

Kumbhakar, S.C., 1996, Estimation of cost efficiency with heteroscedasticity: An application to electric utilities, *Journal of the Royal Statistical Society, Series D (The Statistician)* 45, 319–335.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin, 1992, Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?, *Journal of Econometrics* 54, 159–178.

Lee, Y.H., 1991, Panel data models with multiplicative individual and time effects: Application to compensation and frontier production functions, Ph.D. dissertation, Department of Economics, Michigan State University.

———, and P. Schmidt, 1993, A Production Frontier Model with Flexible Temporal Variation in Technical Efficiency, In H.O. Fried, C.A.K. Lovell, and S.S. Schmidt (eds.), *The Measurement of Productive Efficiency* (New York: Oxford University Press).

Newey, W.K., 1991, Uniform convergence in probability and stochastic equicontinuity, *Econometrica* 59, 1161–1167.

Olson, J.A., P. Schmidt, and D. Waldman, 1980, A Monte Carlo study of estimators of stochastic frontier models, *Journal of Econometrics* 13, 67–82.

Osiewalski, J., and M. Steel, 1998, Numerical tools for the Bayesian analysis of stochastic frontier models, *Journal of Productivity Analysis* 10, 103–117.

Park, B.U., and L. Simar, 1994, Efficient semiparametric estimation in a stochastic frontier model, *Journal of the American Statistical Association* 89, 929–936.

Park, J.Y., and P.C.B. Phillips, 1999, Asymptotics for nonlinear transformations of integrated time series, *Econometric Theory* 15, 269–298.

Phillips, P.C.B., 1987, Time series regression with a unit root, *Econometrica* 55, 277–301.

——— , and P. Perron, 1988, Testing for a unit root in time series regression, *Biometrika* 75, 335–346.

Pitt, M.M., and L.F. Lee, 1981, The measurement and sources of technical inefficiency in the Indonesian weaving industry, *Journal of Development Economics* 9, 43–64.

Reifschneider, D., and R. Stevenson, 1991, Systematic departures from the frontier: A framework for the analysis of firm inefficiency, *International Economic Review* 32, 715–723.

Schmidt, P., and T.F. Lin, 1984, Simple tests of alternative specifications in stochastic frontier models, *Journal of Econometrics* 24, 349–361.

Schmidt, P., and R.C. Sickles, 1984, Production frontiers and panel data, *Journal of Business and Economic Statistics* 2, 367–374.

Simar, L., 1992, Estimating efficiencies from frontier models with panel data: A comparison of parametric, non-parametric and semi-parametric methods with bootstrapping, *Journal of Productivity Analysis* 3, 171–203.

——— , C.A.K. Lovell, and P. Vanden Eeckaut, 1994, Stochastic frontiers incorporating exogenous influences on efficiency, Discussion Papers No. 9403, Institutde Statistique, Université Catholique de Louvain.

Wang, H.J., and P. Schmidt, 2002, One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels, *Journal of Productivity Analysis* 18, 129–144.

White, H., 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48, 817–838.

Wooldridge, J.M., 2002, *Econometric Analysis of Cross Section and Panel Data* (The MIT Press: Cambridge, Massachusetts).