I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

0524

ESTIMATION AND INFERENCE IN CROSS-SECTIONAL, STOCHASTIC FRONTIER MODELS

L. SIMAR and P.W. WILSON

http://www.stat.ucl.ac.be

Estimation and Inference in Cross-Sectional, Stochastic Frontier Models

Léopold Simar Paul W. Wilson*

26 August 2005

Abstract

Conventional approaches for inference about efficiency in parametric stochastic frontier (PSF) models are based on percentiles of the estimated distribution of the one-sided error term, conditional on the composite error, rather than on the sampling distribution of the inefficiency estimator. When used as confidence intervals, these have extraordinarily poor coverage properties that do not improve with increasing sample size. We present a bootstrap method that gives confidence interval estimates based on sampling distributions, and which have good coverage properties that improve with sample size. In addition, researchers who estimate PSF models typically reject models, samples, or both when residuals have skewness in the "wrong" direction, *i.e.,* in a direction that would seem to indicate absence of inefficiency. We show that correctly specified models can generate samples with "wrongly" skewed residuals, even when the variance of the inefficiency process is nonzero. Our bootstrap method provides useful information about inefficiency and model parameters irrespective of whether residuals have the skewness in the desired direction. We also find that a commonly-used Wald test used to test the existence of inefficiency has catastrophic size properties; likelihood-ratio and bootstrap tests are shown in Monte Carlo experiments to perform well both in terms of size and power.

*Simar: Institut de Statistique, Université Catholique de Louvain, Voie du Roman Pays 20, B 1348 Louvain-la-Neuve, Belgium; email simar@stat.ucl.ac.be. Wilson: Department of Economics, University of Texas at Austin, 1 University Station C3100, Austin, Texas 78712 USA; email wilson@eco.utexas.edu. Research support from the "Interuniversity Attraction Pole", Phase V (No. P5/24) from the Belgian Government, and from the Texas Advanced Computing Center is gratefully acknowledged. Any remaining errors are solely our responsibility.

1 Introduction

Parametric stochastic frontier (PSF) models introduced by Aigner et al. (1977) and Meeusen and van den Broeck (1977) specify output, cost, *etc.* in terms of a response function and a composite error term. The composite error term consists of a two-sided error representing random effects and a one-sided term representing technical inefficiency. Since their introduction, several hundred papers, describing either methodological issues or empirical applications of these models, have appeared in the literature. Bauer (1990), Greene (1993), and Kumbhakar and Lovell (2000) provide overviews of developments in this area in varying levels of detail.

PSF models are typically estimated by the maximum likelihood (ML) method. Bauer (1990), Bravo-Ureta and Pinheiro (1993), and Coelli (1995) observe that most applied papers describe estimation of cross-sectional models with errors composed of normal and half-normal random variables. Interest typically lies in making inferences regarding (i) marginal effects, returns to scale, or other features of the response function; (ii) technical efficiency for individual firms, either real or hypothetical; or (iii) mean technical efficiency. With regard to (ii) and (iii), many papers have relied only on point estimates, although interval estimates are possible using a "conventional" approach suggested by Horrace and Schmidt (1996), where one obtains intervals based on percentiles of the estimated distribution of the one-sided error term, conditional on the composite error. These intervals, however, are not based on sampling distributions of the estimators of inefficiency, and consequently should not be expected to have good coverage properties. This is confirmed by our Monte Carlo experiments.

It is apparently common practice for applied researchers to reject models, samples, or both whenever residuals in PSF models have the "wrong" skewness (*e.g.,* positive skewness in the case of a production frontier, or negative skewness in the case of a cost frontier). In fact, even correctly specified models can, and often do, generate samples where residuals are skewed in the "wrong" direction. As demonstrated below, this happens even when the variance of the one-sided inefficiency process is nonzero. In such cases, for a linear response function, Waldman (1982) showed that the ML estimates are equivalent to ordinary least squares (OLS) estimates for the slope parameters and the variance of the two-sided error process, and that the ML estimate of the shape parameter for the one-sided process is zero.

The OLS variance-covariance matrix estimator is invalid, however, and the Hessian of the log-likelihood is singular when residuals have the "wrong" skewness.

We provide a bootstrap method that overcomes many of these difficulties, provided the true variance of the one-sided process is not zero. The method can be used for inference about either individual model parameters or functions of model parameters (*e.g.,* scale elasticities, *etc.*), as well as both mean inefficiency and inefficiency of individual firms. Monte Carlo experiments show that confidence intervals estimated by our bootstrap method have good coverage properties that improve as sample size increases. In addition, in cases where residuals have the "wrong" skewness, our method reveals useful information about model parameters, mean inefficiency, and inefficiency of individual firms. Our method provides such information even when the variance of the one-sided inefficiency process is estimated at zero.

Wald tests are commonly used to test the null hypothesis of no inefficiency, *i.e.,* that the variance of the one-sided process is zero. However, additional Monte Carlo experiments show that the size properties of this test are very poor. A small modification of our bootstrap procedure allows it to be used for testing the null hypothesis of no inefficiency, and the same Monte Carlo experiments show that a bootstrap test as well as a likelihood-ratio test perform well in terms of both size and power.[1]

Asymptotic properties of ML estimators are well-known, but (analytical) finite sample properties in the context of PSF models remain unknown. To date, only a few Monte Carlo studies with frontier models are available. Aigner et al. (1977) provided an examination of the performance of the ML estimator in small samples, while Olson et al. (1980) compared the ML estimator and a modified OLS (MOLS) estimator based on the idea of using moment conditions to correct the intercept term estimated by ordinary least squares. Both studies focused on mean-square error (MSE) properties of the parameter estimates; although limited by today's standards for Monte Carlo experimentation, these papers are remarkable for what they were able to do with the anemic computing power available in the 1970s.

More recently, Banker et al. (1988) compared ML estimators of technical efficiency with

[1]As noted above, our bootstrap does not provide valid confidence intervals when the variance of the one-sided process is zero, since the sampling distribution of one or more estimators will be discontinuous. The problem is similar to the one discussed by Andrews (2000), but does not affect one-sided tests of zero variance in the one-sided process.

non-parametric, data envelopment analysis (DEA) estimators in a Monte Carlo framework. Gong and Sickles (1990, 1992) made similar comparisons, but introduced panel data techniques. The comparisons with DEA estimators seem, in retrospect, rather curious, since DEA estimators are now known to be inconsistent in the presence of two-sided noise as in the stochastic frontier model (see Simar and Wilson, 2000, for discussion). Coelli (1995) considered MSE properties of ML parameter estimates as well as estimates of mean technical efficiency over a sample. Kumbhakar and Löthgren (1998) examine the coverages of classical confidence interval estimates for inefficiency, but failed to hold the points at which inefficiency was estimated constant over their Monte Carlo trials.

Our story develops as follows: notation and a PSF model are defined in Section 2; estimation of technical efficiency is also discussed. Section 3 provides a discussion of inference in PSF models, first about model parameters and then about technical efficiency. Our bootstrap method is described in Section 4, and Section 5 gives details of our Monte Carlo experiments and a discussion of the results. The final section concludes.

2 The Model

The stochastic frontier model can be written in general terms as

$$y = g(\boldsymbol{x} \mid \boldsymbol{\beta})e^{\varepsilon}, \tag{2.1}$$

$y \in \mathbb{R}^1_+$ is the scalar quantity of output produced from (exogenous) input quantities $\boldsymbol{x} \in \mathbb{R}^p_+$, and ε is composed of a two-sided error term v reflecting noise and a one-sided error term $u \geq 0$ reflecting technical inefficiency. We consider production functions and write

$$\varepsilon = v - u, \tag{2.2}$$

but of course the minus sign on the right-hand side of (2.2) could be changed to a plus sign to consider cost functions.

In applications, the two-sided error term is invariably assumed normally distributed:

$$v \sim N(0, \sigma_v^2). \tag{2.3}$$

Various distributions have been assumed for the one-sided term; *e.g.,* half-normal, truncated

(other than at zero) normal, gamma, exponential, *etc.*[2] We assume u is distributed half-normal on the non-negative part of the real number line:

$$u \sim N^{+}(0, \sigma_u^2). \tag{2.4}$$

Following common practice, we assume the v and u are each identically, independently distributed (iid).

The density of ϵ can be found using either the convolution theorem or characteristic functions, yielding

$$f(\varepsilon) = \frac{1}{\sigma\sqrt{\frac{\pi}{2}}}\left[1 - \Phi\left(\frac{\varepsilon}{\sigma}\sqrt{\frac{\gamma}{1-\gamma}}\right)\right]\exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right), \tag{2.5}$$

where $\sigma^2 = \sigma_v^2 + \sigma_u^2$, $\gamma = \sigma_u^2\sigma^{-2}$, and $\Phi(\cdot)$ denotes the standard normal distribution function. Note that the variance of ε is given by $\sigma_\varepsilon^2 = \left(\frac{\pi-2}{\pi}\right)\sigma_u^2 + \sigma_v^2$; the term σ^2 is only used to re-parameterize the model.

Given a sample $\mathcal{S}_n = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ of independent observations from the model represented by (2.1)–(2.4), the log-likelihood

$$\begin{aligned}\text{LLF} = &-\left(\frac{n}{2}\right)\log\left(\frac{\pi}{2}\right) - \left(\frac{n}{2}\right)\log\sigma^2 + \sum_{i=1}^{n}\log\left[1 - \Phi\left(\frac{[\log y_i - \log g(\boldsymbol{x}_i \mid \boldsymbol{\beta})]\gamma^{1/2}}{\sigma(1-\gamma)^{1/2}}\right)\right] \\ &-\frac{1}{2\sigma^2}\sum_{i=1}[\log y_i - \log g(\boldsymbol{x}_i \mid \boldsymbol{\beta})]^2 \end{aligned} \tag{2.6}$$

can be maximized with respect to $\boldsymbol{\beta}$, σ^2, and γ to obtain estimates $\widehat{\boldsymbol{\beta}}$, $\widehat{\sigma}^2$, and $\widehat{\gamma}$. ML estimates of σ_u^2 and σ_v^2 can be recovered from the relationships $\widehat{\sigma}_u^2 = \widehat{\gamma}\widehat{\sigma}^2$ and $\widehat{\sigma}_v^2 = \widehat{\sigma}^2 - \widehat{\sigma}_u^2$.

Here, we use the parameterization of Battese and Corra (1977); alternatively, the density in (2.5) and the log-likelihood in (2.6) may be written in terms of σ^2 and $\lambda = \frac{\sigma_u}{\sigma_v}$, which was the parameterization used by Aigner et al. (1977). Note that $\lambda^2 = \frac{\gamma}{1-\gamma}$. The parameterization used here has the advantage that $\gamma \in [0, 1]$, while $\lambda \in [0, \infty)$; using γ simplifies numerical maximization of the log-likelihood function.

[2]The choice of one-sided distribution is not innocuous. If the one-sided distribution for u can be made to resemble a normal distribution with certain parameter combinations, estimation can be very difficult with sample sizes commonly found in the literature. This was demonstrated for a normal/gamma convolution by Ritter and Simar (1997), but the problem could also be expected if the distribution for u was specified as truncated normal, with the location parameter to be estimated (instead of constrained to zero, as in the half-normal distribution).

Now consider an arbitrary, given point $(\boldsymbol{x}, y) \in \mathbb{R}^p_+ \times \mathbb{R}^1_+$; this point might be one of the observations in $\mathcal{S}_n$, or some other point of interest, perhaps representing input and output quantities of an hypothetical firm. Associated with this point is an error, $\varepsilon = \log y - \log g(\boldsymbol{x} \mid \boldsymbol{\beta})$. If $(\boldsymbol{x}, y)$ is an actual observation, then from (2.2) ε is determined by realizations of the random variables v and u. Otherwise, if $(\boldsymbol{x}, y)$ represents a hypothetical firm, it is appropriate to think of ε as the result of hypothetical draws from $N(0, \sigma_v^2)$ and $N^+(0, \sigma_u^2)$.

Battese and Coelli (1988) note that it is important to clearly define an appropriate measure of technical efficiency that is to be estimated, and that the literature is muddled on this point. Some (*e.g.,* Jondrow et al., 1982) have defined technical inefficiency as "the shortfall of output (y_i) from its maximum possible value"; in ratio form, this amounts to defining technical inefficiency for the ith firm as e^{u_i}. In the context of cross-sections, since realizations of e^u cannot be observed, it is sensible to define technical inefficiency corresponding to the point $(\boldsymbol{x}, y)$ as the conditional expectation

$$\tau \equiv E\left(e^{-u} \mid \varepsilon\right). \tag{2.7}$$

Since $u \geq 0$, $\tau \in (0, 1]$. Unlike e^u, which is not identified, the quantity in (2.7) can be estimated consistently by the ML method. A firm with technical efficiency equal to τ is expected to produce a level of output that is $\tau \times 100$-percent of the technically efficient level of output corresponding to the levels of inputs used by this firm. The error ε is estimated by the residual

$$\widehat{\varepsilon} = \log y - \log g(\boldsymbol{x} \mid \widehat{\boldsymbol{\beta}}). \tag{2.8}$$

Some algebra reveals that the conditional density of u, given ε, is

$$f(u \mid \varepsilon) = \begin{cases} \sigma_*^{-1}\phi\left(\frac{u-\mu_*}{\sigma_*}\right)\left[1 - \Phi\left(\frac{-\mu_*}{\sigma_*}\right)\right]^{-1} & \forall\ u \geq 0; \\ 0 & \text{otherwise,} \end{cases} \tag{2.9}$$

where

$$\mu_* = -\frac{\varepsilon\sigma_u^2}{\sigma^2} = -\varepsilon\gamma, \tag{2.10}$$

$$\sigma_*^2 = \frac{\sigma_u^2\sigma_v^2}{\sigma^2} = \sigma^2\gamma(1-\gamma), \tag{2.11}$$

and $\phi(\cdot)$ denotes the standard normal density function. Hence, conditional on ε, u is distributed $N^+(\mu_*, \sigma_*^2)$.

The conditional density in (2.9) can be used to derive

$$E(u \mid \varepsilon) = \mu_* + \sigma_* \phi\left(\frac{-\mu_*}{\sigma_*}\right)\left[1 - \Phi\left(\frac{-\mu_*}{\sigma_*}\right)\right]^{-1}, \tag{2.12}$$

as in Jondrow et al. (1982), who regard estimates of this conditional expectation as a point estimate for u. However, replacing the unknown parameters in (2.12) with ML estimates yields an ML estimate of $E(u \mid \varepsilon)$, which is different from u. Some authors have estimated technical efficiency by $\widetilde{\tau} = \exp\left(-\widehat{E}(u \mid \widehat{\varepsilon})\right)$, obtained by substituting $\widehat{\varepsilon}$, $\widehat{\sigma}_u^2$, $\widehat{\sigma}_v^2$, and $\widehat{\sigma}^2$ for the corresponding true values ε, σ_u^2, σ_v^2, and σ^2 in (2.10)–(2.11), and then substituting the resulting estimates $\widehat{\mu}_*$ and $\widehat{\sigma}_*$ into (2.12) to obtain $\widehat{E}(u \mid \widehat{\varepsilon})$. Alternatively, it is easy to compute

$$\tau = E\left(e^{-u} \mid \varepsilon\right) = \left[1 - \Phi\left(\sigma_* - \frac{\mu_*}{\sigma_*}\right)\right]\left[1 - \Phi\left(-\frac{\mu_*}{\sigma_*}\right)\right]^{-1} \exp\left(-\mu_* + \frac{\sigma_*^2}{2}\right),$$

which appears in Battese and Coelli (1988). This leads to the ML estimator

$$\widehat{\tau} = \widehat{E}\left(e^{-u} \mid \widehat{\varepsilon}\right) \tag{2.13}$$

of τ obtained by replacing μ_* and σ_*^2 in (2.13) with ML estimators $\widehat{\mu}_*$, and $\widehat{\sigma}_*^2$ (recall that μ_* is defined in terms of ε in (2.10); hence $\widehat{\mu}_*$ depends on $\widehat{\varepsilon}$).

The expectation operator of course has only linear properties; *i.e.,* $E\left(e^u \mid \varepsilon\right) \neq e^{E(u|\varepsilon)}$. The difference between $\exp\left[E(u \mid \varepsilon)\right]$ and $E\left(e^{-u} \mid \varepsilon\right)$ can be substantial. For example, if $\sigma_u^2 = \sigma_v^2 = 1$, then $e^{-E(u|\varepsilon)}/E\left(e^{-u} \mid \varepsilon\right) = 0.6644$, 0.5255, and 0.3798 for $\varepsilon = -1$, 0, and 1, respectively. Consequently, we shall focus on the second estimator, $\widehat{\tau}$, in all that follows.

3 Classical Inference in the Stochastic Frontier Model

3.1 Inference about model parameters:

The standard approach to inference in stochastic frontier models relies on asymptotic normality in the case of the parameter estimates, and estimated percentiles of the distribution of e^{-u} conditional on $\widehat{\varepsilon}$ in the case of technical efficiency. The presence of a boundary in the support of u poses some problems. Lee (1993) derives the asymptotic distribution on the maximum likelihood estimator for the case where $\gamma = 0$ (implying $\sigma_u^2 = 0$). Lee observes (i) that the information matrix is singular in this case; (ii) that the likelihood ratio statistic for

testing the null hypothesis $\gamma = 0$ is asymptotically distributed as $\frac{1}{2}\chi^2(0) + \frac{1}{2}\chi^2(1)$, *i.e.,* as a mixture of a degenerate chi-square distribution with a unit mass at zero and a chi-square distribution with one degree of freedom; and (iii) the convergence rates of the estimators of the intercept term and of γ are slower than the usual $O\left(n^{-1/2}\right)$ parametric rate. For cases where $\gamma > 0$ but close to zero, Lee shows that the first-order normal approximation for γ can be poor as γ approaches zero.[3]

Unfortunately, even if γ is not close to zero, similar problems arise in finite samples. Aigner et al. (1977) noted that in some finite samples, the composite residuals in (2.1) may have positive skewness. In such cases, the ML estimator of γ will be zero, while the ML estimators of σ^2 and slope parameters will be equal to the corresponding OLS estimators in the case of a linear response function, as noted by Waldman (1982). This was discussed briefly in Olson et al. (1980) and confirmed in their Monte Carlo experiments. Both Aigner *et al.* and Olson et al. (1980) as well as others have referred to this as a "failure" of the ML estimator, but this terminology is misleading. As Lee (1993) notes, the estimators are identifiable, even though the negative of the Hessian of the log-likelihood in (2.6) cannot be inverted when $\gamma = 0$ (or when γ is replaced by an estimate $\widehat{\gamma} = 0$) due to singularity. The problem is not a failure of the ML method *per se,* but rather a difficulty for inference.

A simple Monte Carlo experiment illustrates that the problem can be severe for reasonable, plausible values of the model parameters. Table 1 shows the proportion among 1,000 samples of the composite error ε from the normal/half-normal convolution with density given by (2.5) resulting in positive skewness within a given sample, for various samples sizes n and ratios $\lambda^2 = \sigma_u^2/\sigma_v^2$. The results in Table 1 show, for example, that for $\lambda^2 = 1$, almost 1/3 of samples of size $n = 100$ will have skewness in the "wrong" direction. Of course, the problem goes away asymptotically, but at a rate that depends on the value of λ^2. For the case where $\lambda^2 = 1$, even with $n = 1,000$ observations, about 3.5 percent of samples will have positive skewness. If $\lambda^2 = 0.5$—which is far from implausible in applied studies—even with 1,000 observations, one should expect that about 22.5 percent of samples will have positive skewness. For smaller values of λ^2 the problem becomes even more severe. At $\lambda^2 = 0.1$, even with 100,000 observations, about 20 percent of samples will exhibit positive skewness.

[3]Lee (1993) works in terms of the parameterization of Aigner et al. (1977), using λ instead of γ. However, the results described here hold after straightforward substitution.

The results in Table 1 imply that in finite samples, the sampling distribution of the ML estimators for parameters in the model (2.1) are necessarily far from normal in many cases. For example, if $\lambda^2 = 1$ and $n = 100$, the sampling distribution of the ML estimator $\widehat{\gamma}$ will have a mass of approximately 0.3 at zero. Of course, the sampling distribution is normal *asymptotically* provided γ is not equal to either 0 or 1. But, the results in Table 1 indicate that asymptotic normal approximations are, in many cases, not particularly useful for inference about the parameters of the model when samples are finite, even if γ is not very close to 0. In fact, conventional inference is unavailable in many cases, since the negative Hessian of the log-likelihood in (2.5) is singular whenever γ is replaced by an estimate equal to zero. Again for the case where $\lambda^2 = 1$, $n = 100$, conventional inference based on asymptotic normality is undefined for almost 1/3 of all cases.

As noted in Section 1, it is apparent that when applied researchers encounter residuals with the "wrong" skewness, they typically either (i) obtain a new sample, or (ii) to re-specify the model. Indeed, the user's manual for the widely-used LIMDEP software package (Greene, 1995, Section 29.3) advises users that "If this condition emerges, your model is probably not well specified or the data are inconsistent with the model." Moreover, we know of no published papers reporting an estimate of zero for the variance parameter of the one-sided component in a stochastic frontier model. Yet, the results in Table 1 make it clear that even when the PSF being estimated is the *true* model, one should expect to sometimes obtain estimates $\widehat{\gamma} = 0$ (or the equivalent in models where other one-sided distributions are specified).

Presumably, the researcher who seeks to estimate a model such as the one in (2.1)–(2.4) must believe that the model is at least a good approximation of the underlying data-generating process (DGP), if not the correct specification. Otherwise, he would estimate a different model. Changing the model specification simply because residuals are skewed in the wrong direction is likely to lead to a mis-specification, particularly when the DGP produces finite samples yielding positively skewed residuals with some non-zero frequency. Moreover, it is well-known that classical inference assumes that the model specification is chosen independently of any estimates that are obtained; specification-searching introduces problems of bias in both parameter estimates as well as variance-covariance estimates (see Leamer, 1978 for discussion).

Note that if one were to discard estimates $\widehat{\gamma} = 0$ and draw a new sample without changing the model specification, until an estimate $\widehat{\gamma} > 0$ is obtained, this would be tantamount to imposing a complicated conditioning on the underlying model. *This point is crucial.* If the true model is the one represented by (2.1)–(2.4), where the v and u are iid, then we have seen that unless the variance ratio λ^2 is sufficiently large, one should expect to sometimes draw samples with positive skewness in the composite residuals with sample sizes of the same order as those used in published applications. If one were to believe that the true model can only generate finite samples with negatively-skewed residuals, then such a model would be very different from the ones that have been estimated, and different from the one in (2.1)–(2.4). Indeed, to ensure that only samples with negative skewness can be generated, one would have to impose a complicated correlation structure that would ensure that whenever a large, positive value v is drawn, a sufficiently large value of u is also drawn. We know of no instance in the literature where such a model has been estimated. On the contrary, models such as (2.1)–(2.4) have been estimated, and this model is clearly capable of generating samples with positive skewness; moreover, it will do so with a frequency that depends on λ^2 and n. More importantly, values of λ^2 and n that are plausible in applications result in relatively high frequencies of samples with positive skewness.

If a model such as (2.1)–(2.4) is to be estimated, and the researcher has the bad luck to draw a sample that yields positively-skewed residuals, the proper action would be to increase the sample size. Positively skewed residuals should not be taken as evidence that the model is misspecified.

Two software packages are commonly used to estimate stochastic frontier models: (i) the commercial package known as LIMDEP (Greene, 1995), and (ii) a freeware package known as FRONTIER (Coelli, 1996); see Sena (1999) and Herrero and Pascoe (2002) for reviews of these packages. The two packages are rather different in their treatment of cases where composite residuals have the "wrong" skewness. In both packages, ordinary least squares (OLS) estimates are first obtained to serve as starting values after adjusting the intercept and variance terms using the MOLS estimator. In the case of LIMDEP, when the OLS residuals have positive skewness, the program stops with a message stating, "Stoch. Frontier: OLS residuals have wrong skew. OLS is MLE." While it is indeed true that the ML parameter estimates of $\boldsymbol{\beta}$ and σ^2 are equivalent to the OLS estimates in such cases,

the OLS standard error estimates that are reported should not be taken as estimates of the standard error of the ML estimates. The OLS standard error estimates are conditioned on $\gamma = 0$, and consequently understate the true standard errors since uncertainty about γ is ignored.[4] Indeed, conventional standard error estimates of the ML estimates are unavailable due to singularity of the negative Hessian of the log-likelihood in this case.

With FRONTIER, estimation proceeds even if the OLS residuals have positive skewness. After the OLS estimates have been obtained, a grid search procedure is used to find a starting value for γ; then these starting values are used in the DFP algorithm (Davidon, 1959; Fletcher and Powell, 1963). If the OLS residuals have positive skewness, FRONTIER returns a very small estimate for γ, but typically not zero. In addition, FRONTIER does not use the inverse negative Hessian to estimate the variance-covariance matrix, but rather the DFP direction matrix, which is an approximation of the inverse negative Hessian. The DFP algorithm is based in part on approximating the objective function by a quadratic; the accuracy of the approximation of the Hessian by the DFP direction matrix will suffer if the algorithm iterates only a few times or if the objective function is far from quadratic. Inattentive users may be misled by the fact that FRONTIER returns estimates of variance for the parameter estimates in all cases, even though these are clearly invalid when the OLS residuals have positive skewness.

Another common practice when estimating PSF models is to test for the existence of inefficiency. In the model in Section 2, this is equivalent to testing $H_0 : \gamma = 0$ against $H_1 : \gamma > 0$. Coelli (1995, p. 251) notes that

> "The first test of this hypotheses was reported in Aigner et al. (1977), where the ratio of the ML estimate of σ_u^2 to its estimated standard error was observed to be quite small.... This Wald test, or a slight variant, has been explicitly or implicitly conducted in almost every application of this stochastic frontier model since this first application."

Indeed, Coelli (1995) reports results obtained for Monte Carlo experiments using code

[4]Neither the output from the LIMDEP program nor the accompanying manual suggest that the OLS standard error estimates should be taken as estimates of the standard error of the ML estimates when the OLS residuals are positively skewed. Nor do we know of any cases where one has done so; rather, as discussed earlier, we expect that many applied users are tempted to draw a new sample at this point, or to re-specify their model, even though residuals with positive skewness can easily arise when (2.1)–(2.4) is the correct specification.

from the FRONTIER package with numbers of observations $n \in \{50,\ 100,\ 400,\ 800\}$ and 11 different values of γ ranging from 0 to 1. Rejection rates for Wald tests of the null hypothesis $H_0 : \gamma = 0$ are reported. To implement the Wald tests, estimates of the standard error for $\widehat{\gamma}$ are needed, and Coelli (1995, p. 251) reports that these were approximated by taking the square root of the appropriate diagnonal element of the DFP direction matrix. The entire exercise is curious, since under the null, the Cramér-Rao regularity conditions are not satisfied and the Hessian of the log-likelihood is singular as discussed above.[5] With little surprise, Coelli (1995) finds that this Wald test has poor size properties.[6]

3.2 Inference regarding technical efficiency

The problems surrounding inference about model parameters also affect inferences about technical efficiency. In addition, making inferences about inefficiency presents further problems as discussed below. In applications, researchers are typically interested in inefficiencies corresponding to individual firms as well as the mean level of inefficiency.

Inefficiencies for specific firms are estimated by $\widehat{\tau}$ defined in (2.13); these estimates are necessarily conditional on an estimated residual $\widehat{\varepsilon}$. Recall from (2.9) that $u \mid \varepsilon \sim N^+(\mu_*, \sigma_*^2)$. Simple algebra reveals that percentiles ρ_α defined by $\Pr(u \leq \rho_\alpha \mid \varepsilon) = \alpha$ are given by

$$\rho_\alpha(\varepsilon) = \mu_* + \sigma_* \Phi^{-1}\left[1 - (1-\alpha)\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right], \tag{3.1}$$

where $\Phi^{-1}(\cdot)$ denotes the standard normal quantile function. Necessarily, the interval $\big(\rho_{\alpha/2}(\varepsilon),\ \rho_{1-\alpha/2}(\varepsilon)\big)$ or

$$\left(\mu_* + \sigma_* \Phi^{-1}\left[1 - \frac{\alpha}{2}\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right],\ \ \mu_* + \sigma_* \Phi^{-1}\left[1 - \left(1 - \frac{\alpha}{2}\right)\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right]\right) \tag{3.2}$$

gives a $(1-\alpha) \times 100$-percent *probability* interval for $u \mid \epsilon$, meaning that asymptotically, $(1-\alpha) \times 100$-percent of all draws from $N^+(\mu_*, \sigma_*^2)$ will fall within this interval. Since the exponential function is monotonic, it also follows that

$$\left(\exp\left\{-\mu_* - \sigma_* \Phi^{-1}\left[1 - \frac{\alpha}{2}\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right]\right\},\ \exp\left\{-\mu_* - \sigma_* \Phi^{-1}\left[1 - \left(1 - \frac{\alpha}{2}\right)\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right]\right\}\right) \tag{3.3}$$

[5] Note that Aigner et al. (1977) stopped short of explicitly testing σ_u^2, perhaps because they recognized the violation of regularity conditions.

[6] In principle, one could use the re-parameterization of Lee (1993) to make inference about γ, but there is no mention of this in Coelli (1995).

gives an $\alpha \times 100$-percent probability interval for $e^{-u} \mid \varepsilon$.

Horrace and Schmidt (1996, pp. 261–262) describe the intervals in (3.2) and (3.3) as $\alpha \times 100$-percent *confidence intervals* for $u \mid \varepsilon$ and $\tau = E\left(e^{-u} \mid \varepsilon\right)$. The intervals can be estimated by substituting estimates $\widehat{\mu}_*$, and $\widehat{\sigma}_*^2$ for the corresponding true values in (3.3). Hjalmarsson et al. (1996) and Bera and Sharma (1999) reported estimates of (3.2); Bera and Sharma (1999) also estimated the interval in (3.3).

Mean efficiency for the model defined by (2.1)–(2.4) is given by

$$\begin{aligned} \overline{\tau} = E(\tau) = E\left(e^{-u}\right) &\equiv \int_0^\infty e^{-u} f(u)\, du \\ &= 2e^{\frac{1}{2}\sigma_u^2}\left[1 - \Phi(\sigma_u)\right], \end{aligned} \tag{3.4}$$

which is estimated by replacing σ_u with $\widehat{\sigma}_u$ in (3.4) to obtain $\widehat{\overline{\tau}}$. Using the assumption in (2.4) that u is distributed half-normal to solve $\Pr(u \leq \rho_\alpha) = \alpha$ yields the $\alpha \times 100$ percentile of u, *i.e.*,

$$\rho_\alpha = \sigma^u \Phi^{-1}\left(\frac{1+\alpha}{2}\right). \tag{3.5}$$

Following the approach of Horrace and Schmidt (1996), one might claim that an estimator of the interval

$$\left[\exp\left(-\sigma_u \Phi^{-1}\left(1 - \frac{\alpha}{4}\right)\right),\ \exp\left(-\sigma_u \Phi^{-1}\left(\frac{1}{2} + \frac{\alpha}{4}\right)\right)\right]. \tag{3.6}$$

gives an $\alpha \times 100$-percent confidence interval for $\widehat{\overline{\tau}}$; this interval can be estimated by replacing σ_u with an estimate $\widehat{\sigma}_u$ in (3.6). As with the intervals in (3.2) or (3.3), however, the interval in (3.6) is not based the sampling distribution of $\widehat{\overline{\tau}}$. Instead, (3.6) gives an interval within which the random variable e^{-u} will fall with probability $(1 - \alpha)$.

It is important to recall that the quantities of interest are $E\left(e^{-u} \mid \varepsilon\right)$ in the case of firm-specific inefficiency and $E\left(e^{-u}\right)$ in the case of mean inefficiency. The intervals in (3.3) and (3.6) describe something different, however: they are based on percentiles of the distributions of $\left(e^{-u} \mid \varepsilon\right)$ and e^{-u}, respectively, instead of the sampling distributions of the estimators $\widehat{\tau}$ and $\widehat{\overline{\tau}}$. *This distinction is important.* To provide an illustration by analogy, consider the more familiar example where $z \sim N(\mu_z, \sigma_z^2)$; the ML estimator of $\mu_z = E(z)$ is $\widehat{\mu}_z = \widehat{E}(z) = n^{-1}\sum_{i=1}^n z_i$, with $\widehat{\mu}_z \sim N\left(\mu_z, \frac{\sigma_z^2}{n}\right)$. Using the intervals in (3.3) or (3.6) is analogous to using the $\frac{\alpha}{2} \times 100$- $\left(1 - \frac{\alpha}{2}\right) \times 100$-percentiles of $N(\mu_z, \sigma_z^2)$ instead of $N\left(\mu_z, \frac{\sigma_z^2}{n}\right)$ to form an

$\alpha \times 100$-percent confidence interval for μ_z. The distribution of $\widehat{E}(z)$ is not the same as the distribution of the z, just as the distributions of estimators of $E\left(e^u \mid \varepsilon\right)$ and $E\left(e^u\right)$ are not the same as the distributions of $(e^u \mid \varepsilon)$ and e^u.

The classical approach to inference about inefficiency suffers from at least two additional problems. First, interval estimates obtained from (3.3) and (3.6) are based on the idea of replacing the true, but unknown parameters with estimates. The resulting interval estimates do not reflect the uncertainty about the original parameters. Second, the unknown conditioning event ε is replaced by one observed residual $\widehat{\varepsilon}$ to construct $\widehat{\mu}_*$. Consequently, it is reasonable to expect the estimator $\widehat{\mu}_*$ to be very sensitive to the noise (with variance σ_v^2) contained in v in (2.2).

As seen in the next section, the bootstrap is well-suited to overcome some of these difficulties.

4 Inference Using the Bootstrap

Given a point $(\boldsymbol{x}_0, y_0) \in \mathbb{R}_+^p \times \mathbb{R}_+^1$, the corresponding efficiency estimate based on (2.13) is simply a function of parameter estimates and $(\boldsymbol{x}_0, y_0)$. The point $(\boldsymbol{x}_0, y_0) \in \mathbb{R}_+^p \times \mathbb{R}_+^1$ could correspond to an observation in the sample data $\mathcal{S}_n = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$, or it could represent some other point of interest (*e.g.,* the mean or median of observations in $\mathcal{S}_n$). For purposes of the discussion that follows, it is useful to write the efficiency estimate as $\widehat{\tau} = \tau(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2, \widehat{\gamma} \mid \boldsymbol{x}_0, y_0)$, and to write the quantity in (2.13) that is to be estimated as $\tau = \tau(\boldsymbol{\beta}, \sigma^2, \gamma \mid \boldsymbol{x}_0, y_0)$. In any applied setting, the point $(\boldsymbol{x}_0, y_0)$ will be known, while $\boldsymbol{\beta}$, σ^2, and γ are unknown and consequently must be estimated. Finally, write mean efficiency defined in (3.4)—which depends only on σ_u^2—as $\overline{\tau}(\sigma_u^2)$, and write the corresponding estimator as $\widehat{\overline{\tau}} = \overline{\tau}(\widehat{\sigma}_u^2)$.

It is straightforward to implement parametric bootstrap methods in the case of stochastic frontier models. For the model defined by (2.1)–(2.4), a parametric bootstrap consists of the following steps:

[1] Using the sample data $\mathcal{S}_n = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$, maximize the log-likelihood in (2.6) to obtain estimates $\widehat{\boldsymbol{\beta}}$, $\widehat{\sigma}^2$, and $\widehat{\gamma}$; recover $\widehat{\sigma}_v^2$, $\widehat{\sigma}_u^2$ from $\widehat{\sigma}^2$ and $\widehat{\gamma}$.

[2] For $i = 1, \ldots, n$ draw $v_i^* \sim N(0, \widehat{\sigma}_v^2)$ and $u_i^* \sim N^+(0, \widehat{\sigma}_u^2)$, and compute $y_i^* = g(\boldsymbol{x}_i \mid \widehat{\boldsymbol{\beta}}) e^{v_i^* - u_i^*}$.

[3] Using the pseudo-data $\mathcal{S}_n^* = \{\boldsymbol{x}_i,\ y_i^*\}_{i=1}^n$, obtain bootstrap estimates $\widehat{\boldsymbol{\beta}}^*$, $\widehat{\sigma}^{*2}$, and $\widehat{\gamma}^*$ by maximizing (2.6) after replacing $\log y_i$ with $\log y_i^*$.

[4] Repeat steps [2]–[3] B times to obtain estimates $\mathcal{B} = \{(\widehat{\boldsymbol{\beta}}_b^*,\ \widehat{\sigma}_b^{*2},\ \widehat{\gamma}_b^*)\}_{b=1}^B$.

The estimates obtained in step [1] can be used to compute estimates $\widehat{\tau} = \tau(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2, \widehat{\gamma} \mid \boldsymbol{x}_0, y_0)$ and $\widehat{\overline{\tau}} = \overline{\tau}(\widehat{\sigma}_u^2)$ of $\tau = \tau(\boldsymbol{\beta}, \sigma^2, \gamma \mid \boldsymbol{x}_0, y_0)$ and $\overline{\tau}(\widehat{\sigma}_u^2)$, respectively. Similarly, the bootstrap estimates obtained in $\mathcal{B}$ can be used to compute bootstrap estimates $\widehat{\tau}_b^* = \tau(\widehat{\boldsymbol{\beta}}_b^*, \widehat{\sigma}_b^{*2}, \widehat{\gamma}_b^* \mid \boldsymbol{x}_0, y_0)$ and $\widehat{\overline{\tau}}_b^* = \overline{\tau}(\widehat{\sigma}_u b^{*2})$. If one wishes to make inference about σ_u^2 and σ_v^2, bootstrap values σ_{ub}^{*2}, σ_{vb}^{*2} can be recovered from the bootstrap estimates in $\mathcal{B}$.

The bootstrap estimates in $\mathcal{B}$ (or the additional bootstrap estimates of inefficiency) can be used to estimate confidence intervals for the parameters of the model or for the efficiencies in any of several ways. We consider two possibilities. To illustrate, consider an initial estimate $\widehat{\theta}$ of a quantity θ, and a corresponding set of bootstrap estimates $\{\widehat{\theta}_b^*\}_{b=1}^B$. For nominal size α, the percentile intervals discussed by Efron (1979, 1982) are given by $\left(\widehat{\theta}_{(\frac{\alpha}{2})},\ \widehat{\theta}_{(\frac{1-\alpha}{2})}\right)$ where $\widehat{\theta}_{(\alpha)}$ denotes the $\alpha \times 100$-percentile of the elements of $\{\widehat{\theta}_b^*\}_{b=1}^B$. Alternatively, bias-corrected (BC) intervals described by Efron and Tibshirani (1993) are given (again, for nominal size α) by $\left(\widehat{\theta}_{(\alpha_1)},\ \widehat{\theta}_{(\alpha_2)}\right)$ where $\alpha_1 = \Phi\left(2\widehat{z}_0 + z^{\frac{\alpha}{2}}\right)$, $\alpha_2 = \Phi\left(2\widehat{z}_0 + z^{1-\frac{\alpha}{2}}\right)$, $\widehat{z}_0 = \Phi^{-1}\left(\frac{\#\{\widehat{\theta}_b^* < \widehat{\theta}\}}{B}\right)$.

The idea of the bootstrap method is to approximate the unknown distribution $(\widehat{\theta} - \theta)$ by an empirical approximation of the distribution of $(\widehat{\theta}^* - \widehat{\theta}) \mid \mathcal{S}_n$. Consequently, in the case of inefficiency estimates, the bootstrap is able to estimate confidence intervals based on the sampling distribution of the actual estimators, as opposed to the distributions of $e^{-u} \mid \varepsilon$ or e^{-u} (recall the discussion following (3.6)). Moreover, the bootstrap explicitly allows for uncertainty in the parameter estimates, which is ignored in the classical method when estimates are substituted for true values in (3.3) and (3.6).

5 Monte Carlo Experiments and Results

5.1 Design of Experiments

In our Monte Carlo experiments, we take (2.1)–(2.4) as the "true" model, with $g(x \mid \boldsymbol{\beta}) = e^{\beta_1} x^{\beta_2}$ and $x \sim N^+(60, 10)$ so that $x > 0$. After taking logs, the model can be written as

$$\log y_i = \beta_1 + \beta_2 \log x_i + v_i - u_i, \tag{5.1}$$

with v_i and u_i distributed as described in (2.3)–(2.4). Olson et al. (1980) note that only two parameters among the set $\{\sigma_u,\ \sigma_v,\ \sigma^2,\ \gamma,\ \sigma_\varepsilon^2\}$ (where $\sigma_\varepsilon^2 = \mathrm{VAR}(\varepsilon) = \frac{(\pi-2)\sigma_u^2}{\pi} + \sigma_v^2$) are independent. Following the reasoning of Olson et al. (1980), we set $\sigma_\varepsilon^2 = 1$ without loss of generality, and consider various values of the ratio $\lambda^2 = \frac{\sigma_u^2}{\sigma_v^2}$ of shape parameters. In addition, we set $\beta_1 = \log(10) \approx 2.3026$ and $\beta_2 = 0.8$.

On each Monte Carlo trial, we generated n observations $\{(x_i,\ y_i)\}_{i=1}^n$ and use these to estimate the parameters of the model by maximizing the log-likelihood (2.6). We then use the parameter estimates to estimate mean efficiency $\overline{\tau}$ defined in (3.4). We also estimate inefficiency for a set of five "hypothetical" firms, which do not necessarily correspond to any observations in our samples. For each of these hypothetical firms, the input level is fixed at 60.

Plausible values of the dependent variable in (5.1) vary across experiments, depending on λ^2. For each experiment, we compute the 0.1, 0.2, 0.5, 0.7, and 0.9 quantiles of $\varepsilon = v - u$; denote these by $\varepsilon_{(.1)},\ \ldots,\ \varepsilon_{(.9)}$. We then obtain log-output values for the five hypothetical firms by computing $\log y_{(.1)} = \beta_1 + \beta_2 \log(60) + \varepsilon_{(.1)}, \ldots, \log y_{(.9)} = \beta_1 + \beta_2 \log(60) + \varepsilon_{(.9)}$. The input/output values for these hypothetical firms are held constant across Monte Carlo trials, and across bootstrap replications within each Monte Carlo trial, reflecting the fact that in applied settings, points of interest in the input-output space are always taken as given.

Experiments were run with six different values of $\lambda^2 = \sigma_u^2/\sigma_v^2$, with $\lambda^2 \in \{0.1,\ 0.5,\ 1,\ 2,\ 10,\ 100\}$. These values of λ^2 correspond to $\gamma = 0.0909,\ 0.3333,\ 0.5,\ 0.6667,\ 0.9091$, and 0.9901, respectively. Each experiment consisted of 1,024 Monte Carlo trials; on each trial, we used 2,000 bootstrap replications.[7] Within a given experiment, conventional intervals based on (3.3) and (3.6) as well as intervals using the bootstrap method discussed in Section 4 were computed on each trial. Estimated coverages were computed by recording the proportion among the 1,024 Monte Carlo trials where the computed intervals include the underlying true values. In all cases, the percentile and BC bootstrap intervals were found to give similar results; to conserve space, we report only results based on the percentile intervals described in Section 4.

[7]The Monte Carlo experiments were performed on a massively parallel machine, where the number of processors for a particular job are conveniently and efficiently chosen as a power of 2. Choosing $2^{10} = 1,024$ Monte Carlo trials makes efficient use of the processors.

5.2 Coverage of Intervals for Mean Efficiency

Table 2 shows estimated coverages of mean inefficiency $\overline{\tau}$, defined in (3.4), by conventional and bootstrap confidence interval estimates for each value of λ^2 at nominal significance levels of 90, 95, and 99 percent, with sample sizes $n \in \{100,\ 1000\}$. With sample size $n = 100$, the conventional confidence intervals have poor coverage for all the values of λ^2 that were considered. The bootstrap intervals perform much better, though at .90 and .95 significance levels, coverages are slightly too large at smaller values of λ^2 and slightly too small at larger values of λ^2. At .99 significance, the bootstrap intervals have very good coverage. When the sample size is increased to $n = 1000$, coverages by the bootstrap intervals improve in almost every case. Coverages of the conventional intervals increase, but remain far too small for $\lambda^2 \leq 0.5$; for $\lambda \geq 2$, these intervals *always* cover the true value $\overline{\tau}$.

Careful inspection of the interval in (3.6) reveals why the conventional intervals fail so miserably. To illustrate, consider the case where $\lambda^2 = 0.1$; the results in Table 1 indicate roughly 46 percent of samples of size $n = 1000$—about 471 of 1,024 Monte Carlo trials—will have residuals with positive skewness, ensuring that $\widehat{\gamma} = 0$ and hence $\widehat{\sigma}_u^2 = 0$. Replacing σ_u^2 in (3.6) with $\widehat{\sigma}_u^2 = 0$, the interval estimate collapses to $[1, 1]$. But, simple algebra reveals that when $\lambda^2 = 0.1$ in our experiments, $\sigma_u^2 \approx 0.09649$ and hence $\overline{\tau} \approx 0.7935$. Hence, interval estimates based on (3.6) will cover $\overline{\tau}$ in *at most* about 54-percent, or 553 of 1,024 Monte Carlo trials, *regardless of the significance level that is chosen.*[8]

Now consider what happens when the bootstrap procedure described in Section 4 is used. In the roughly 471 trials where residuals have positive skewness, it remains true that $\widehat{\sigma}_u^2 = 0$. Then in step [2] of the procedure, $u_i^* = 0\ \forall\ i = 1,\ \ldots,\ n$; consequently, the residuals are composed only of the normally-distributed v_i^*. Although the normal distribution has no skewness, when finite samples are drawn, roughly half will have small but positive skewness, while the remaining samples will have small but negative skewness. Therefore, on each Monte Carlo trial where $\widehat{\sigma}_u^2 = 0$, the distribution of the B bootstrap estimates σ_{ub}^{*2} obtained in step [4] will exhibit a probability mass of about 0.5 at zero. This reflects rather closely the sampling distribution of $\widehat{\sigma}_u^2$, which—due to the results in Table 1—necessarily must contain

[8] As noted in Section 3.2, (3.6) gives a probability interval for the random variable e^{-u}. With $\lambda^2 = 0.1$ and hence $\sigma_u^2 \approx 0.09649$, $\Pr\left(e^{-u} = 1\right) = 0$. Thus, even if the estimator of (3.6) is interpreted as a probability interval for e^{-u}, instead of as a confidence interval for $\overline{\tau}$, it can include e^{-u} in at most about 54-percent of all cases when $\lambda^2 = 0.1$.

a probability mass of about 0.46 for the case described here.

In cases where sample sizes or λ^2 are large enough to ensure that few if any samples could ever be drawn that would yield residuals with positive skewness, interval estimates based on (3.6) also result in poor coverage properties; in Table 2, estimated coverages of the conventional intervals are much too large in such instances. This too is to be expected when one realizes that the intervals are based not on the sampling distribution of $\widehat{\overline{\tau}}$, but instead on percentiles of the distribution of $u \mid \varepsilon$.

5.3 Coverage of Intervals for Individual Efficiency

Turning to efficiency estimates based on (2.13) for the individual, hypothetical firms described above in Section 5.1, Table 3 gives results on coverages by conventional intervals based on (3.3) and bootstrap confidence interval estimates for sample size $n = 100$. Table 3 contains three sections, corresponding to nominal significance levels of .90, .95, and .99. Within each section, the first column indicates the quantiles of ε. The next six columns give estimated coverages by the conventional intervals for each of six values of λ^2, and the last six columns give similar results for the bootstrap intervals.

As with the results in Table 2 for mean efficiency, Table 3 indicates that coverages of the conventional intervals for individual efficiencies are poor in almost every instance. Although the bootstrap intervals give too much coverage with the smaller values of λ^2, their coverages are much closer to nominal significance than is the case with the conventional interval estimates. The results indicate that coverages depend somewhat dramatically on the signal-to-noise ratio as reflected by λ^2. Coverages also depend, though less severely, on where the output value lies for a given point of interest, as reflected by the quantiles of ε.

Table 4 gives similar results for $n = 1000$. Increasing the sample size causes coverages of the conventional intervals to increase in every case, but the coverage remains poor. In particular, for larger values of λ^2, the coverage is near or equal to 100-percent even at 90-percent nominal significance. The bootstrap intervals, by contrast, perform well with the larger sample size when $\lambda^2 \geq 1$. For smaller values of λ^2, coverage by the bootstrap intervals is too large for 90-percent and 95-percent nominal significance. But, small values of λ^2 represent low signal-to-noise ratios; in other words, when λ^2 is small, extracting information from a given body of data is necessarily more difficult than when λ^2 is larger. The bootstrap is

neither magic nor a panacea, but it still out-performs the conventional approach to inference regarding efficiencies of individual units.

The reasons for the failure of the conventional intervals for individual efficiencies are analogous to the reasons for their failure in the case of mean efficiency discussed in Section 5.2. In particular, for samples where residuals have positive skewness, $\widehat{\gamma} = 0$, and consequently $\widehat{\mu}_* = \widehat{\sigma}_*^2 = 0$, causing estimates of the interval in (3.3) to collapse to $[1, 1]$.

5.4 Coverage of Intervals for Model Parameters

The bootstrap algorithm given in Section 4 can also be used to estimate confidence intervals for the model parameters. Conventional inference about model parameters meanwhile relies on the asymptotic normality result from the theory of maximum likelihood, and employs the inverse negative Hessian of the log-likelihood function in (2.6) as an estimator of the variance-covariance matrix. Although the model is parameterized in terms of $\boldsymbol{\beta}$, σ^2, and γ, straightforward algebra yields estimators $\widehat{\sigma}_v^2$ and $\widehat{\sigma}_v^2$ as functions of $\widehat{\boldsymbol{\beta}}$, $\widehat{\sigma}^2$, and $\widehat{\gamma}$; then variances of the the estimators $\widehat{\sigma}_v^2$ and $\widehat{\sigma}_u^2$ can be estimated by $\nabla_v' \boldsymbol{\Sigma} \nabla_v$ and $\nabla_u' \boldsymbol{\Sigma} \nabla_u$ where $\boldsymbol{\Sigma}$ is the estimated variance covariance matrix for $(\boldsymbol{\beta},\ \sigma^2,\ \gamma)$ and ∇_v and ∇_u are vectors of derivatives of $\widehat{\sigma}_v^2$ and $\widehat{\sigma}_u^2$ with respect to $\widehat{\boldsymbol{\beta}}$, $\widehat{\sigma}^2$, and $\widehat{\gamma}$. Conventional variance estimates can be used to estimate confidence intervals in the usual way.

Table 5 gives estimated coverages of both conventional and bootstrap intervals for model parameters with sample size $n = 100$ and significance levels .90, .95, and .99. The first row of the table gives the number of trials—out of 1,024—where confidence intervals for model parameters could be estimated. Bootstrap intervals could be computed for every trail, but the negative Hessian of the log-likelihood function is singular whenever $\widehat{\gamma} = 0$ or 1. Due to the problems discussed in Section 3.1, $\widehat{\gamma} = 0$ with frequencies similar to those in Table 1. In addition, with a sample size of only 100 observations, large values of λ (10 and 100) yield small numbers of Monte Carlo trials with estimates $\widehat{\gamma} = 1$. Of course, the conventional confidence interval estimates cannot be computed when the negative Hessian is singular.

Estimated coverages in Table 5 for conventional intervals were computed by dividing the number of instances where computed intervals covered the corresponding true values not by the total number of Monte Carlo trials (1,024), but by the number of cases given in the first row of the table where conventional intervals could be computed. Consequently, the

coverage estimates shown for conventional intervals in Table 5 may be overly optimistic. Although coverage of the intercept and slope parameters (β_1 and β_2) by the conventional intervals is close to nominal significance levels in Table 5, coverage of σ^2, γ, σ_v^2, and σ_u^2 is poor for all values of λ^2. Coverage by the bootstrap interval estimates is similar to that of the conventional intervals for β_1 and β_2, and is more somewhat more accurate than with the conventional intervals for parameters σ^2, γ, σ_v^2, and σ_u^2.

Table 6 gives similar results for sample size $n = 1000$. Coverage by conventional intervals for σ^2, γ, σ_v^2, and σ_u^2 remains poor when the sample size is increased to 1,000, but coverage of these parameters by the bootstrap intervals improves for all but the smallest values of λ^2. For the smallest values of λ^2, coverage by bootstrap intervals for σ_v^2 is very accurate in Table 6, but too large for σ^2, γ, and σ_u^2 at significance levels .90 and .95. Coverages by the bootstrap intervals, however, are closer to nominal values than coverages by the conventional intervals. Once again, when the signal-to-noise ratio is low—reflected by small values of λ^2—it is apparently difficult to extract information about efficiency and noise from the model in 5.1.

5.5 Size and Power of Tests of $H_0 : \gamma = 0$

As noted above in Section 3.1, it is common in applications to test for the presence of inefficiency in a global sense. In the context of the Battese and Corra (1977) parameterization used here, this amounts to testing the null hypothesis $H_0 : \gamma = 0$. Also as discussed previously in Section 3.1, Wald tests are commonly used to test $H_0 : \gamma = 0$. A Wald statistic for testing $H_0 : \gamma = 0$ may be written as $\widehat{W} = \widehat{\gamma}^2/\widehat{\sigma}_{\widehat{\gamma}}^2$, where $\widehat{\sigma}_{\widehat{\gamma}}^2$ denotes an estimate of the variance of $\widehat{\gamma}$. Given the usual Cramér-Rao regularity conditions, $\widehat{W}$ would posess a limiting chi-square distribution with one degree of freedom, but as noted before, the Cramér-Rao regularity conditions do not hold under H_0. Consequently, the distribution of $\widehat{W}$ cannot be $\chi^2(1)$, but rather is a mixture of chi-square distributions, namely $\frac{1}{2}\chi^2(0) + \frac{1}{2}\chi^2(1)$ (as implied by the results in Lee, 1993). Moreover, the variance of $\widehat{\gamma}$ cannot be esimated by the corresponding diagonal of the inverse negative-Hessian, since the Hessian is singular under H_0. As noted previously in Section 3.1, Coelli (1995) used the DFP direction matrix to approximate the variance-covariance matrix in his Monte Carlo experiments to estimate rejection rates for Wald tests of $H_0 : \gamma = 0$. The DFP direction matrix is always positive

definite, even when the Hessian of the log-likelihood is singular as when $\widehat{\gamma} = 0$. In other words, in cases where $\widehat{\gamma} = 0$, the DFP direction matrix approximates a matrix that does not exist.

To capture all the possibilities that might be employed by reseachers using Wald statistics to test $H_0 : \gamma = 0$, we consider four versions of the test in our Monte Carlo experiments. Researchers could obtain the required variance estimates from the DFP direction matrix, or from the inverse negative-Hessian in cases where $\widehat{\gamma} > 0$. In addition, researchers might determine critical values either from the $\chi^2(1)$-distribution, or from the mixture $\frac{1}{2}\chi^2(0) + \frac{1}{2}\chi^2(1)$. We name the four possible tests accordingly:

- Wald Test #1: $\widehat{\sigma}^2_{\widehat{\gamma}}$ from DFP direction matrix, critical values from $\chi^2(1)$;
- Wald Test #2: $\widehat{\sigma}^2_{\widehat{\gamma}}$ from inverse negative-Hessian, critical values from $\chi^2(1)$;
- Wald Test #1: $\widehat{\sigma}^2_{\widehat{\gamma}}$ from DFP direction matrix, critical values from $\frac{1}{2}\chi^2(0) + \frac{1}{2}\chi^2(1)$;
- Wald Test #2: $\widehat{\sigma}^2_{\widehat{\gamma}}$ from inverse negative-Hessian, critical values from $\frac{1}{2}\chi^2(0) + \frac{1}{2}\chi^2(1)$.

Test #1 is equivalent to the Wald test considered by Coelli (1995). Test #2 and #4 can only be performed in cases where $\widehat{\gamma} > 0$ due to singularity of the Hessian matrix when $\widehat{\gamma} = 0$.

Of course, the likelihood ratio (LR) test can also be used to test $H_0 : \gamma = 0$. This test has the advantage that it does not require variance-covariance estimates, and consequently its size and power can be determined in Monte Carlo experiments provided one remembers that under the null, its distribution is not the usual chi-square but rather a mixture of chi-square distributions (see Lee, 1993, and the discussion in Section 3.1).

In addition, bootstrap methods can be used to test $H_0 : \gamma = 0$, although the algorithm given in Section 4 requires some modification to ensure that the bootstrap samples are generated *under the null*. This is accomplished by setting $u_i^* = 0 \; \forall \; i = 1, \; \ldots, \; n$ in step [2] of the algorithm, instead of drawing u_i^* from $N^+(0, \widehat{\sigma}_u^2)$. In addition, only the bootstrap estimates $\widehat{\gamma}_b^*$, $b = 1, \; \ldots, \; B$ need to be retained in step [4] of the algorithm. Then, after B bootstrap replications in step [4], the estimated p-value for the test is given by $\widehat{p} = \frac{\#\{\widehat{\gamma}_b^* > \widehat{\gamma}\}}{B}$. If $\widehat{p}$ is reasonably small, *e.g.* smaller than a chosen nominal size α, one may reject $H_0 : \gamma = 0$.

A series of Monte Carlo experiments were conducted with the same six values of λ^2 as in the previous experiments. In addition, an experiment where $\lambda^2 = 0$ was also conducted in

order to estimate the size of the various tests for $H_0 : \gamma = 0$. The experiments where $\lambda^2 > 0$ allow estimation of the power of the tests for various departures from the null. Results of these experiments are given in Table 7 for sample size $n = 100$, and in Table 8 for sample size $n = 1000$. Three significance levels (.90, .95, and .99) are considered; for each significance level, as well as for each of the seven values of λ^2, the Tables give the estimated rejection rates obtained with the Wald test using either the DFP direction matrix or the Hessian of the log-likelihood, the LR test, and the bootstrap test described above.

The first row in either Table 7 or 8 gives the number of Monte Carlo trials where the Hessian matrix was non-singular; only in these cases could the Wald tests (#2 and #3) based on the Hessian be computed.[9] Estimated coverages for these tests were obtained by dividing the number of trials where the null could be rejected by the number of trials where the test could be computed. For each of the other tests, division was by the number of Monte Carlo trials (1,024), since the other tests do not depend on non-singularity of the Hessian.

The results in the column of Table 7 where $\lambda^2 = 0$ reveal that, at least with sample size $n = 100$, the size properties of all versions of the Wald test are quite poor, especially for the versions (#1 and #3) based on variance estimates from the DFP direction matrix. Given the poor size properties, there is little reason to consider the tests' power. Both the LR and bootstrap tests appear to have good size properties as well as reasonable power to reject $H_0 : \gamma = 0$ when the null is in fact false.[10] For each of the three nominal sizes that were considered, the realized size of the bootstrap test is slightly smaller than nominal size in each case, while the realized size of the LR test is slightly larger than nominal size in each case. Both tests have realized sizes that are quite close to nominal sizes, however. In addition, the power of the LR and bootstrap tests are similar for nominal sizes .1 and .05, but the LR test appears to have greater power when the test size is .01 except in the case where $\lambda^2 = 100$.

When the sample size is increased to $n = 1000$, as in Table 8, the Wald test based on the DFP direction matrix again performs miserably in terms of size properties. The Wald test based on the Hessian has reasonable size when the nominal size is set at .05, but this

[9] The number of trials where the Hessian was non-singular differs slightly in several cases from the numbers reported in Tables 5 and 6, reflecting variation across Monte Carlo experiments.

[10] For sample size $n = 100$, Coelli (1995) found results for the Wald test using the DFP direction matrix and for the LR test similar to those reported here.

is apparently a fluke, since the test's realized size is far too large when nominal size is .1, and zero when nominal size is .01. By contrast, both the LR and bootstrap tests have very good size properties in Table 8. Moreover, both tests have very similar power for all values of $\lambda^2 > 0$ and each of the three nominal sizes.

Given these results, and the fact that the LR test requires less computation to implement than the bootstrap test, the LR test seems a better choice for applied work.

6 Conclusions

It has been known for some time that PSF models sometimes yield residuals with the "wrong" skewness. The implications of this phenomenon for inference, however, have not been carefully considered. This paper makes clear that, depending on the signal-to-noise ratio reflected by λ^2, the problem can arise with alarming frequency, even when the model is correctly specified. Consequently, "wrongly" skewed residuals in an observed sample do not provide meaningful evidence of specification error. Common practice among practitioners, however, has been to change the model specification whenever "wrongly" skewed residuals are encountered.

Classical inference assumes that the model specification is chosen independently of any estimates that are obtained; specification-searching introduces problems of bias in both parameter estimates as well as variance-covariance estimates. The bootstrap procedure given in Section 4 can be used to estimate confidence intervals with good coverage properties, and can extract useful information about efficiency levels even from samples where residuals have the "wrong" skewness. In such cases, one of course relies heavily on parametric assumptions, but this is probably always true when PSF models are estimated.

References

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

Andrews, Donald K. (2000), "Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space," *Econometrica,* 20, 399–405.

Banker, R.D., Charnes, A., Cooper, W.W., and Maindiratta, A. (1988), "A comparison of DEA and translog estimates of production functions using simulated data from a known technology," In *Applications of Modern Production Theory: Efficiency and Productivity,* ed. A. Dŏgramaçi and R. Färe, Boston: Kluwer Publishing Co., Inc.

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

Battese, G.E., and Corra, G.S. (1977), "Estimation of a production frontier model: with application to the pastoral zone off Eastern Australia," *Australian Journal of Agricultural Economics,* 21, 169–179.

Bauer, P.W. (1990), "Recent developments in the econometric estimation of frontiers," *Journal of Econometrics,* 46, 39–56.

Bera, A.K., and Sharma, S.C. (1999), "Estimating production uncertainty in stochastic frontier production function models," *Journal of Productivity Analysis,* 12, 187–210.

Bravo-Ureta, B.E., and Pinheiro, A.E. (1993), "Efficiency analysis of developing country agriculture: A review of the frontier function literature," *Agricultural and Resource Economics Review,* 22, 88–101.

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

Coelli, Tim (1996), "A guide to frontier version 4.1: A computer program for stochastic frontier production and cost function estimation," CEPA working paper #96/07, Centre for Efficiency and Productivity Analysis, University of New England, Arimidale, NSW 2351, Australia.

Davidon, W.C. (1959), "Variable metric method for minimization," Technical Report. Argonne National Laboratory Research and Development report #5990; reprinted as (Davidon 1991).

——— (1991), "Variable metric method for minimization," *SIAM Journal on Optimization,* 1, 1–17.

Efron, B. (1979), "Bootstrap methods: another look at the jackknife," *Annals of Statistics,* 7, 1–16.

——— (1982), *The Jackknife, the Bootstrap and Other Resampling Plans,* Philadelphia: Society for Industrial and Applied Mathematics. CBMS-NSF Regional Conference Series in Applied Mathematics, #38.

Efron, B., and Tibshirani, R.J. (1993), *An Introduction to the Bootstrap,* London: Chapman and Hall.

Fletcher, R., and Powell, M.J.D. (1963), "A rapidly convergent descent method for minimization," *Computer Journal,* 6, 163–168.

Gong, B.H., and Sickles, R.C. (1990), "Finite sample properties of stochastic frontier models using panel data," *Journal of Productivity Analysis,* 1, 229–261.

____ (1992), "Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data," *Journal of Econometrics,* 51, 259–284.

Greene, W.H. (1993), "The econometric approach to efficiency analysis," In *The Measurement of Productive Efficiency,* ed. H.O. Fried, C.A.K. Lovell, and S.S. Schmidt, New York: Oxford University Press pp. 68–119.

Greene, William H. (1995), *LIMDEP Version 7.0 User's Manual,* New York: Econometric Software, Inc.

Herrero, I., and Pascoe, S. (2002), "Estimation of technical efficiency: A review of some of the stochastic frontier and DEA software," *Computers in Higher Education Economics Review.* Virtual Edition (http://www.economics.ltsn.ac.uk/cheer.htm).

Hjalmarsson, L., Kumbhakar, S.C., and Heshmati, A. (1996), "DEA, DFA and SFA: A comparison," *Journal of Productivity Analysis,* 7, 303–327.

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

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

Kumbhakar, S.C., and Löthgren, M. (1998), "A Monte Carlo analysis of technical inefficiency predictors," Working Paper Series in Economics and Finance, No. 229, Stockholm School of Economics, P.O. Box 6501, S–113 83 Stockholm, Sweden.

Kumbhakar, S.C., and Lovell, C.A.K. (2000), *Stochastic Frontier Analysis,* Cambridge: Cambridge University Press.

Leamer, Edward E. (1978), *Specification Searches: Ad Hoc Inference with Nonexperimental Data,* New York: John Wiley & Sons, Inc.

Lee, L.F. (1993), "Asymptotic distribution of the maximum likelihood estimator for a stochastic frontier function model with a singular information matrix," *Econometric Theory,* 9, 413–430.

Meeusen, W., and van den Broeck, J. (1977), "Efficiency estimation from Cobb-Douglas production functions with composed error," *International Economic Review,* 18, 435–444.

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

Ritter, C., and Simar, L. (1997), "Pitfalls of normal–gamma stochastic frontier models," *Journal of Productivity Analysis,* 8, 167–182.

Sena, V. (1999), "Stochastic frontier estimation: A review of the software options," *Journal of Applied Econometrics,* 14, 579–586.

Simar, L., and Wilson, P.W. (2000), "Statistical inference in nonparametric frontier models: The state of the art," *Journal of Productivity Analysis,* 13, 49–78.

Waldman, D. (1982), "A stationary point for the stochastic frontier likelihood," *Journal of Econometrics,* 18, 275–279.

Table 1: Proportion of 1,000 Normal-Half Normal Samples with Positive Skewness

	$\lambda^2 = \frac{\sigma_u^2}{\sigma_v^2}$								
n	.01	.05	.1	.5	1	2	10	20	100
25	0.517	0.510	0.512	0.464	0.390	0.320	0.114	0.048	0.020
50	0.494	0.498	0.493	0.455	0.381	0.219	0.023	0.005	0.003
100	0.498	0.486	0.489	0.405	0.301	0.142	0.000	0.000	0.000
200	0.475	0.509	0.465	0.372	0.228	0.060	0.000	0.000	0.000
500	0.503	0.497	0.489	0.320	0.106	0.007	0.000	0.000	0.000
1,000	0.510	0.488	0.460	0.220	0.032	0.000	0.000	0.000	0.000
10,000	0.513	0.458	0.386	0.004	0.000	0.000	0.000	0.000	0.000
100,000	0.498	0.379	0.215	0.000	0.000	0.000	0.000	0.000	0.000
1,000,000	0.463	0.148	0.003	0.000	0.000	0.000	0.000	0.000	0.000

Table 2: Estimated Coverages of Confidence Intervals for Mean Inefficiency ($\overline{\tau}$)

	$\lambda^2 = \frac{\sigma_u^2}{\sigma_v^2}$					
Test	.1	.5	1	2	10	100
$n = 100$						
			$(1-\alpha) = .90$			
Conventional	0.480	0.541	0.643	0.811	0.996	1.000
Bootstrap	0.949	0.955	0.951	0.942	0.861	0.859
			$(1-\alpha) = .95$			
Conventional	0.483	0.544	0.649	0.817	0.996	1.000
Bootstrap	0.975	0.977	0.974	0.974	0.917	0.934
			$(1-\alpha) = .99$			
Conventional	0.487	0.549	0.654	0.822	0.998	1.000
Bootstrap	0.993	0.994	0.993	0.989	0.982	0.981
$n = 1000$						
			$(1-\alpha) = .90$			
Conventional	0.520	0.741	0.950	1.000	1.000	1.000
Bootstrap	0.951	0.955	0.905	0.898	0.896	0.903
			$(1-\alpha) = .95$			
Conventional	0.522	0.745	0.951	1.000	1.000	1.000
Bootstrap	0.982	0.981	0.969	0.945	0.942	0.942
			$(1-\alpha) = .99$			
Conventional	0.526	0.748	0.951	1.000	1.000	1.000
Bootstrap	0.998	0.996	0.994	0.988	0.987	0.985

Table 3: Estimated Coverages of Confidence Intervals for True Efficiencies $\tau_j = e^{-u_j}$ $(n = 100)$

	Conventional CIs $\lambda^2 = \frac{\sigma_u^2}{\sigma_v^2}$						Bootstrap CIs $\lambda^2 = \frac{\sigma_u^2}{\sigma_v^2}$					
quantile	.1	.5	1	2	10	100	.1	.5	1	2	10	100
						$(1-\alpha) = .90$						
0.1	0.233	0.399	0.513	0.675	0.878	0.692	0.949	0.951	0.946	0.934	0.850	0.886
0.3	0.383	0.502	0.608	0.767	0.916	0.700	0.949	0.952	0.952	0.938	0.852	0.876
0.5	0.461	0.536	0.639	0.800	0.940	0.704	0.949	0.953	0.954	0.936	0.850	0.869
0.7	0.480	0.542	0.651	0.816	0.964	0.705	0.949	0.963	0.963	0.880	0.832	0.863
0.9	0.483	0.548	0.654	0.823	0.979	0.711	0.980	0.963	0.818	0.711	0.832	0.841
						$(1-\alpha) = .95$						
0.1	0.305	0.444	0.546	0.724	0.911	0.703	0.974	0.975	0.972	0.964	0.904	0.955
0.3	0.434	0.521	0.621	0.784	0.938	0.705	0.974	0.975	0.975	0.966	0.908	0.944
0.5	0.477	0.539	0.647	0.810	0.948	0.707	0.974	0.976	0.977	0.969	0.905	0.940
0.7	0.484	0.547	0.653	0.820	0.972	0.710	0.974	0.979	0.979	0.947	0.910	0.930
0.9	0.485	0.551	0.656	0.826	0.980	0.715	0.995	0.983	0.894	0.790	0.884	0.901
						$(1-\alpha) = .99$						
0.1	0.414	0.503	0.613	0.771	0.938	0.713	0.990	0.992	0.992	0.983	0.972	0.990
0.3	0.473	0.543	0.646	0.805	0.955	0.716	0.990	0.993	0.993	0.986	0.972	0.988
0.5	0.483	0.547	0.653	0.818	0.967	0.718	0.990	0.994	0.994	0.990	0.974	0.987
0.7	0.488	0.551	0.656	0.826	0.979	0.720	0.991	0.994	0.996	0.988	0.982	0.987
0.9	0.488	0.553	0.659	0.831	0.982	0.719	1.000	0.995	0.957	0.916	0.947	0.955

Table 4: Estimated Coverages of Confidence Intervals for True Efficiencies $\tau_j = e^{-u_j}$ $(n = 1000)$

	Conventional CIs $\lambda^2 = \frac{\sigma_u^2}{\sigma_v^2}$						Bootstrap CIs $\lambda^2 = \frac{\sigma_u^2}{\sigma_v^2}$					
quantile	.1	.5	1	2	10	100	.1	.5	1	2	10	100
						$(1-\alpha) = .90$						
0.1	0.512	0.727	0.933	0.997	1.000	1.000	0.955	0.957	0.900	0.896	0.886	0.921
0.3	0.520	0.736	0.946	0.999	1.000	1.000	0.955	0.954	0.903	0.895	0.882	0.917
0.5	0.521	0.741	0.950	1.000	1.000	1.000	0.955	0.954	0.902	0.890	0.887	0.917
0.7	0.522	0.745	0.951	1.000	1.000	1.000	0.955	0.955	0.902	0.881	0.890	0.913
0.9	0.522	0.748	0.951	1.000	1.000	1.000	0.955	0.960	0.900	0.875	0.893	0.911
						$(1-\alpha) = .95$						
0.1	0.521	0.732	0.942	0.999	1.000	1.000	0.983	0.979	0.979	0.948	0.948	0.957
0.3	0.522	0.742	0.950	1.000	1.000	1.000	0.983	0.979	0.975	0.946	0.948	0.956
0.5	0.522	0.745	0.951	1.000	1.000	1.000	0.983	0.980	0.969	0.944	0.947	0.956
0.7	0.524	0.748	0.951	1.000	1.000	1.000	0.983	0.983	0.957	0.941	0.948	0.957
0.9	0.526	0.748	0.951	1.000	1.000	1.000	0.983	0.984	0.945	0.928	0.945	0.963
						$(1-\alpha) = .99$						
0.1	0.525	0.743	0.950	0.999	1.000	1.000	0.999	0.996	0.994	0.987	0.985	0.983
0.3	0.526	0.746	0.951	1.000	1.000	1.000	0.999	0.996	0.994	0.987	0.986	0.984
0.5	0.526	0.748	0.951	1.000	1.000	1.000	0.999	0.996	0.995	0.988	0.989	0.984
0.7	0.526	0.749	0.951	1.000	1.000	1.000	0.999	0.996	0.996	0.987	0.992	0.983
0.9	0.526	0.750	0.953	1.000	1.000	1.000	0.999	0.997	0.992	0.976	0.984	0.992

Table 5: Estimated Coverages of Confidence Intervals for Model Parameters ($n = 100$) (percentile bootstrap)

	Conventional CIs						Bootstrap CIs					
	$\lambda^2 = \frac{\sigma_u^2}{\sigma_v^2}$						$\lambda^2 = \frac{\sigma_u^2}{\sigma_v^2}$					
	.1	.5	1	2	10	100	.1	.5	1	2	10	100
#	503	576	688	861	1020	966	1024	1024	1024	1024	1024	1024
						$(1-\alpha) = .90$						
β_1	0.891	0.889	0.894	0.890	0.874	0.883	0.893	0.890	0.883	0.884	0.871	0.859
β_2	0.889	0.875	0.887	0.897	0.874	0.883	0.896	0.879	0.880	0.882	0.876	0.856
σ^2	0.579	0.630	0.655	0.627	0.151	0.019	0.943	0.965	0.956	0.909	0.854	0.855
γ	0.962	1.000	1.000	1.000	1.000	1.000	0.949	0.950	0.950	0.930	0.844	0.708
σ_v^2	0.984	0.988	0.997	0.987	0.995	1.000	0.774	0.866	0.906	0.909	0.829	0.704
σ_u^2	0.980	0.997	1.000	0.998	0.970	0.984	0.949	0.955	0.958	0.944	0.858	0.865
						$(1-\alpha) = .95$						
β_1	0.940	0.939	0.951	0.945	0.930	0.914	0.940	0.940	0.945	0.940	0.932	0.921
β_2	0.932	0.934	0.938	0.944	0.927	0.916	0.934	0.940	0.940	0.939	0.932	0.918
σ^2	0.620	0.667	0.685	0.665	0.183	0.027	0.969	0.983	0.984	0.961	0.912	0.929
γ	0.998	1.000	1.000	1.000	1.000	1.000	0.973	0.975	0.972	0.960	0.897	0.995
σ_v^2	0.988	0.997	0.997	0.997	0.995	1.000	0.841	0.909	0.940	0.939	0.892	0.986
σ_u^2	0.994	1.000	1.000	1.000	0.985	0.992	0.974	0.975	0.974	0.972	0.919	0.935
						$(1-\alpha) = .99$						
β_1	0.986	0.988	0.990	0.986	0.985	0.943	0.982	0.986	0.985	0.987	0.987	0.982
β_2	0.990	0.990	0.988	0.987	0.983	0.945	0.982	0.985	0.985	0.983	0.987	0.983
σ^2	0.668	0.721	0.721	0.713	0.249	0.038	0.994	0.996	0.997	0.995	0.976	0.977
γ	1.000	1.000	1.000	1.000	1.000	1.000	0.990	0.992	0.991	0.983	0.963	0.997
σ_v^2	0.996	0.998	1.000	1.000	0.997	1.000	0.933	0.970	0.984	0.976	0.952	0.997
σ_u^2	1.000	1.000	1.000	1.000	0.997	0.998	0.990	0.993	0.993	0.988	0.980	0.979

Table 6: Estimated Coverages of Confidence Intervals for Model Parameters ($n = 1000$) (percentile bootstrap)

	Conventional CIs						Bootstrap CIs					
	$\lambda^2 = \frac{\sigma_u^2}{\sigma_v^2}$						$\lambda^2 = \frac{\sigma_u^2}{\sigma_v^2}$					
	.1	.5	1	2	10	100	.1	.5	1	2	10	100
#	542	778	981	1024	1024	1024	1024	1024	1024	1024	1024	1024
						$(1-\alpha) = .90$						
β_1	0.904	0.910	0.895	0.888	0.887	0.870	0.905	0.901	0.884	0.878	0.890	0.889
β_2	0.887	0.888	0.891	0.884	0.889	0.875	0.891	0.880	0.889	0.878	0.887	0.888
σ^2	0.738	0.802	0.797	0.528	0.152	0.024	0.939	0.964	0.902	0.890	0.897	0.901
γ	0.801	0.986	1.000	1.000	1.000	1.000	0.955	0.957	0.896	0.902	0.885	0.913
σ_v^2	0.958	0.969	0.951	0.994	1.000	1.000	0.906	0.947	0.912	0.892	0.896	0.910
σ_u^2	0.873	0.978	0.992	0.980	0.998	0.998	0.955	0.954	0.899	0.901	0.895	0.903
						$(1-\alpha) = .95$						
β_1	0.945	0.945	0.945	0.942	0.935	0.935	0.945	0.946	0.941	0.936	0.938	0.942
β_2	0.948	0.946	0.942	0.940	0.935	0.939	0.943	0.946	0.943	0.939	0.938	0.944
σ^2	0.788	0.854	0.833	0.621	0.174	0.027	0.978	0.985	0.956	0.938	0.943	0.940
γ	0.898	1.000	1.000	1.000	1.000	1.000	0.983	0.979	0.977	0.948	0.953	0.974
σ_v^2	0.983	0.982	0.959	0.997	1.000	1.000	0.947	0.978	0.966	0.950	0.945	0.977
σ_u^2	0.948	0.992	0.999	0.992	0.999	1.000	0.983	0.981	0.967	0.948	0.944	0.940
						$(1-\alpha) = .99$						
β_1	0.985	0.983	0.986	0.989	0.985	0.980	0.988	0.985	0.987	0.987	0.985	0.984
β_2	0.987	0.983	0.986	0.987	0.986	0.982	0.986	0.983	0.985	0.986	0.986	0.984
σ^2	0.860	0.905	0.894	0.769	0.217	0.038	1.000	0.997	0.996	0.986	0.986	0.986
γ	0.991	1.000	1.000	1.000	1.000	1.000	0.999	0.996	0.995	0.988	0.983	0.995
σ_v^2	1.000	0.995	0.973	0.997	1.000	1.000	0.990	0.994	0.993	0.986	0.984	0.995
σ_u^2	0.996	1.000	1.000	1.000	1.000	1.000	0.999	0.996	0.994	0.987	0.987	0.985

Table 7: Estimated Rejection Rates for $H_0: \gamma = 0$ ($n = 100$)

	$\lambda^2 = \frac{\sigma_u^2}{\sigma_v^2}$						
Test	0.0	.1	.5	1	2	10	100
#	(490)	(503)	(577)	(689)	(861)	(1018)	(964)
				$\alpha = .1$			
Wald Test #1	0.2314	0.2373	0.3008	0.4023	0.6084	0.9717	0.9844
Wald Test #2	0.0918	0.0815	0.0953	0.1132	0.1882	0.7073	0.9098
Wald Test #3	0.2764	0.2832	0.3359	0.4463	0.6494	0.9795	0.9863
Wald Test #4	0.0939	0.0815	0.0953	0.1132	0.1882	0.7132	0.9098
LR	0.1113	0.1094	0.1436	0.2197	0.4043	0.9287	0.9971
Bootstrap	0.0977	0.0879	0.1270	0.1992	0.3701	0.9326	0.9971
				$\alpha = .05$			
Wald Test #1	0.2061	0.2139	0.2705	0.3730	0.5723	0.9678	0.9824
Wald Test #2	0.0204	0.0159	0.0295	0.0334	0.0499	0.3468	0.7635
Wald Test #3	0.2314	0.2373	0.3008	0.4023	0.6084	0.9717	0.9844
Wald Test #4	0.0000	0.0000	0.0000	0.0000	0.0012	0.0157	0.1712
LR	0.0635	0.0547	0.0820	0.1416	0.2744	0.8887	0.9971
Bootstrap	0.0488	0.0479	0.0713	0.1104	0.2236	0.8721	0.9971
				$\alpha = .01$			
Wald Test #1	0.1709	0.1670	0.2139	0.2979	0.5068	0.9590	0.9824
Wald Test #2	0.0000	0.0000	0.0000	0.0000	0.0012	0.0354	0.2396
Wald Test #3	0.1855	0.1846	0.2295	0.3271	0.5234	0.9619	0.9824
Wald Test #4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LR	0.0166	0.0166	0.0225	0.0459	0.1152	0.7344	0.9912
Bootstrap	0.0088	0.0098	0.0137	0.0215	0.0586	0.5840	0.9922

Table 8: Estimated Rejection Rates for $H_0: \gamma = 0$ ($n = 1000$)

	$\lambda^2 = \frac{\sigma_u^2}{\sigma_v^2}$						
Test	0.0	.1	.5	1	2	10	100
#	(517)	(542)	(778)	(981)	(1024)	(1024)	(1024)
				$\alpha = .1$			
Wald Test #1	0.1914	0.2100	0.4443	0.7959	0.9922	1.0000	0.9883
Wald Test #2	0.3288	0.3303	0.5450	0.7890	0.9902	1.0000	1.0000
Wald Test #3	0.2354	0.2656	0.5107	0.8281	0.9961	1.0000	0.9883
Wald Test #4	0.3288	0.3303	0.5488	0.7900	0.9912	1.0000	1.0000
LR	0.0977	0.1152	0.3037	0.6533	0.9824	1.0000	1.0000
Bootstrap	0.1006	0.1211	0.3066	0.6553	0.9844	1.0000	1.0000
				$\alpha = .05$			
Wald Test #1	0.1553	0.1670	0.4043	0.7471	0.9893	1.0000	0.9883
Wald Test #2	0.1954	0.2325	0.4190	0.6972	0.9854	1.0000	1.0000
Wald Test #3	0.1914	0.2100	0.4443	0.7959	0.9922	1.0000	0.9883
Wald Test #4	0.0426	0.0480	0.1375	0.4200	0.9268	1.0000	1.0000
LR	0.0498	0.0615	0.1885	0.5312	0.9639	1.0000	1.0000
Bootstrap	0.0527	0.0625	0.1895	0.5381	0.9688	1.0000	1.0000
				$\alpha = .01$			
Wald Test #1	0.1006	0.1143	0.3047	0.6572	0.9814	1.0000	0.9883
Wald Test #2	0.0677	0.0683	0.1774	0.4842	0.9482	1.0000	1.0000
Wald Test #3	0.1230	0.1348	0.3467	0.6875	0.9844	1.0000	0.9883
Wald Test #4	0.0000	0.0000	0.0000	0.0000	0.0000	0.9102	1.0000
LR	0.0059	0.0059	0.0566	0.2832	0.8555	1.0000	1.0000
Bootstrap	0.0068	0.0059	0.0527	0.2812	0.8594	1.0000	1.0000