Wrong Skewness and Finite Sample Correction in the Normal-Half Normal Stochastic Frontier Model

Jun Cai, Qu Feng, William C. Horrace, Guiying Laura Wu*

October 13, 2020

Abstract

In parametric stochastic frontier models, the composed error is specified as the sum of a two-sided noise component and a one-sided inefficiency component, which is usually assumed to be half-normal, implying that the error distribution is skewed in one direction. In practice, however, estimation residuals may display skewness in the wrong direction. Model respecification or pulling a new sample is often prescribed. Since wrong skewness may manifest as a finite sample problem, this paper proposes a finite sample adjustment to existing estimators to obtain the desired direction of residual skewness. This provides an alternative empirical approach to deal with the wrong skewness problem that does not require respecification of the model.

JEL Classifications: C13, C23, D24

Keywords: Stochastic frontier model, skewness, MLE, constrained estimators, BIC

---

*Cai and Horrace: Syracuse University. Feng and Wu: Nanyang Technological University. Horrace corresponding author: whorrace@maxwell.syr.edu.

1   Introduction

In parametric stochastic frontier models for cross-sectional data, the error term is composed as the sum of a two-sided noise component and a one-sided inefficiency component. The canonical model of Aigner Lovell and Schmidt (1977) assumes that the noise distribution is zero-mean normal and the inefficiency distribution is half-normal. Other common distributional assumptions on the inefficiency term include the exponential (Meeusen and van den Broeck, 1977; Aigner, Lovell and Schmidt, 1977), the truncated normal (Stevenson, 1980), and the gamma (Stevenson, 1980; Greene, 1980). For surveys, see Greene (2007) and Kumbhakar and Lovell (2000). The beauty of the canonical model is it allows for continuous inefficiency, does not require estimation of a support bound (in what amounts to a deconvolution problem), and only has two unknown distributional parameters, making maximum likelihood estimation simple relative to other specifications, except for (perhaps) the normal-exponential model. Consequently, the normal-half normal specification of the model has found widespread use.

In the normal-half normal production function specification, the skewness of the composed error is negative, and parameters can be estimated by maximum likelihood estimation (MLE) or corrected ordinary least squares (COLS).[1] Waldman (1982) shows that when the skewness of the ordinary least squares (OLS) residuals is positive, OLS is a local maximum of the likelihood function, and estimated inefficiency is zero in the sample.[2] Horrace and Wright (2020) generalize the Waldman result to the case where the inefficiency distribution is from a scalable parametric class, like the exponential distribution. This "wrong skewness" phenomenon is widely documented in the literature and is often regarded as an estimation failure.[3] When it occurs, researchers are advised to either obtain a new sample (which is rarely feasible) or respecify the model. While there are many ways one can respecify a model, a reasonable approach would be to chose an inefficiency distribution that accommodates residuals of either positive or negative skew. For example, Li (1996), Carree (2002), Almanidis and Sickles (2011), Almanidis, Qian and Sickles (2014), and Hafner, Manner and Simar (2019) develop models with new distributional assumptions on the inefficiency component that allow for skew in either direction.

---

[1]The skewness of the composed error is positive in the stochastic frontier cost function model. We use the terminology COLS following Olson, Schmidt and Waldman (1982). COLS is also called MOLS. See Greene (2007).

[2]Greene (2007, p.131) claims "In this instance, the OLS results are the MLEs, and consequently, one must estimate the one-sided terms as 0."

[3]For example, estimating the variance parameters in COLS is invalid in this case.

An alternative "solution" to the problem that does not require respecification is that of Simar and Wilson (2010), who argue that "wrong skewness" is not an estimation or misspecification failure but a finite sample problem that most likely occurs when the ratio of the inefficiency variance to the noise variance (the inefficiency variance ratio) is small. That is, wrong skewness may not indicate that the normal-half normal model is wrong or that inefficiency does not exist in the population. They propose a bootstrap method (called "bagging") to construct confidence intervals for model parameters and expected inefficiency which have higher coverage than traditional intervals, regardless of residual skewness direction. Then, the sample under study can still be used to infer the model parameters. While bagging can be applied to any parametric form of the model, their technique is specifically intended to salvage the canonical model when the residual skew has the wrong sign. Such is the spirit of this research.

We take Simar and Wilson's (2010) view that wrong skewness may be a consequence of a small inefficiency variance ratio in finite samples, even when the canonical model is properly specified.[4] However, instead of the bagging approach of Simar and Wilson (2010), this paper provides a finite sample adjustment to existing estimators in the presence of wrong skewness. That is, we impose a negative residual skewness constraint in the MLE (or COLS) algorithm. A natural candidate for this constraint is the upper bound of the population skew, which is a monotonic function of the positive lower bound of the inefficiency variance ratio in the half-normal model. However, the constraint is non-linear in the parameters of interest, complicating computation of the optimum. Therefore, a linearization approximation of the constraint is proposed. Additionally, a model selection approach is proposed to determine the lower bound of the inefficiency variance ratio used in the constraint. Monte Carlo experiments suggest that our correction becomes more reliable as the true inefficiency variance ratio increases. A shortcoming of the approach is that in finite samples the linear approximation may not be accurate enough to guarantee a negative residual skew, but it will always give us non-zero estimates for firm-level inefficiencies. The possible failure of correct residual skewness using the linearized constraint illustrates a trade-off between computational complexity and accuracy. Using the original non-linear constraint avoids this issue, but the computational simplicity of our approach, as shown in our Monte Carlo experiments and empirical example, would be lost.

The proposed finite sample adjustment provides a non-zero point estimate for technical (or cost)

---

[4]It may also be a consequence of a misspecified model, but that is not our focus here.

inefficiency that may be used in applied research and that serves as a compliment to the bagging intervals of Simar and Wilson (2010). That is, if the wrong skew exists in the canonical model, then one may use Simar and Wilson's bagging technique to construct confidence intervals on efficiency while the proposed methodology provides non-zero point estimates without having to respecify the model. Since wrong skewness can occur fairly regularly (even when efficiency may exist in the population under study), the finite sample adjustment is particularly attractive in cases where the half-normal inefficiency assumption is maintained. It is noteworthy that the proposed adjustment is only needed *in finite samples*, for as the sample size increases wrong skewness is less likely to be an issue when the ratio of inefficiency variance to the noise variance is sizable.

This rest of this paper is organized as follows. The next section discusses the wrong skewness issue in the literature. In Section 3, we propose a finite sample correction approach. To simplify computation of the proposed constrained estimation, a linearized version of the constraint is used, so that constrained MLE (or COLS) can be easily implemented in most software packages (like STATA). The constrained estimators are discussed in Section 4. In Section 5, Monte Carlo experiments are conducted to study the properties of constrained COLS. An empirical example is used to illustrate the proposed approach in Section 6, and all the point estimates of inefficiency lie within the confidence intervals of Simar and Wilson (2010). The last section concludes.

2 Wrong Skewness Issue

A stochastic production frontier (SPF) model for a cross-sectional sample of size N is:

$$y_i = x_i'\beta + \varepsilon_i, \quad i = 1, \cdots, N, \tag{1}$$

with composed error $\varepsilon_i = v_i - u_i$. The disturbance v_i is assumed $iidN(0, \sigma_v^2)$. Inefficiency of firm i is characterized by $u_i \geq 0$. In the SPF literature, u_i is usually assumed half-normal $|iidN(0, \sigma_u^2)|$ (Aigner, Lovell and Schmidt, 1977; Wang and Schmidt, 2009), and independent of v_i, with variance $Var(u_i) = \frac{\pi-2}{\pi}\sigma_u^2$. The first component of the $p \times 1$ vector x_i is 1, so the intercept term is contained into the $p \times 1$ slope parameter vector β. As in Aigner, Lovell and Schmidt (1977) and Simar and Wilson (2010), let $\sigma^2 = \sigma_u^2 + \sigma_v^2$ and $\lambda = \sigma_u/\sigma_v$. The parameters to be estimated are $\theta = (\beta, \lambda, \sigma^2)$.

There are two primary estimators suggested in the literature: the maximum likelihood estimator and corrected least squares (Aigner, Lovell and Schmidt, 1977; Olson, Schmidt and Waldman, 1980). Under the normal-half normal specification, the MLE of $(\beta, \lambda, \sigma^2)$ is the set of parameters values

3

maximizing the likelihood function:

$$\ln L(\beta, \lambda, \sigma^2 | (y_i, x_i), i = 1, ..., N) \tag{2}$$

$$= \frac{N}{2} \ln(\frac{2}{\pi}) - \frac{N}{2} \ln \sigma^2 + \sum_{i=1}^{N} \ln \left[1 - \Phi \left(\frac{\lambda}{\sqrt{\sigma^2}} (y_i - x_i'\beta) \right) \right]$$

$$- \frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - x_i'\beta)^2,$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. The COLS estimate of β is simply the least squares slope estimate in the regression of y_i on x_i. However, the mean of $\varepsilon_i = v_i - u_i$ is negative due to the term $-u_i$, so the COLS estimate needs to be adjusted by adding the bias, $\sqrt{2\sigma_u^2/\pi}$, back into the intercept estimator. The bias can be consistently estimated using the variance estimates:

$$\hat{\sigma}_u^2 = \left[\sqrt{\frac{\pi}{2}} \left(\frac{\pi}{\pi - 4} \right) \hat{\mu}_3' \right]^{2/3}, \quad \hat{\sigma}_v^2 = \hat{\mu}_2' - \frac{\pi - 2}{\pi} \hat{\sigma}_u^2, \tag{3}$$

where $\hat{\mu}_2'$ and $\hat{\mu}_3'$ are the estimates of second and third sample moments of the least squares residuals.

Both MLE and COLS are consistent. The Monte Carlo experiments in Olson, Schmidt and Waldman (1980) show that there is little difference between MLE and COLS for the slope coefficients in finite samples. For the intercept and variance parameters, however, MLE and COLS differ. In addition to MLE and COLS, Olson, Schmidt and Waldman (1980) also consider a third consistent estimator, the two-step Newton-Raphson estimator, which has different finite sample properties than MLE and COLS.

Waldman (1982) discovers an important property of MLE: for the likelihood function (2) above, the point $(b, 0, s^2)$ is a stationary point, where b and s^2 are the OLS estimates of β and σ^2. Intuitively, when $\lambda = 0$, the term u_i disappears, so the likelihood function of the SPF model (2) boils down to one of a linear model with $u_i = 0$. A salient result in Waldman (1982) is that when the skewness of the OLS residuals is positive, i.e., $\hat{\mu}_3' > 0$, then $(b, 0, s^2)$ is a local maximum in the parameter space of the likelihood function.[5] This is the so-called "wrong skewness issue" in the literature, because $\mu_3' < 0$ in the normal-half normal model. Olson, Schmidt and Waldman (1980) refer to this phenomenon as "Type I failure" since the COLS estimator defined in (3) does not exist when $\hat{\mu}_3' > 0$.

---

[5]Waldman (1982, p. 278) also suggests that $(b, 0, s^2)$ may be a global maximum. There are two roots in this normal-half normal model: OLS $(b, 0, s^2)$ and one at the MLE with positive λ. When the residual skewness is positive, the first is superior to the second (Greene, 2007, note 28).

The Monte Carlo studies in Simar and Wilson (2010) show that the wrong skewness issue is not rare, even when the inefficiency variance ratio is considerably large. For example, the frequency of wrong skewness could be 30% for a sample of size of 100 when $\lambda = \sigma_u/\sigma_v = 1$. Wrong skewness casts doubt on the specification of the SPF model (Greene, 2007). Moreover, it invalidates the calculation of standard errors of parameter estimates (Simar and Wilson, 2010).

Greene (2007) considers OLS residual skewness a useful diagnostic tool for the normal-half normal model. Wrong skewness suggests there is little evidence of inefficiency in the sample, implying that firms in the sample are "super efficient". Thus, λ and σ_u^2 are assumed to be zero, and the stochastic frontier model reduces to a production function without the inefficiency term.[6] Another interpretation of the wrong skewness issue is that the normal-half normal model is not the correct specification. Other specifications may well reveal the presence of inefficiency and reconcile the distribution of one-sided inefficiency with the data. The binomial distribution considered by Carree (2002) and doubly truncated normal distribution proposed by Almanidis and Sickles (2011) and Almanidis, Qian and Sickles (2014) could have either negative or positive skewness. They argue that models with ambiguous skewness may be more appropriate in applied research.

Simar and Wilson (2010) argue that wrong skewness is a finite sample problem, even when the model is correctly specified.[7] They show that a bootstrap aggregating method provides useful information about inefficiency and the model parameters, regardless of whether residuals are skewed in the desired direction. We also consider wrong skewness to be a consequence of estimation in finite samples when the inefficiency variance ratio $Var(u_i)/Var(\varepsilon_i)$ is small.[8] Since the OLS residuals of a production function regression with $u_i = 0$ display skewness in either direction with probability 50%, a sample drawn from an SPF model with small inefficiency variance ratio could generate positively skewed residuals with high probability.[9]

---

[6]Kumbhakar, Parmeter and Tsionas (2013) propose a stochastic frontier model to accommodate the presence of both efficient and inefficient firms in the sample.

[7]Waldman (1982, p.278) notes that for $\sigma_u > 0$ "as the sample size increases the probability that $\sum e_t^3 > 0$ and hence that $(b, 0, s^2)$ locates a local maximum goes to zero."

[8]Badunenko, Henderson and Kumbhakar (2012) find that the estimation of efficiency scores depends on the estimated ratio of the variation in efficiency to the variation in noise. As discussed by Kim, Kim and Schmidt (2007) and Feng and Horrace (2012) in fixed effects stochastic frontier models, small signal-to-noise ratio leads to inaccurate inference.

[9]As pointed out by Simar and Wilson (2010, p.72), this problem could happen in other one-sided specifications. In a previous version of this paper, our Monte Carlo experiments suggest that wrong skewness could also occur with high probability in exponential and binomial SPF models, when the signal-to-noise ratio is small.

3 Finite Sample Correction

As illustrated by Simar and Wilson (2010), wrong skewness may occur when the inefficiency variance ratio is sizable, so simply setting $\sigma_u^2 = 0$ when the skewness is positive could be a mistake. Instead of improved interval estimates proposed by Simar and Wilson (2010), this paper proposes a finite sample adjustment to existing estimators in the presence of wrong skewness.

For MLE, a constraint with non-positive residuals skewness is imposed:

$$\max \ln L(\beta, \lambda, \sigma^2 | (y_i, x_i), i = 1, ..., N)$$

$$s.t. \ \frac{1}{N} \sum_{i=1}^{N} \left[\frac{y_i - \bar{y} - x_i'\beta + \bar{x}'\beta}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - x_i'\beta - \bar{y} + \bar{x}'\beta)^2}} \right]^3 \leq 0, \tag{4}$$

where $\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$ and $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$. Unfortunately, when implementing maximum likelihood estimation with the inequality constraint defined by (4), there is a practical issue. As pointed out by Waldman (1982), in the case of positive skewness of residuals, OLS $(b, 0, s^2)$ is a local maximum and the unconstrained MLE is equal to $(b, 0, s^2)$. Since, OLS is a local maximum in the parameter space of unconstrained MLE, the constraint (4) is always binding at the maximum, leading to zero skewness of the constrained MLE residuals.[10]

If we regard the sign of residual skewness as an important indicator of model specification, the constrained MLE above seems unsatisfactory. We, therefore, propose a (negative) upper bound of skewness instead of zero in (4). This is relevant for empirical modeling. As in the empirical example below, when there is evidence of technical inefficiency in the data (Greene 2007, p.202), its variance cannot be too small, relative to that of the composed error ε_i. Denote the inefficiency variance ratio by

$$k = Var(u_i)/Var(\varepsilon_i).$$

That is, a lower bound on the inefficiency variance ratio is implicitly imposed, $k \geq k_0$. From this perspective, to impose a positive value of k_0 is to obtain a non-zero estimated inefficiency.

To develop the relationship between the upper bound of skewness and the lower bound of the inefficiency variance ratio, consider the second and third moment of ε_i. Under the normal-half

---

[10]This stems from the fact that Waldman (1982) shows that OLS is local maximum in the parameter space of MLE when the OLS residuals are positively skewed. In fact, the non-positivity contraint will bind globally (when the OLS residuals are positively skewed), if OLS is a global maximum, as the Monte Carlo studies of Olsen, Waldman and Schmidt (1980) suggest.

normal specification, Olson, Schmidt and Waldman (1980) show that

$$Var(\varepsilon_i) = \sigma_v^2 + \frac{\pi - 2}{\pi}\sigma_u^2 \tag{5}$$

and

$$E[\varepsilon_i - E(\varepsilon_i)]^3 = \sigma_u^3\sqrt{2/\pi}[(\pi - 4)/\pi]. \tag{6}$$

Using $\gamma_1(\varepsilon_i)$ to denote the skewness of ε_i, we have

$$\gamma_1(\varepsilon_i) = E\left[\frac{\varepsilon_i - E(\varepsilon_i)}{\sqrt{Var(\varepsilon_i)}}\right]^3 = -\left[\frac{Var(u_i)}{Var(\varepsilon_i)}\right]^{3/2}\gamma_1(u_i) = -k^{3/2}\gamma_1(u_i),$$

where $\gamma_1(u_i) = \frac{4-\pi}{\pi-2}\sqrt{\frac{2}{\pi-2}} \simeq 0.9953$, a constant for a half-normal distribution of u_i. Denote $g(k) = \gamma_1(\varepsilon_i) = -k^{3/2}\gamma_1(u_i)$. Since $\gamma_1(u_i) > 0$, $g(k) < 0$ (e.g., $g(0.1) \approx -0.0315$, $g(0.2) \approx -0.0890$ and $g(0.3) \approx -0.1635$) and $g'(k) = -\frac{3}{2}k^{1/2}\gamma_1(u_i) < 0$. An important property of $g(k)$ is that it is a monotonically decreasing function of k. This implies that any upper bound, say g_0, of the population skewness, $g(k) = \gamma_1(\varepsilon_i) \le g_0$, is equivalent to a lower bound, denoted by k_0, of the inefficiency variance ratio, $k \ge k_0$, i.e., $g_0 = g(k_0) < 0$.

We impose this upper bound on the sample skewness, by replacing 0 in the constraint (4) with the negative upper bound of the population skewness, $g(k_0)$. Consequently, a modified constraint

$$\frac{1}{N}\sum_{i=1}^{N}\left[\frac{y_i - \bar{y} - x_i'\beta + \bar{x}'\beta}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i'\beta - \bar{y} + \bar{x}'\beta)^2}}\right]^3 \le g(k_0)$$

is used in the constrained MLE in the event of wrong skewness of the OLS residuals.

Based on Waldman's (1982) argument, the constraint above will also be binding at a maximum in the neighborhood of OLS. The constraint becomes

$$\frac{1}{N}\sum_{i=1}^{N}\left[\frac{y_i - \bar{y} - x_i'\beta + \bar{x}'\beta}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i'\beta - \bar{y} + \bar{x}'\beta)^2}}\right]^3 = g(k_0) \tag{7}$$

This finite sample adjustment gives a constrained estimator of parameter vector $(\beta, \lambda, \sigma^2)$.

The constrained COLS slope coefficients can be similarly defined. We use constraint (7), but replace the likelihood (2) with the sum of squared residuals as the objective function of a minimization problem. Since COLS reduces to OLS in the presence of wrong skewness and OLS is a local maximum of likelihood, as a finite sample adjustment to OLS, the constrained COLS slope coefficients are expected be close to their constrained MLE counterparts.

3.1 Linearizing the constraint

The non-linearity of β in the constraint (7) creates computational difficulties in calculating the constrained MLE. To simplify computation, a linearized version of the constraint (7) is considered. Given that OLS is a local maximum of likelihood in the presence of wrong skewness, empiricists normally start by estimating OLS with $u_i = 0$. This is the first step in LIMDEP (Greene, 1995) and FRONTIER (Coelli, 1996). If the skewness of the OLS residuals is positive, then OLS is the optimum and the point of departure for our linearization concept.

Since the primary concern is skewness correction, we impose the additional restriction that the MLE residual variance $\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i'\beta - \bar{y} + \bar{x}'\beta)^2$ is equal to that of OLS residuals, $\hat{\mu}_2'$, which is a consistent estimator of the error variance. Thus, the linearized constraint becomes:

$$\frac{1}{N}\sum_{i=1}^{N}[y_i - \bar{y} - (x_i + \bar{x})'\beta]^3 = g(k_0) \cdot (\hat{\mu}_2')^{3/2}.$$

Denote $f(\beta) = \frac{1}{N}\sum_{i=1}^{N}[y_i - \bar{y} - (x_i + \bar{x})'\beta]^3$. The first-order Taylor expansion of $f(\beta)$ at the OLS estimate $\hat{\beta}_{OLS}$ is:

$$f(\beta) \approx f(\hat{\beta}_{OLS}) + \left[\frac{\partial f(\beta)}{\partial \beta}_{|\hat{\beta}_{OLS}}\right]' \cdot (\beta - \hat{\beta}_{OLS}),$$

where $\frac{\partial f(\beta)}{\partial \beta}_{|\hat{\beta}_{OLS}}$ is the derivative of $f(\beta)$ with respect to β evaluated at $\hat{\beta}_{OLS}$. $f(\hat{\beta}_{OLS})$ is the 3rd central moment of OLS residuals, i.e., $\hat{\mu}_3'$. Now,

$$\frac{\partial f(\beta)}{\partial \beta} = -\frac{3}{N}\sum_{i=1}^{N}[y_i - \bar{y} - (x_i + \bar{x})'\beta]^2(x_i - \bar{x}),$$

and

$$\frac{\partial f(\beta)}{\partial \beta}_{|\hat{\beta}_{OLS}} = -\frac{3}{N}\sum_{i=1}^{N}e_i^2(x_i - \bar{x}),$$

where e_i denotes the OLS residual $y_i - x_i'\hat{\beta}_{OLS}$. Its sample mean is equal to zero since a constant term is included in the regression. Hence, an approximation of the constraint (7) is

$$\hat{\mu}_3' - \frac{3}{N}\sum_{i=1}^{N}e_i^2(x_i - \bar{x})'(\beta - \hat{\beta}_{OLS}) = g(k_0) \cdot (\hat{\mu}_2')^{3/2}, \tag{8}$$

or

$$[\frac{1}{N}\sum_{i=1}^{N}e_i^2(x_i - \bar{x})]'(\beta - \hat{\beta}_{OLS}) = \frac{\hat{\mu}_3'}{3} - \frac{g(k_0)}{3}(\hat{\mu}_2')^{3/2}. \tag{9}$$

8

Letting the $N \times 1$ vector $\tilde{e}$ be the squared OLS residual vector $(e_1^2, ..., e_N^2)'$, the constraint above can be written in matrix form as

$$\frac{1}{N}\tilde{e}'M_0 X(\beta - \hat{\beta}_{OLS}) = \frac{\hat{\mu}_3'}{3} - \frac{g(k_0)}{3}(\hat{\mu}_2')^{3/2},$$

where $M_0 = I - \frac{1}{N}\iota\iota'$ and $\iota = (1, ..., 1)'$. Thus, the linear constraint above can be written as

$$R\beta = q(k_0) \tag{10}$$

with $R = \frac{1}{N}\tilde{e}'M_0 X$ and $q(k_0) = R\hat{\beta}_{OLS} + \frac{\hat{\mu}_3'}{3} + \frac{\Pi}{3}k_0^{3/2}(\hat{\mu}_2')^{3/2}$, depending on the value of k_0.[11]

Therefore, the proposed finite sample correction for MLE of $(\beta, \lambda, \sigma^2)$, i.e., the constrained MLE, is defined as the solution to maximizing the likelihood (2) subject to the linear constraint (10). The corresponding estimators of σ_u^2 and σ_v^2 can be obtained by using the relationship $\sigma^2 = \sigma_u^2 + \sigma_v^2$ and $\lambda = \sigma_u/\sigma_v$.

Similarly, the constrained COLS of β is defined to minimize the sum of squared residuals subject to (10). As in the unconstrained estimation, the constrained estimators of σ_u^2 and σ_v^2 can be obtained by formula (3).

If $k_0 = 0$, then $g(k_0) = 0$ and the constraint above becomes $R(\beta - \hat{\beta}_{OLS}) = \hat{\mu}_3'/3$. This implies that the constrained and unconstrained estimators would be similar, since $\hat{\mu}_3'$ is usually very small in the presence of wrong skewness. In the extreme case of $\hat{\mu}_3' = 0$, the constrained estimator reduces to OLS, which is a local maximum of the likelihood.

Using the linearized constraint (10), the estimates, standard errors and confidence intervals of the constrained MLE and constrained COLS can be easily obtained using Stata or other existing software.[12]

However, since (10) does not guarantee a negative residual skewness in finite samples, there is a possibility that wrong skewness could still occur after our correction. The Monte Carlo experiments below show that this may only be a concern when the underlying inefficiency variance ratio is very small. However, as stressed above, by setting $k_0 > 0$, a non-zero estimated inefficiency may be obtained even in the presence of wrong residual skewness.

---

[11]It is worth noting that (10) is not a direct linearization of (7). Alternatively, a full linearization of (7) can be similarly obtained by replacing $R = \frac{1}{N}\tilde{e}'M_0 X$ with $R = \frac{1}{N}(\tilde{e}'M_0 - \sqrt{\hat{\mu}_2'}\hat{\mu}_3'e')X$. The additional term $-\frac{1}{N}\sqrt{\hat{\mu}_2'}\hat{\mu}_3'e'X$ is from the effect of the denominator of the constraint in (7). Monte Carlo simulations suggest that the estimation results are robust to this choice. Details are available upon request.

[12]In the empirical example below, the command *Frontier* in Stata, which allows for a linear constraint, is employed.

3.2 Choosing the value of k_0

The idea of the proposed constrained estimators is to adjust the slope coefficients to obtain a correct sign of residual skewness using the constraint (10), which is a function of k_0. It is expected that when the chosen value of k_0 is small, a slight adjustment results in the constrained MLE (or constrained COLS), and its value will be close to the unconstrained MLE.

Choosing a specific value of k_0 is an empirical issue. On the one hand, when there is a *priori* evidence of inefficiency, the inefficiency variance ratio cannot be too small. On the other hand, as illustrated by the Monte Carlo study in Simar and Wilson (2010), wrong skewness is less likely to occur as the inefficiency variance ratio increases.[13] In the spirit of this trade-off we develop a model selection criteria to choose k_0. The idea is to incorporate a penalty function, so that as k_0 increases the penalty decreases. Hence, the fit of the model and effect of the constraint on the optimum can be balanced.

For constrained MLE we propose a Bayesian information criterion (BIC) via the likelihood to choose the value of k_0:

$$BIC(k_0) = -2l_r(k_0) - k_0 \ln N,$$

where $l_r(k_0)$ is the log-likelihood evaluated at the constrained MLE of $(\beta, \lambda, \sigma^2)$, depending on k_0. Since OLS $(b, 0, s^2)$ is a local maximum of the log-likelihood function in the presence of positive skewness with a restriction on k_0, the value of $l_r(k_0)$ decreases with k_0 in the neighborhood of $(b, 0, s^2)$.[14] Different from the usual BIC, here we use a negative sign in front of the penalty term $k_0 \ln N$ so that $-2l_r(k_0)$ and $-k_0 \ln N$ move in opposite directions with k_0. An optimal value of k_0 is chosen to minimize $BIC(k_0)$:

$$\tilde{k}_0 = \arg \min_{k_0 \in [0,1)} BIC(k_0).$$

---

[13]Table 1 in Simar and Wilson (2010) provides some guidance. On the one hand, when $\lambda^2 \leq 0.1$ (i.e., $k = 1/(\frac{\pi}{\pi-2}\frac{1}{\lambda^2} + 1) < 0.035$) for samples with size less than 200, the proportion of wrong skewness is close to 50%, implying that the inefficiency term is hard to distinguish from noise. On the other hand, when $\lambda^2 \geq 1$ ($k \geq 0.267$), the wrong skewness probability decreases dramatically. For example, only 6% of samples display wrong residual skewness for $\lambda^2 = 2$ ($k = 0.421$) and $N = 200$. We have a similar finding both for Simar and Wilson's design and the design in Section 5 of this paper. Results are available upon request.

[14]The constraint $k \geq k_0$ is always binding in the neighborhood of OLS. And a restriction on k is equivalent on λ, which is a monotonic increasing function of k in the half-normal model,

$$\lambda = \sqrt{\frac{\sigma_u^2}{\sigma_v^2}} = \sqrt{\frac{Var(u_i)}{\frac{\pi-2}{\pi}\sigma_v^2}} = \sqrt{\frac{k}{\frac{\pi-2}{\pi}(1-k)}} = \sqrt{\frac{\pi}{\pi-2}\frac{1}{(1/k-1)}}.$$

10

Similarly, for the constrained COLS, a criterion based on sum of squared residuals is proposed to select the value of k_0:

$$C(k_0) = \frac{1}{N} SSR_r(k_0) - k_0 \hat{\sigma}_\varepsilon^2 \frac{\ln N}{N},$$

where $SSR_r(k_0)$ is the sum of squared residuals of OLS with the constraint (10). $C(k_0)$ is a Mallows' C_p-type criterion, similar to the expression proposed by Bai and Ng (2002) to choose the number of factors in the approximate factor models, except that the penalty term takes a negative sign. By applying the properties of the usual restricted least squares, it can be shown that $SSR_r(k_0)$ increases with k_0. (See the appendix.) Hence, the effect of increasing k_0 on the model fit can be balanced by the penalty term, thus an appropriate value of k_0 is chosen to minimize $C(k_0)$:

$$\hat{k}_0 = \arg \min_{k_0 \in [0,1)} C(k_0).$$

The estimated error variance $\hat{\sigma}_\varepsilon^2$ provides an appropriate scaling to the penalty term. Here, we use $\hat{\sigma}_\varepsilon^2 = \frac{1}{N} SSR$, where SSR is the sum of squared residuals of OLS without constraint.

In practice, to find the value of $\tilde{k}_0$ (or $\hat{k}_0$) a grid search can be applied to $BIC(k_0)$ (or $C(k_0)$) starting from a small positive value, e.g., 0.05.

Since the measures of the model fit in the constrained MLE and COLS, i.e., the objective functions in the penalized least squares and penalized maximum likelihood, are different, $\tilde{k}_0$ is not necessarily equal to $\hat{k}_0$. However, in the neighborhood of OLS $(b, 0, s^2)$ with a small value of λ, when the term $\sum_{i=1}^{N} \ln \left[1 - \Phi \left(\frac{\lambda}{\sqrt{\sigma^2}} (y_i - x_i'\beta) \right) \right]$ in $l(\beta, \lambda, \sigma^2)$ has small values of partial derivatives in the first-order conditions, $\tilde{k}_0$ should be close to $\hat{k}_0$.

It is worthwhile to note that k_0 is not a model parameter here, and is selected by the proposed selection criteria only for finite sample correction. Thus, choosing k_0 is inherently different from model selection in the literature, such as, choosing the number of model parameters, where consistency is a primary requirement for the penalty term. Therefore, we could use different penalty terms in $BIC(k_0)$ or $C(k_0)$ above as long as a unique value of k_0 can be chosen. The Monte Carlo experiments and empirical example below suggest that the proposed selection criteria work well.[15]

---

[15]Per a referee's advice, we experimented with selecting k_0 by minimizing the mean integrated squared error of the difference between the constrained and unconstrained residual densities, but the selection performed poorly in terms of the RMSE of the estimated coefficients in finite samples.

4 Constrained Estimators

With the proposed finite sample adjustment, the sample can still be used to construct a point estimate for inferring population parameters in the presence of wrong skewness. This is similar in spirit to Simar and Wilson (2010), who still rely on the MLE estimation results, but provide more accurate interval estimates using improved inference (bagging) methods.

As previously mentioned, any negative constraint on sample skewness is binding in the presence of wrong skewness. This result implies that estimated λ (or k) is implicitly determined by the constraint (10). Consequently, it is biased when the selected value of k_0, the lower bound of k, is not equal to the true value of k. Inconsistency of the proposed constrained estimators might be a concern. However, this concern may be overstated. Under the true specification, as the sample size increases, wrong skewness is less likely to appear. Thus, the proposed finite sample adjustment becomes unnecessary, and asymptotics are less of a concern. In addition, with the nature of finite sample adjustment, the proposed method is regarded as an adjustment to existing estimators, rather than a new estimator.[16]

In the next subsection, properties of constrained estimators are studied. Since the constrained COLS is essentially restricted least squares, which has an analytical solution, we mainly focus on it.

4.1 Constrained COLS

The proposed constrained COLS, denoted by $\hat{\beta}_r$, is a 2-step estimator. In the first step, for a given k_0, the constrained COLS $\hat{\beta}_r(k_0)$ is defined as the solution of

$$
\begin{aligned}
\min_{\beta} SSR(\beta) &= \min_{\beta}(Y - X\beta)'(Y - X\beta) \\
s.t. \ R\beta &= q(k_0).
\end{aligned}
$$

In the second step, k_0 is selected such that $\hat{k}_0 = \arg\min_{k_0} C(k_0)$, where $C(k_0) = \frac{1}{N}(Y - X\hat{\beta}_r(k_0))'(Y - X\hat{\beta}_r(k_0)) - k_0\hat{\sigma}_\varepsilon^2 \frac{\ln N}{N}$. The proposed constrained COLS is defined as $\hat{\beta}_r = \hat{\beta}_r(\hat{k}_0)$.

This 2-step estimator is equivalent to a 1-step penalized least squares with the linear constraint:

$$
\min_{\beta, k_0} \frac{1}{N}(Y - X\beta)'(Y - X\beta) - k_0\hat{\sigma}_\varepsilon^2 \frac{\ln N}{N}
$$
$$
s.t. \ R\beta = q(k_0).
$$

---

[16]In this sense, our approach is different from the literature on models with moment conditions, e.g., Moon and Schorfheide (2009).

This equivalence comes from the fact that in the objective function k_0 only appears in the penalty term $-k_0 \hat{\sigma}_\varepsilon^2 \frac{\ln N}{N}$. Thus, β can be concentrated out for a given k_0.

For a given k_0, $\hat{\beta}_r(k_0)$ is the restricted least square. By Amemiya (1985) or Greene (2012),

$$\hat{\beta}_r(k_0) = \hat{\beta}_{OLS} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}[R\hat{\beta}_{OLS} - q(k_0)],$$

and

$$SSR_r(k_0) = SSR + [R\hat{\beta}_{OLS} - q(k_0)]'[R(X'X)^{-1}R']^{-1}[R\hat{\beta}_{OLS} - q(k_0)].$$

Thus, the criterion is

$$C(k_0) = \frac{1}{N}SSR + \frac{1}{N}[R\hat{\beta}_{OLS} - q(k_0)]'[R(X'X)^{-1}R']^{-1}[R\hat{\beta}_{OLS} - q(k_0)] - k_0 \hat{\sigma}_\varepsilon^2 \frac{\ln N}{N}.$$

Minimizing $C(k_0)$ defines $\hat{k}_0$. The follow proposition proves the existence and uniqueness of $\hat{k}_0$.

Proposition 1 *In the presence of positive skewness of OLS residuals, i.e., $\hat{\mu}'_3 > 0$, (i) $\frac{dSSR_r(k_0)}{dk_0} > 0$; (ii) for a relatively large sample size N, there exists a solution for $\hat{k}_0$ such that $\hat{k}_0$ minimizes $C(k_0)$; (iii) $\frac{d^2 C(k_0)}{dk_0^2} > 0$, implying that $\hat{k}_0$ is the unique solution.*

The proof in the Appendix shows that a relatively large N guarantees the existence of $\hat{k}_0$. Since $\frac{\ln N}{N} \to 0$, when $N \to \infty$, compared with the first term $\frac{1}{N}SSR_r(k_0)$, which converges to a non-zero constant, the penalty term $-k_0 \hat{\sigma}_\varepsilon^2 \frac{\ln N}{N}$ in $C(k_0)$ can be ignored asymptotically. This implies that $\hat{k}_0 \to 0$ as $N \to \infty$. Hence, when N is large the proposed constrained COLS approaches the OLS with constraint $R(\beta - \hat{\beta}_{OLS}) = \hat{\mu}'_3/3$, which is very close to OLS in the presence of wrong skewness.

This property also implies that in a sample with a large number of firms, the selected $\hat{k}_0$ could be 0. In this case, to obtain non-zero inefficiency estimates, a small positive value, say, 0.05, is suggested.[17]

For a given sample, the difference between OLS and the constrained COLS

$$\hat{\beta}_{OLS} - \hat{\beta}_r = (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}[R\hat{\beta}_{OLS} - q(\hat{k}_0)]$$

depends on $\hat{k}_0$, and

$$\frac{d[\hat{\beta}_{OLS} - \hat{\beta}_r]}{d\hat{k}_0} = -(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}\frac{\Pi}{2}\hat{k}_0^{1/2}(\hat{\mu}'_2)^{3/2}$$

implying that the magnitude of this difference is positively correlated with the chosen value $\hat{k}_0$.

---

[17]When N is large, the wrong skewness problem is less likely to occur unless the inefficiency variance ratio is very small. When it does occur in this setting, alternative approaches including respecification may be considered.

4.2 Constrained MLE

For a given k_0, the constrained MLE $(\hat{\beta}_{CMLE}(k_0), \hat{\lambda}_{CMLE}(k_0), \hat{\sigma}^2_{CMLE}(k_0))$ depends on k_0. Minimizing $BIC(k_0)$ determines the value of k_0, i.e., $\tilde{k}_0 = \arg\min_{k_0 \in [0,1)} BIC(k_0)$. Similar to the constrained COLS, $(\hat{\beta}_{CMLE}, \hat{\lambda}_{CMLE}, \hat{\sigma}^2_{CMLE})$ is defined as $(\hat{\beta}_{CMLE}(\tilde{k}_0), \hat{\lambda}_{CMLE}(\tilde{k}_0), \hat{\sigma}^2_{CMLE}(\tilde{k}_0))$. It can also be written as a penalized maximum likelihood estimator with a constraint,

$$\min_{\beta,\lambda,\sigma^2,k_0} -2l(\beta,\lambda,\sigma^2) - k_0 \ln N$$

$$s.t. R\beta = q(k_0),$$

where $l(\beta,\lambda,\sigma^2) = \frac{N}{2}\ln(\frac{2}{\pi}) - \frac{N}{2}\ln\sigma^2 + \sum_{i=1}^{N}\ln\left[1 - \Phi\left(\frac{\lambda}{\sqrt{\sigma^2}}(y_i - x_i'\beta)\right)\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - x_i'\beta)^2$ defined in (2).

Since there is no analytical solution to the constrained optimization problem above, it is difficult to derive the properties of constrained MLE.

However, dividing by N, $\frac{1}{N}BIC(k_0) = -\frac{2}{N}l_r(k_0) - k_0\frac{\ln N}{N}$, compared with $-\frac{2}{N}l_r(k_0)$, which does not converge to zero, the penalty term $-k_0\frac{\ln N}{N}$ can be asymptotically ignored as $N \to \infty$, implying that $\tilde{k}_0$ tends to 0 as $N \to \infty$. Since $\tilde{k}_0$ is small when N is large, the proposed constrained MLE is expected be close to MLE. Since the MLE of slope parameters is very close to OLS, the constrained MLE and constrained COLS are expected to be close. Similar to the constrained COLS, the selected $\tilde{k}_0$ could be 0 in a sample with a large N. In this case, we also impose a lower bound of, say, 0.05, to obtain non-zero estimated inefficiency.

We now consider the difference between constrained MLE and OLS by examining the first-order conditions of (2). Aigner, Lovell and Schmidt (1977) show that:

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{N}(y_i - x_i'\beta)^2 + \frac{\lambda}{2\sigma^3}\sum_{i=1}^{N}\frac{\phi(\cdot)}{1 - \Phi(\cdot)}(y_i - x_i'\beta) = 0; \tag{11}$$

$$\frac{\partial \ln L}{\partial \lambda} = -\frac{1}{\sigma^2}\sum_{i=1}^{N}\frac{\phi(\cdot)}{1 - \Phi(\cdot)}(y_i - x_i'\beta) = 0; \tag{12}$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{1}{\sigma^2}\sum_{i=1}^{N}(y_i - x_i'\beta)x_i + \frac{\lambda}{\sigma}\sum_{i=1}^{N}\frac{\phi(\cdot)}{1 - \Phi(\cdot)}x_i = 0, \tag{13}$$

where $\phi(\cdot)$ is the standard normal density function. $\phi(\cdot)$ and $\Phi(\cdot)$ are evaluated at $\frac{\lambda}{\sigma}(y_i - x_i'\beta) = \frac{\lambda}{\sigma}\varepsilon_i$. Waldman (1982) shows that in the presence of wrong skewness $\lambda = 0$ and OLS is a local maximum of the log-likelihood.

For our constrained MLE, the constraint (7) or (9) involves the value of k_0, not λ directly. Since λ is a monotonic increasing function of k, $k \geq k_0$ implies

$$\lambda \geq \sqrt{\frac{\pi}{\pi - 2} \frac{1}{(1/k_0 - 1)}}. \tag{14}$$

To show how restricting λ affects the estimation result and how the constrained MLE of β is different from the OLS, consider equation (13).[18] Taking the first-order Taylor expansion at $\lambda = 0$ gives

$$\frac{\phi(\frac{\lambda}{\sigma}\varepsilon_i)}{1 - \Phi(\frac{\lambda}{\sigma}\varepsilon_i)} \approx \sqrt{\frac{2}{\pi}} + \frac{2}{\pi}\frac{\lambda}{\sigma}\varepsilon_i.$$

Thus, (13) becomes

$$
\begin{aligned}
0 &= \sum_{i=1}^{N}(y_i - x_i'\beta)x_i + \lambda\sigma\sum_{i=1}^{N}\frac{\phi(\frac{\lambda}{\sigma}\varepsilon_i)}{1 - \Phi(\frac{\lambda}{\sigma}\varepsilon_i)}x_i \approx \sum_{i=1}^{N}(y_i - x_i'\beta)x_i + \lambda\sigma\sum_{i=1}^{N}(\sqrt{\frac{2}{\pi}} + \frac{2}{\pi}\frac{\lambda}{\sigma}\varepsilon_i)x_i \\
&= (1 + \frac{2}{\pi}\lambda^2)\sum_{i=1}^{N}(y_i - x_i'\beta)x_i + \sqrt{\frac{2}{\pi}}\lambda\sigma\sum_{i=1}^{N}x_i.
\end{aligned}
$$

That is,

$$\sum_{i=1}^{N}(y_i - x_i'\beta)x_i + \frac{\sqrt{\frac{2}{\pi}}\lambda}{(1 + \frac{2}{\pi}\lambda^2)}\sqrt{\sigma^2}\sum_{i=1}^{N}x_i = 0. \tag{15}$$

In matrix form, the equation (15) above can be written as

$$X'y - X'X\beta + \varphi(\lambda)\sqrt{\sigma^2}X'\iota = 0 \tag{16}$$

where $\varphi(\lambda) = \sqrt{\frac{2}{\pi}}\lambda/(1 + \frac{2}{\pi}\lambda^2)$ and ι is the $N \times 1$ vector of ones. Equivalently,

$$\hat{\beta}_{CMLE} \simeq (X'X)^{-1}X'y + \varphi(\lambda)\sqrt{\sigma^2}(X'X)^{-1}X'\iota. \tag{17}$$

In the presence of wrong skewness, OLS (i.e., $\lambda = \varphi = 0$) is a local maximum of the log-likelihood. Under the constraint (14), the estimator of β is adjusted by the second term in equation (17).[19]

Given the fact that $\varphi(\lambda)$ is monotonically increasing in λ in the range $[0, \sqrt{\pi/2} \approx 1.2533]$, the difference between the constrained MLE and the OLS of β is positively related to the value

---

[18]Strictly speaking, restricting λ as a constraint yields a different result from constraint (7). Though the population skewness is equal to $g(k_0)$ and thus a monotonic function of λ, the sample skewness is not a function of λ. However, the insights derived here on the effect of the chosen value of k_0 on estimation still apply.

[19]As pointed out by a referee, the second term on the right-hand side of (17) is a constant in X. See Papadopoulos (2018), pp 338-339. This is due to the fact we only take a first-order Taylor expansion of $\phi(\frac{\lambda}{\sigma}\varepsilon_i)/[1 - \Phi(\frac{\lambda}{\sigma}\varepsilon_i)]$. With higher-order terms included, constrained MLE and COLS involve additional terms.

of λ.[20] The larger λ (or k_0) is imposed, the bigger is the difference between the OLS and the constrained MLE. Furthermore, in a given sample this difference depends not only on $\varphi(\lambda)$, but also on the sample value of the regressors and σ^2 jointly determined by first-order equations. We conjecture that constraint (10) with a small value of k_0, slightly adjusts the estimators of β and σ_v^2, but has a much larger effect on the estimated σ_u^2 and λ. This point is confirmed in the Monte Carlo experiments and empirical example below.

5   Monte Carlo Experiments

In this section, Monte Carlo experiments are conducted to study how the proposed constraints affect the estimates and how the chosen value of k_0, the imposed lower bound of k, is affected by the sample size. Since respecification is always a solution to wrong skew, we also study our approach under model misspecification. We consider two forms of misspecification: when the inefficiency distribution is binomial (Carree, 2002) and when it is exponential. The binomial case is particularly relevant since it is a specification that allows for residual skewness in either direction. We compare our approach under misspecification to the correctly specified estimators in term of the Root Mean Squared Error (RMSE) of the estimated efficiency variance and average efficiency, and the proposed method's performance is surprisingly good (given the misspecification) when the sample size is small. Due to its computational convenience, our focus is on the constrained COLS estimator.

We consider a specification

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i, \ \varepsilon_i = -u_i + v_i, \ \ i = 1, \cdots, N,$$

where $\beta_0 = 1$, $\beta_1 = 0.8$, $\beta_2 = 0.2$, $x_{1i} \sim \log(|N(4, 100)|)$, $x_{2i} \sim \log(|N(2, 60)|)$, $v_i \sim N(0, \sigma_v^2)$ and $u_i \sim |N(0, \sigma_u^2)|$. $k = Var(u_i)/Var(\varepsilon_i)$ is the inefficiency variance ratio.[21] $\sigma_u^2 = \frac{\pi}{\pi-2}Var(u_i) = \frac{\pi-2}{\pi}kVar(\varepsilon_i)$ and $\sigma_v^2 = (1-k)Var(\varepsilon_i)$. We set $Var(\varepsilon_i) = \sigma_v^2 + Var(u_i) = 0.026$, so the variance of x_{1i} and $Var(\varepsilon_i)$ are comparable to those in the empirical example below.

Since the focus is the proposed correction for samples with wrong residual skewness, we report the frequency of wrong skewness in the replications and drop the samples with correct skewness. The number of replications is $4,000$ before dropping the samples with correct skewness. We conduct

---

[20] For a small value of k_0, e.g., $k_0 \in [0.1, 0.3]$, λ lies in the interval $[0.5530, 1.0860]$.

[21] Coelli (1995) also uses this signal-to-noise ratio measure, denoted by γ^*, in his Monte Carlo experiments.

experiments with $k = 0.1, 0.2, 0.3, 0.5, 0.7$ and $N = 50, 100, 200$. For our binomial misspecification experiments we generate $u_i \sim B(10, p)$ with $Var(u) = p(1 - p)$. For exponential misspecification, we generate $u_i \sim Exp(\sigma_u)$ with $Var(u) = \sigma_u^2$. The same inefficiency variance ratio and sample size combinations are considered in both cases.

Table 1 reports the simulation results. Column (2) gives the average value of $\hat{k}_0$. To obtain $\hat{k}_0$ for each sample, a grid search is conducted to minimize $C(k_0)$ on the interval $[0.05, 0.9]$. As expected, the average value of $\hat{k}_0$ decreases with N. Column (3) shows that there is still a possibility of wrong skewness after the proposed finite sample correction. The frequency depends on the inefficiency variance ratio and sample size, varying from 16.3% to 39.9%. For example, for $k = 0.5$, $N = 100$, our finite sample correction approach could fail with a possibility of 28.4%. This failure is a cost of the linearization approximation (8). When k_0 is small, $g(k_0)(\hat{\mu}_2')^{3/2}$ could be a small negative value close to zero. Consequently, due to approximation error, a linearized constraint does not guarantee a negative third moment of residuals or skewness. However, as k increases, the failure frequency can be greatly reduced, e.g., to 16.3% for $k = 0.7$, $N = 200$.

For parameter estimators, columns (4)-(7) indicate that with the correction of $\hat{\sigma}_u\sqrt{2/\pi}$, constrained COLS of β_0 is less biased than OLS, but with a much larger root mean squared errors (RMSE). But when k and N increase, the RMSE of constrained COLS is comparable to that of OLS. (Bias and RMSE of OLS of β_0 (and β_1) are included in columns (6), (8) (and (10), (12)) for comparison). In addition, compared with OLS, the constrained COLS of β_1 is slightly upward biased with bigger RMSE, and the bias and RMSE decrease with k and N.

In the presence of wrong skewness, σ_u^2 is typically assumed to be zero. Using our correction, column (12) shows that the estimated σ_u^2 tends to be overestimated for a small value of k and underestimated for a big value of k. Compared with σ_u^2, σ_v^2 can be estimated more accurately in terms of bias, as indicated in column (14).

Columns (14) and (15) report the bias and RMSE of the mean technical efficiency $E[\exp(-u_i)] = 2\exp(\sigma_u^2/2)[1 - \Phi(\sigma_u)]$. In the presence wrong skewness, traditional practice suggests that the estimated σ_u^2 is 0, implying that the estimated mean technical efficiency is 1. This practice obviously overestimates the true mean technical efficiency. Column (14) shows that the mean technical efficiency estimator using the proposed correction could be unbiased with a reasonable RMSE for a sizable value of k, say, 0.2 here under the current design. It is downward biased for a small value of k, and upward biased for $k > 0.2$.

17

Under binomial misspecification we focus on the performance of the variance estimate of σ_u^2 and the mean technical efficiency score $E(\mathbf{e^{-u}})$. The binomial case is particularly relevant since it is a specification that accommodates residual skewness in either direction. The results are reported in Table 2. Columns (1)-(3) are very similar to those in Table 1 except that the frequency of wrong skewness after applying the proposed finite sample correction is higher. This makes sense considering the misspecified model in the current setting. Columns (4)-(7) report the coefficient estimates and Columns (12)-(15) for Carree's (2002) approach, which is OLS estimated with a binomial inefficiency. Columns (8)-(11) report the bias and RMSE of the variance estimates and mean technical efficiency estimates for the proposed constrained COLS, and those using the Carree (2002) approach are in columns (16)-(19). We find that the proposed method works well (i.e., smaller bias and RMSE) when the sample size or k is small.[22] As sample size increases, the precision of binomial estimation improves very fast, and its is similar to that of the proposed method.

We also conduct misspecified simulations, assuming that inefficiency follows an exponential distribution, while applying our proposed finite sample correction in Section 4 based on a half-normal distribution. The results are reported in Table 3. In general, the performance of proposed estimator is similar to that in Table 1 (the normal-half normal design), except for that of the mean technical efficiency, which depends heavily on the distribution of the data generating process. The bias of the variance estimator is small, indicating a degree of robustness of our approach.

6 Empirical Example: the US Airline Industry

In this section, an airlines example is used to illustrate our approach. This is an unbalanced panel data set with 256 observations. See Greene (2007) for detailed information of this data set. In this example, the dependent variable is the logarithm of output and the independent variables include the logarithm of fuel, materials, equipment, labor and property. Here, the unbalanced panel is treated as a cross section for 256 firms to ensure that the wrong skewness issue arises.[23] Column (2) of Table 4 presents the OLS estimates along with standard errors (column 3). Except for the

---

[22]Due to misspecification and small sample, the variance estimates with proposed method do not show a consistency property and may yield extreme estimates in some cases, for instance $N = 200$ with $k = 0.5$.

[23]With the exception of perhaps Green and Mayes (1991), Mester (1997) and Parmeter and Racine (2012), there appear to be very few empirical studies with wrong skewness in the literature. As in Greene (2007, Table 2.11), we use this panel data example as a cross-sectional one only for the purpose of illustration.

constant term, the slope coefficients are consistent with Table 2.11 in Greene (2007). The OLS residual skewness (0.0167) is in the wrong direction for the estimated normal-half normal model. Thus, the estimates of λ and σ_u^2 are set to zero and firms are considered to be "super efficient". However, Greene (2007, footnote 84) does suggest that there is evidence of technical inefficiency in the data. The second root of the likelihood with positive λ is reported in the second section of Table 4. This MLE yields a small positive residual skewness 0.0093.[24] Usually, in the presence of "wrong" skewness, researchers are advised to obtained a new sample or respecify the model. For the purpose of comparison, we also report the 95% confidence (prediction) intervals for MLE obtained using the bagging procedure of Simar and Wilson (2010) in column (4) and corresponding standard errors in column (5).

Instead, we use the constrained MLE (and constrained COLS), a finite sample adjustment to the existing MLE (and COLS). The optimal value of k_0 can be chosen by $BIC(k_0)$ (and $C(k_0)$ for the constrained COLS) proposed above. For purposes of illustration, we present constrained MLE results of $k_0 = 0.05$, 0.1, 0.15, and 0.2 in columns (8)-(15) of Table 4 and compare the values of $BIC(k_0)$, showing that $\tilde{k}_0 = 0.15$ achieves the minimum of $BIC(k_0)$. Thus, the constrained MLE of λ and σ_u are positive, 0.6829 and 0.1015 respectively. We also report the mean technical efficiency, i.e., $E(e^{-u})$ for comparison. Furthermore, consistent with the negative population skewness of the composed error, the skewness of constrained MLE residuals (-0.0599) has the desired sign.

Since the constraint slightly adjusts the coefficients of constrained MLE, as expected, the rest of the coefficients are very close to the unconstrained MLE and OLS. For example, the constrained estimated coefficient of variable *Log fuel* is 0.3907 (column (12)), while its unconstrained counterpart is 0.3836 (in column 4) and OLS coefficient is 0.3828 (in column (2)). All estimates obtained from constrained MLE stay within the bagging confidence intervals. In addition, the bagging standard errors are closer to those from the constrained MLE in column (12) than those from unconstrained MLE in column (5) or OLS in column (3). These provide a degree of comfort for the validity of our proposed method.

Consistent with the analysis in Section 4.2, the difference between the constrained MLE slope coefficients and its OLS (and unconstrained MLE) counterparts is positively related to the mag-

---

[24]Inconsistent with the statements of Waldman (1982) and Greene (2007) the MLE with positive λ achieves a slightly bigger value of log-likelihood than OLS for this dataset. Similarly, the inconsistency between OLS and MLE in the presence of positive OLS residual skewness by using FRONTIER is discussed by Simar and Wilson (2009). Greene (2007, p.202) notes: "... for this data set, and more generally, when the OLS residuals are positively skewed, then there is a second maximizer of the log-likelihood, OLS, that may be superior to the stochastic frontier."

nitude of k_0. The bigger the value of k_0, the larger is the difference. However, this difference is relatively small. For example, the constrained estimated coefficients of variable *Log fuel* using $k_0 = 0.2$ is 0.3939 (in column (14) of Table 4), compared with the OLS 0.3828 and the unconstrained MLE 0.3836 (in columns (2) and (4) of Table 4). This is also the case for σ_v and σ^2. In stark contrast to this small difference in slope coefficients, the residual skewness and estimated k change significantly, since they are implicitly determined by the chosen value of k_0 in the constraint. Another important point observed in Table 4 is that the value of the likelihood decreases with k_0.[25]

The results of constrained COLS are reported in columns (8)-(15) of Table 5 and are very close to their constrained MLE counterparts for given values of $k_0 = 0.05$, 0.1, 0.15, and 0.2.[26] However, for the constrained COLS, the optimal value of k_0 is 0.1 by applying Mallows' C_p-type criterion $C(k_0)$ proposed above. (Table 5 reports $N \times C(k_0)$ instead of $C(k_0)$.) This is slightly different from $\tilde{k}_0 = 0.15$ by minimizing $BIC(k_0)$ in the constrained MLE. Therefore, the constrained COLS of σ_u is 0.0853 and skewness is -0.0325 in column (10).

It is worth mentioning that the value of criterion $C(0.15)$ is nearly equal to $C(0.1)$ in this empirical example, implying that $BIC(k_0)$ for the constrained MLE and $C(k_0)$ for the constrained COLS result in similar optimal values of k_0. Again, all the estimates obtained from constrained COLS stay within the bagging confidence intervals by Simar and Wilson (2010) and the bagging standard errors are very close to those from constrained COLS.

Since the proposed finite sample adjustment restricts the inefficiency variance ratio, it indirectly affects the estimated σ_u^2. In this example, it is 0.1015^2, for the constrained MLE. Consequently, the mean technical efficiency estimate, $2\exp(\hat{\sigma}_u^2/2)[1 - \Phi(\hat{\sigma}_u)]$, depends on the chosen value of k_0. However, efficiency rankings appear to be preserved under different choices of k_0. For the unconstrained MLE, the least efficient firm is the 79th with technical efficiency .8958. If we impose $k_0 = 0.05$, 0.1, 0.15, 0.2 in the constraint, the technical efficiency becomes .8583, .8308, .8015, .7722 respectively, and it remains lowest among the 256 firms. The most efficient firm is the 250th with

---

[25]This property can be obtained by the equation (3) in Waldman (1982, p.278):

$$\Delta l = \frac{\mu^3}{6s^3}\sqrt{\frac{2}{\pi}}\frac{\pi - 4}{\pi}\sum_{i=1}^{N}e_i^3$$

where μ can be regarded as λ changing from 0 as in the analysis in Section 3.1. Since $\pi - 4 < 0$, in the presence wrong skewness ($\sum_{i=1}^{N}e_i^3 > 0$), the log-likelihood decreases with the imposed value of λ (and k_0).

[26]The constant term is calculated by OLS intercept plus $\sqrt{2\hat{\sigma}_u^2/\pi}$. The standard errors formula of the COLS estimators of constant term, σ^2 and γ (not λ) can be found in Coelli (1995).

technical efficiency .9696, 0.9669, .9655, .9644, .9636 for the unconstrained MLE and constrained MLE with $k_0 = 0.05, 0.1, 0.15, 0.2$, respectively. This is also the case for the median firm.

Finally, we also calculate the bagging prediction intervals for efficiency scores of the least efficient firm (i.e., the 79th firm) and the most efficient firm (i.e., the 250th firm): [0.5366, 0.9987] and [0.8865,0.9988], respectively. All above estimates fall safely into these prediction intervals.

7   Conclusions

This paper studies the wrong skewness issue in parametric stochastic frontier models. Following Simar and Wilson's (2010), we consider wrong skewness to be a consequence of estimation in finite samples when the inefficiency variance ratio is small. (Another possibility is that the model is misspecified.) In finite samples the data may fail to be informative enough to detect the existence of inefficiency term in stochastic frontier models, even though the population inefficiency variance ratio could be fairly large. Thus, the resulting residuals could display skewness in either direction with probability of as high as 50%.

As an alternative to the usual "solutions" to the wrong skew problem, we propose a feasible finite sample adjustment to existing estimates. When there is evidence of inefficiency, it is reasonable to impose a lower bound on the inefficiency variance ratio in the normal-half normal model, equivalent to a negative upper bound on the residual skewness. Thus, we propose to use this negative bound on residual skewness as a constraint in the MLE and COLS in the event of wrong skewness.

The idea of the proposed constrained estimators is to slightly adjust the slope coefficients in finite samples. They provide a point estimate that yields a negative residual skewness, though a correct sign of residual skewness is not always guaranteed. Since the constraint is based on k_0, the choice of k_0 affects estimation results. A model selection approach is proposed to select k_0. Monte Carlo experiments show that the bias of constrained estimates is less of a concern when sample size is large and inefficiency variance ratio increases. The proposed method is comparable with existing methods in the literature such as the binomial estimation by Carree (2002). The empirical example in this paper also shows that the value k_0 has little effect on the estimated slope coefficients and σ_v, σ^2, while the residual skewness and estimated k are implicitly determined by the value of k_0. In this sense, the proposed method can be regarded as a finite sample adjustment to existing estimators, rather than a new estimator. When the sample size is large, since wrong skewness is less likely to

occur, such adjustment becomes unnecessary.

References

[1] Aigner, D.J., C.A.K. Lovell, and P. Schmidt, 1977, Formulation and Estimation of Stochastic Frontier Production Function Models, *Journal of Econometrics* 6, 21-37.

[2] Amemiya, T., 1985. *Advanced Econometrics*, Cambridge, MA: Harvard University Press.

[3] Almanidis, P. and R. C. Sickles, 2011, The Skewness Issue in Stochastic Frontier Models: Fact of Fiction? In I. van Keilegom and P. W. Wilson (Eds.), Exploring Research Frontiers in Contemporary Statistics and Econometrics. Springer Verlag, Berlin Heidelberg.

[4] Almanidis, P., J. Qian and R. Sickles, 2014, Stochastic Frontier with Bounded Efficiency, In *Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications*. RC Sickles and WC Horrace (eds.) Springer Science & Business Media, New York, NY, (2014): 47-81..

[5] Badunenko, O., D. Henderson and S. Kumbhakar, 2012, When, Where and How to Perform Efficiency Estimation, *Journal of the Royal Statistical Society*, *Series A*, 175, 863-892.

[6] Bai, J. and S. Ng, 2002. Determining the Number of Factors in Approximate Factor Models, *Econometrica*, 70, 191-221.

[7] Carree, M., 2002, Technological inefficiency and the skewness of the error component in stochastic frontier analysis, *Economics Letters* 77, 101-107.

[8] Coelli, T., 1996, A guide to frontier version 4.1: A computer program for stochastic frontier production and cost function estimation, CEPA working paper No. 96/07, Centre for Efficiency and Productivity Analysis, University of New England, Arimidale, NSW 2351, Australia.

[9] Feng, Q. and W. C. Horrace, 2012, Alternative Technical Efficiency Measures: Skew, Bias and Scale, *Journal of Applied Econometrics* 27, 253-268.

[10] Green, A. and Mayes, D., 1991, Technical Inefficiency in Manufacturing Industries, *Economic Journal* 101, 523-538.

[11] Greene, W., 1980, On the Estimation of a Flexible Frontier Production Model, *Journal of Econometrics*, 3, 101-115.

[12] Greene, W., 1995, LIMDEP Version 7.0 User's Manual, New York: Econometric Software, Inc.

[13] Greene, W., 2007, The Econometric Approach to Efficiency Analysis, in H.O. Fried, C. A. K. Lovell and S. Schmidt, eds., *The Measurement of Productive Efficiency: Techniques and Applications*. New York: Oxford University Press.

[14] Greene, W., 2012, *Econometric Analysis*, 7th edition, Pearson.

[15] Hafner, C., H. Manner and L. Simar, 2019, The "Wrong Skewness" Problem in Stochastic Frontier Models: A New Approach, *Econometric Reviews*, 37, 380-400.

[16] Horrace, W.C. and I.A. Wright, 2020, Stationary Points for Stochastic Frontier Models, *Journal of Business and Economic Statistics*, 38:3, 516-526

[17] Horrace, W.C. and C. Parmeter, 2018, A Laplace Stochastic Frontier Model, *Econometric Reviews* 37 (3), 260-280

[18] Kim, M., Kim, Y. and P. Schmidt, 2007, On the accuracy of bootstrap confidence intervals for efficiency Levels in Stochastic Frontier Models with Panel Data, *Journal of Productivity Analysis*, 28, 165-181.

[19] Kumbhakar, S. and K. Lovell, 2000, *Stochastic Frontier Analysis*, Cambridge University Press, Cambridge, UK.

[20] Q. Li, 1996, Estimating a Stochastic Production Frontier When the Adjusted Error is Symmetric, *Economics Letters*, 52, 221-228.

[21] Kumbhakar, S., C. Parmeter and E. Tsionas, 2013, A Zero Inefficient Stochastic Frontier Model, *Journal of Econometrics*, 172, 66-76.

[22] Meeusen,W., and J. van den Broeck, 1977, Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error, *International Economic Review*, 18, 435–444.

[23] Mester, L.J., 1997, Measuring Efficiency at US Banks: Accounting for Heterogeneity is Important, *European Journal of Operational Research*, 98, 230–242.

[24] Moon, H. R., and F. Schorfheide, 2009, Estimation with Overidentifying Inequality Moment Conditions, *Journal of Econometrics*, 153, 136-154.

[25] Olson, J., Schmidt, P. and D. M. Waldman, 1980, A Monte Carlo Study of Estimators of Stochastic Frontier Production Functions, *Journal of Econometrics*,13, 67-82.

[26] Papadopoulos, A., 2018, The two-tier stochastic frontier framework: theory and applications, models and tools. Ph.D. Thesis, Athens University of Economics and Business.

[27] Parmeter, C. F. and J. S. Racine, 2012, Smooth Constrained Frontier Analysis, *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis: Essays in*

Honor of Halbert L. White Jr., edited by X. Chen and N.E. Swanson, Chapter 18, 463-489, Springer-Verlag: New York.

[28] Simar, L. and P. W. Wilson (2010), Inferences from Cross-Sectional Stochastic Frontier Models, *Econometric Reviews,* 29, 62-98.

[29] Stevenson, R. 1980, Likelihood Functions for Generalized Stochastic Frontier Estimation, *Journal of Econometrics*, 13, 58–66.

[30] Waldman, D. (1982), A Stationary Point for the Stochastic Frontier Likelihood, *Journal of Econometrics* 18, 275-279.

[31] Wang, W.S. and P. Schmidt (2009), On the Distribution of Estimated Technical Efficiency in Stochastic Frontier Models, *Journal of Econometrics* 148, 36-45.

Appendix: Constrained COLS

Proof of Proposition 1.

As defined in Section 3, the constrained COLS is the restricted least squares with the linear constraint

$$R\beta = q(k_0)$$

where $R = \frac{1}{N}\tilde{e}'M_0X$ and $q(k_0) = R\hat{\beta}_{OLS} + \frac{\hat{\mu}_3'}{3} + \frac{\Pi}{3}k_0^{3/2}(\hat{\mu}_2')^{3/2}$, with the slope estimator

$$\hat{\beta}_r(k_0) = \hat{\beta}_{OLS} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}[R\hat{\beta}_{OLS} - q(k_0)]$$

and the corresponding sum of squared residuals

$$SSR_r(k_0) = SSR + [R\hat{\beta}_{OLS} - q(k_0)]'[R(X'X)^{-1}R']^{-1}[R\hat{\beta}_{OLS} - q(k_0)],$$

where $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$. Thus,

$$\frac{dSSR_r(k_0)}{dk_0} = -2[R\hat{\beta}_{OLS} - q(k_0)]'[R(X'X)^{-1}R']^{-1}\frac{dq(k_0)}{dk_0}.$$

Given the facts that $\hat{\mu}_3' > 0$ in the presence of wrong skewness, the scalar

$$
\begin{aligned}
R\hat{\beta}_{OLS} - q(k_0) &= R\hat{\beta}_{OLS} - [R\hat{\beta}_{OLS} + \frac{\hat{\mu}_3'}{3} + \frac{\Pi}{3}k_0^{3/2}(\hat{\mu}_2')^{3/2}] \\
&= -[\frac{\hat{\mu}_3'}{3} + \frac{\Pi}{3}k_0^{3/2}(\hat{\mu}_2')^{3/2}] < 0.
\end{aligned}
$$

In addition,

$$\frac{dq(k_0)}{dk_0} = \frac{1}{2}(\hat{\mu}_2')^{3/2}\Pi k_0^{1/2} > 0,$$

and the scalar $R(X'X)^{-1}R' > 0$ since the matrix $X'X$ is positive definite. Therefore,

$$\frac{dSSR_r(k_0)}{dk_0} > 0.$$

Since $C(k_0) = \frac{1}{N}SSR_r(k_0) - k_0\hat{\sigma}_\varepsilon^2\frac{\ln N}{N}$, the FOC is

$$\frac{1}{N}\frac{dSSR_r(k_0)}{dk_0} - \hat{\sigma}_\varepsilon^2\frac{\ln N}{N} = 0,$$

or

$$-2[R\hat{\beta}_{OLS} - q(k_0)]'[R(X'X)^{-1}R']^{-1}\frac{dq(k_0)}{dk_0} - \hat{\sigma}_\varepsilon^2 \ln N = 0.$$

Substituting $R\hat{\beta}_{OLS} - q(k_0) = -[\frac{\hat{\mu}_3'}{3} + \frac{\Pi}{3}k_0^{3/2}(\hat{\mu}_2')^{3/2}]$ and $\frac{dq(k_0)}{dk_0} = \frac{1}{2}(\hat{\mu}_2')^{3/2}\Pi k_0^{1/2}$ into the equation above, we obtain

$$[\frac{1}{\Pi}\frac{\hat{\mu}_3'}{(\hat{\mu}_2')^{3/2}} + k_0^{3/2}]k_0^{1/2} = \frac{3}{\Pi^2}\frac{\hat{\sigma}_\varepsilon^2}{(\hat{\mu}_2')^3}\ln N \cdot [R(X'X)^{-1}R'].$$

26

The LHS $k_0^2 + \frac{1}{\Pi}\frac{\hat{\mu}_3'}{(\hat{\mu}_2')^{3/2}}k_0^{1/2}$ is a monotonic increasing function of k_0, with a minimum 0 at $k_0 = 0$ and a maximum of $1 + \frac{1}{\Pi}\frac{\hat{\mu}_3'}{(\hat{\mu}_2')^{3/2}}$ at $k_0 = 1$. Since the OLS residual skewness $\frac{\hat{\mu}_3'}{(\hat{\mu}_2')^{3/2}}$ is usually a very small positive number in the presence of wrong skewness, $1 + \frac{1}{\Pi}\frac{\hat{\mu}_3'}{(\hat{\mu}_2')^{3/2}}$ is slightly bigger than 1.

Consider the RHS $\frac{3}{\Pi^2}\frac{\hat{\sigma}_\varepsilon^2}{(\hat{\mu}_2')^3}\frac{\ln N}{N} \cdot N[R(X'X)^{-1}R']$. It is positive. In addition, The positive scalar

$$
\begin{aligned}
N \cdot R(X'X)^{-1}R' &= \frac{1}{N}\tilde{e}'M_0X(X'X)^{-1}X'M_0\tilde{e} = \frac{1}{N}(\tilde{e} - \hat{\mu}_2'\iota)'X(X'X)^{-1}X'(\tilde{e} - \hat{\mu}_2'\iota) \\
&= \frac{1}{N}[\tilde{e}'P_X\tilde{e} - 2\hat{\mu}_2'\tilde{e}'P_X\iota + (\hat{\mu}_2')^2\iota'P_X\iota] \\
&= \frac{1}{N}[\tilde{e}'P_X\tilde{e} - N(\hat{\mu}_2')^2] = \frac{1}{N}\tilde{e}'P_X\tilde{e} - (\hat{\mu}_2')^2,
\end{aligned}
$$

where $P_X\iota = \iota$ and $\tilde{e}'\iota = N\hat{\mu}_2'$. We normalize $\tilde{e}$ by dividing it by its average $\hat{\mu}_2'$, i.e., $\mathring{e} = \tilde{e}/\hat{\mu}_2'$, such that $\mathring{e}'\iota = \tilde{e}'\iota/\hat{\mu}_2' = N$. Thus,

$$
\begin{aligned}
&\frac{3}{\Pi^2}\frac{\hat{\sigma}_\varepsilon^2}{(\hat{\mu}_2')^3}\ln N \cdot [R(X'X)^{-1}R'] \\
&= \frac{3}{\Pi^2}\frac{\hat{\sigma}_\varepsilon^2}{(\hat{\mu}_2')^3}\ln N \cdot \frac{1}{N^2}\tilde{e}'M_0X(X'X)^{-1}X'M_0\tilde{e}] \\
&= \frac{3}{\Pi^2}\frac{\ln N}{N} \cdot \frac{1}{N}\mathring{e}'M_0X(X'X)^{-1}X'M_0\mathring{e}].
\end{aligned}
$$

Since for a large N,

$$
\mathring{e}'M_0X(X'X)^{-1}X'M_0\mathring{e} = \frac{1}{N}\mathring{e}'M_0X(\frac{X'X}{N})^{-1}X'M_0\mathring{e} = O_p(1),
$$

we obtain,

$$
RHS = \frac{3}{\Pi^2}\frac{\ln N}{N} \cdot \frac{O_p(1)}{N}.
$$

For a relatively large sample size N, RHS falls into the unity interval, implying the existence of $\hat{k}_0$ as the solution to $\min_{k_0 \in [0,1)} C(k_0)$.

Uniqueness of $\hat{k}_0$ is guaranteed by the second order condition. The second order derivative of $C(k_0)$ is

$$
\begin{aligned}
\frac{d^2C(k_0)}{dk_0^2} &= \frac{1}{N}\frac{d^2SSR_r(k_0)}{dk_0^2} \\
&= \frac{\Pi^2(\hat{\mu}_2')^3}{3N[R(X'X)^{-1}R']}\frac{d[\frac{1}{\Pi}\frac{\hat{\mu}_3'}{(\hat{\mu}_2')^{3/2}}k_0^{1/2} + k_0^2]}{dk_0} \\
&= \frac{\Pi^2(\hat{\mu}_2')^3}{3N[R(X'X)^{-1}R']}[\frac{1}{2\Pi}\frac{\hat{\mu}_3'}{(\hat{\mu}_2')^{3/2}}k_0^{-1/2} + 2k_0] > 0
\end{aligned}
$$

for any $0 < k_0 < 1$ since OLS residual skewness $\frac{\hat{\mu}_3'}{(\hat{\mu}_2')^{3/2}} > 0$ in the presence of wrong skewness.

Table 1: Monte Carlo Results: Constrained COLS

N	k_0 average	frequency	β_0 bias	OLS bias	RMSE	OLS RMSE	β_1 bias	OLS bias	RMSE	OLS RMSE	σ_u^2 bias	σ_v^2 bias	mean efficiency bias	RMSE
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
							$k=0.1$							
50	0.059	0.399	-0.025	-0.067	0.458	0.088	0.009	0.000	0.112	0.022	0.082	0.006	-0.025	0.698
100	0.051	0.394	-0.037	-0.068	0.572	0.078	0.017	0.000	0.270	0.015	0.091	0.000	-0.023	0.785
200	0.050	0.374	-0.017	-0.067	0.533	0.072	0.005	0.000	0.140	0.011	0.064	-0.001	-0.024	0.718
							$k=0.2$							
50	0.059	0.395	-0.063	-0.096	0.344	0.111	0.009	0.000	0.105	0.022	0.028	0.003	0.006	0.728
100	0.052	0.379	-0.057	-0.095	0.363	0.103	0.006	0.000	0.111	0.015	0.034	0.003	0.007	0.706
200	0.050	0.353	-0.041	-0.095	0.437	0.098	0.004	0.000	0.102	0.011	0.080	0.000	0.006	0.721
							$k=0.3$							
50	0.061	0.366	-0.075	-0.117	0.297	0.130	0.005	0.000	0.087	0.022	0.018	0.004	0.027	0.639
100	0.052	0.340	-0.076	-0.117	0.373	0.123	0.006	0.000	0.097	0.015	0.035	0.006	0.030	0.666
200	0.050	0.312	-0.065	-0.117	0.325	0.120	0.000	0.000	0.091	0.011	0.031	0.002	0.036	0.618
							$k=0.5$							
50	0.062	0.319	-0.110	-0.150	0.272	0.160	0.006	0.000	0.086	0.022	-0.010	0.009	0.061	0.615
100	0.053	0.284	-0.105	-0.150	0.262	0.155	0.003	0.000	0.069	0.015	-0.005	0.008	0.070	0.586
200	0.050	0.248	-0.108	-0.151	0.326	0.153	0.004	0.000	0.087	0.010	0.006	0.009	0.073	0.481
							$k=0.7$							
50	0.068	0.270	-0.138	-0.176	0.215	0.185	0.003	0.000	0.076	0.023	-0.038	0.013	0.092	0.510
100	0.054	0.208	-0.138	-0.179	0.215	0.183	0.003	0.000	0.057	0.016	-0.039	0.013	0.099	0.442
200	0.050	0.163	-0.138	-0.178	0.188	0.180	0.002	0.000	0.045	0.011	-0.041	0.014	0.105	0.385

Notes:

1. Column (3) represents the proportions of wrong skewness after the finite sample correction using a linearized constraint.

2. Columns (5), (7), (9) and (11) refer to the bias and root mean squared error of OLS of β_0 and β_1.

3. The mean technical efficiency is calculated as $E(e^{-u})$.

Table 2: Monte Carlo Results: Comparison between Constrained COLS VS Carree (2002)

			Constrained COLS								Carree's (2002) approach							
	k_0	frequency	β_0		β_1		σ_u^2		mean efficiency		β_0		β_1		σ_u^2		mean efficiency	
N	average		bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
							$k=0.1$											
50	0.058	0.400	0.048	0.319	0.005	0.096	0.048	0.318	-0.084	0.825	0.001	0.057	0.000	0.022	0.713	0.931	1.150	1.449
100	0.052	0.393	0.027	0.251	0.007	0.075	0.022	0.108	-0.078	0.810	-0.003	0.037	0.001	0.015	0.531	0.672	0.600	0.970
200	0.050	0.413	0.017	0.284	0.011	0.090	0.026	0.125	-0.081	0.844	0.000	0.025	0.000	0.010	0.396	0.502	0.195	0.814
							$k=0.2$											
50	0.059	0.395	0.052	0.272	0.003	0.075	0.032	0.198	-0.082	0.815	0.002	0.052	0.000	0.020	0.688	0.871	1.083	1.303
100	0.051	0.403	0.047	0.235	0.004	0.076	0.024	0.095	-0.085	0.828	-0.001	0.033	0.000	0.013	0.547	0.675	0.658	0.928
200	0.050	0.358	0.047	0.294	0.006	0.091	0.032	0.209	-0.081	0.754	-0.001	0.025	0.001	0.010	0.357	0.450	0.086	0.769
							$k=0.3$											
50	0.057	0.418	0.021	0.241	0.010	0.093	0.019	0.173	-0.072	0.852	0.001	0.047	-0.001	0.018	0.788	0.989	1.393	1.480
100	0.051	0.380	0.043	0.255	0.004	0.086	0.022	0.152	-0.076	0.789	0.001	0.032	-0.001	0.013	0.519	0.655	0.582	0.963
200	0.050	0.400	0.037	0.355	0.012	0.093	0.064	0.914	-0.076	0.824	0.002	0.023	0.000	0.009	0.368	0.463	0.128	0.790
							$k=0.5$											
50	0.058	0.443	0.022	0.228	0.006	0.077	0.010	0.130	-0.069	0.895	-0.001	0.042	0.000	0.016	0.839	1.065	1.561	1.610
100	0.051	0.425	0.030	0.399	0.002	0.098	0.047	0.954	-0.063	0.864	-0.001	0.027	0.000	0.011	0.565	0.716	0.739	1.029
200	0.050	0.395	0.100	0.937	-0.013	0.264	0.375	1.893	-0.071	0.816	-0.001	0.019	0.001	0.008	0.346	0.435	0.078	0.783
							$k=0.7$											
50	0.058	0.405	0.041	0.199	0.000	0.057	0.004	0.146	-0.058	0.829	-0.003	0.032	0.001	0.012	0.723	0.892	1.229	1.323
100	0.051	0.383	0.028	0.164	0.004	0.055	-0.007	0.049	-0.050	0.789	0.003	0.022	-0.001	0.008	0.511	0.652	0.591	0.993
200	0.050	0.340	0.031	0.212	-0.002	0.067	-0.003	0.076	-0.053	0.721	-0.001	0.014	0.000	0.006	0.334	0.433	0.060	0.792

Notes:

1. Column (3) represents the proportions of wrong skewness after the finite sample correction using a linearized constraint.

2. The coefficient estimates in Carree (2002) is OLS estimates with a binomial inefficiency.

3. The mean technical efficiency is calculated as $E(e^{-u})$.

Table 3: Monte Carlo Results: Constrained COLS with Exponential Inefficiency

	k_0	frequency	β_0				β_1				σ_u^2	σ_v^2	mean efficiency	
N	average		bias	OLS bias	RMSE	OLS RMSE	bias	OLS bias	RMSE	OLS RMSE	bias	bias	bias	RMSE
(1)	(2)	(3)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
						$k=0.1$								
50	0.062	0.413	-0.030	-0.052	0.471	0.077	0.000	-0.001	0.143	0.022	0.084	0.000	0.870	0.876
100	0.052	0.395	-0.020	-0.050	0.352	0.062	0.012	-0.001	0.094	0.015	0.042	0.000	0.871	0.876
200	0.050	0.348	-0.004	-0.049	0.257	0.056	0.006	-0.001	0.077	0.011	0.027	0.002	0.870	0.875
						$k=0.2$								
50	0.058	0.385	-0.038	-0.078	0.290	0.097	0.008	0.002	0.086	0.023	0.047	0.001	0.855	0.860
100	0.052	0.360	-0.038	-0.073	0.209	0.082	0.003	-0.001	0.059	0.015	0.016	0.004	0.856	0.861
200	0.050	0.270	-0.019	-0.072	0.205	0.077	0.002	0.001	0.058	0.010	0.013	0.005	0.858	0.862
						$k=0.3$								
50	0.062	0.330	-0.066	-0.087	0.349	0.102	0.014	0.001	0.114	0.021	0.011	0.004	0.846	0.852
100	0.053	0.363	-0.027	-0.090	0.573	0.098	-0.002	0.001	0.163	0.015	0.132	0.011	0.848	0.853
200	0.050	0.288	-0.036	-0.085	0.246	0.089	0.000	0.000	0.077	0.010	0.025	0.005	0.849	0.853
						$k=0.5$								
50	0.065	0.290	-0.072	-0.104	0.213	0.119	0.012	-0.001	0.195	0.022	0.002	0.007	0.833	0.838
100	0.054	0.245	-0.061	-0.108	0.156	0.114	0.001	0.000	0.043	0.014	-0.003	0.007	0.833	0.838
200	0.050	0.228	-0.051	-0.108	0.560	0.111	-0.005	0.000	0.144	0.011	0.114	0.001	0.833	0.838
						$k=0.7$								
50	0.075	0.210	-0.091	-0.125	0.133	0.137	0.005	0.000	0.038	0.021	-0.011	0.008	0.820	0.827
100	0.056	0.170	-0.079	-0.128	0.138	0.133	0.003	0.001	0.039	0.014	-0.010	0.009	0.820	0.827
200	0.051	0.143	-0.083	-0.123	0.117	0.126	0.000	0.000	0.023	0.010	-0.013	0.009	0.820	0.827

Notes:

1. Column (3) represents the proportions of wrong skewness after the finite sample correction using a linearized constraint.

2. An exponential inefficiency term is assumed.

3. The mean technical efficiency is calculated as $E(e^{-u})$.

Table 4: Estimates of Airlines Example: Constrained MLE

Dependent variable: Log output

Variables	OLS		MLE				Constrained MLE							
							k_0=0.05		k_0=0.1		k_0=0.15		k_0=0.2	
	estimates	se	estimates	se	bagging inteval	bagging se	estimates	se	estimates	se	**estimates**	**se**	estimates	se
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	**(12)**	**(13)**	(14)	(15)
Log fuel	0.3828	0.0712	0.3836	0.0707	[0.2465, 0.5094]	0.0690	0.3857	0.0696	0.3879	0.0697	**0.3907**	**0.0698**	0.3939	0.0700
Log materials	0.7192	0.0773	0.7167	0.0800	[0.5836, 0.8907]	0.0800	0.7097	0.0685	0.7028	0.0685	**0.6940**	**0.0685**	0.6836	0.0685
Log equipment	0.2192	0.0739	0.2196	0.0731	[0.0709, 0.3554]	0.0736	0.2201	0.0730	0.2206	0.0729	**0.2210**	**0.0729**	0.2215	0.0729
Log labor	-0.4101	0.0645	-0.4114	0.0648	[-0.5368, -0.2802]	0.0638	-0.4146	0.0618	-0.4178	0.0618	**-0.4219**	**0.0618**	-0.4267	0.0618
Log property	0.1880	0.0298	0.1897	0.0336	[0.1386, 0.2436]	0.0278	0.1945	0.0184	0.1992	0.0184	**0.2053**	**0.0184**	0.2126	0.0185
constant	-0.9105		-0.8562	0.1835	[-0.9864, -0.6804]	0.0767	-0.8442	0.1103	-0.8361	0.0959	**-0.8279**	**0.0870**	-0.8199	0.0814
σ_v	0.1608		0.1553	0.0338			0.1527	0.0219	0.1508	0.0196	**0.1486**	**0.0180**	0.1463	0.0169
σ_u	0.0000		0.0676	0.2098			0.0820	0.1079	0.0917	0.0848	**0.1015**	**0.0692**	0.1110	0.0586
σ^2	0.0259		0.0287	0.0182	[0.0228, 0.0485]	0.0085	0.0301	0.0116	0.0311	0.0103	**0.0324**	**0.0095**	0.0337	0.0089
λ	0.0000		0.4351	0.2428	[0.0346, 0.8637]	0.2846	0.5371	0.1288	0.6082	0.1032	**0.6829**	**0.0858**	0.7587	0.0740
$E(e^{-u})$	1.0000		0.9483		[0.5366, 0.9988]		0.9378		0.9308		**0.9239**		0.9172	
skewness	0.0167		0.0093				-0.0115		-0.0325		**-0.0599**		-0.0927	
k	0.0000		0.0644				0.0949		0.1185		**0.1449**		0.1730	
log-likelihood	105.0588		105.0617		105.0617		105.047		105.002		**104.898**		104.707	
BIC(k_0)			-210.123				-210.371		-210.558		**-210.6272**		-210.5235	

Notes:

1. Columns (6) and (7) report bagging intervals and standard errors proposed by Simar and Wilson (2010).

2. Columns (12) and (13) are the constrained MLE based on the selected value of k_0 using the proposed BIC-type criterion.

Table 5: Estimates of Airlines Example: Constrained COLS

Dependent variable: Log output

Variables	OLS estimates	se	MLE bagging interval	bagging se	Constrained MLE k_0=0.1 estimates	se	Constrained COLS k_0=0.05 estimates	se	k_0=0.1 estimates	se	k_0=0.15 estimates	se	k_0=0.2 estimates	se
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Log fuel	0.3828	0.0712	[0.2465, 0.5094]	0.0690	0.3879	0.0697	0.3860	0.0701	**0.3883**	**0.0701**	0.3912	0.0702	0.3947	0.0703
Log materials	0.7192	0.0773	[0.5836, 0.8907]	0.0800	0.7028	0.0685	0.7098	0.0692	**0.7029**	**0.0692**	0.6940	0.0692	0.6835	0.0693
Log equipment	0.2192	0.0739	[0.0709, 0.3554]	0.0736	0.2206	0.0729	0.2197	0.0737	**0.2200**	**0.0738**	0.2204	0.0738	0.2210	0.0739
Log labor	-0.4101	0.0645	[-0.5368, -0.2802]	0.0638	-0.4178	0.0618	-0.4145	0.0625	**-0.4177**	**0.0625**	-0.4218	0.0625	-0.4266	0.0626
Log property	0.1880	0.0298	[0.1386, 0.2436]	0.0278	0.1992	0.0184	0.1944	0.0185	**0.1991**	**0.0186**	0.2052	0.0186	0.2123	0.0186
constant	-0.9105		[-0.9864, -0.6804]	0.0767	-0.8361	0.0959	-0.8619		**-0.8417**		-0.8258		-0.8121	
σ_v	0.1608				0.1508	0.0196	0.1567		**0.1525**		0.1481		0.1437	
σ_u	0.0000				0.0917	0.0848	0.0604		**0.0853**		0.1046		0.1212	
σ^2	0.0259		[0.0228, 0.0485]	0.0085	0.0311	0.0103	0.0282		**0.0305**		0.0329		0.0353	
λ	0.0000		[0.0346, 0.8637]	0.2846	0.6082	0.1032	0.3854		**0.5593**		0.7064		0.8436	
$E(e^{-u})$	1.0000		[0.5366, 0.9988]		0.9308		0.9536		**0.9354**		0.9217		0.9102	
skewness	0.0167				-0.0325		-0.0115		**-0.0325**		-0.0599		-0.0927	
k	0.0000				0.1185		0.0512		**0.1021**		0.1535		0.2055	
SSR	6.596		105.0617				6.598		**6.602**		6.611		6.625	
$N*C(k_0)$	6.596						6.591		**6.588**		6.589		6.597	

Notes:

1. Columns (4) and 5) report bagging intervals and standard errors proposed by Simar and Wilson (2010).

2. Columns (10) and (11) are the constrained COLS based on the selected value of k_0 using the proposed C-type criterion.