

Syracuse University

SURFACE at Syracuse University

Center for Policy Research

Maxwell School of Citizenship and Public
Affairs

10-2015

Wrong Skewness and Finite Sample Correction

Qu Feng

Nanyang Technological University, Singapore, qfeng@ntu.edu.sg

William C. Horrace

Syracuse University, whorrace@maxwell.syr.edu

Guiying Laura Wu

Nanyang Technological University, Singapore, guiying.wu@ntu.edu.sg

Follow this and additional works at: <https://surface.syr.edu/cpr>



Part of the [Economic Policy Commons](#), and the [Economics Commons](#)

Recommended Citation

Feng, Qu; Horrace, William C.; and Wu, Guiying Laura, "Wrong Skewness and Finite Sample Correction" (2015). *Center for Policy Research*. 389.

<https://surface.syr.edu/cpr/389>

This Working Paper is brought to you for free and open access by the Maxwell School of Citizenship and Public Affairs at SURFACE at Syracuse University. It has been accepted for inclusion in Center for Policy Research by an authorized administrator of SURFACE at Syracuse University. For more information, please contact surface@syr.edu.



CENTER FOR POLICY RESEARCH
THE MAXWELL SCHOOL

WORKING PAPER SERIES

Wrong Skewness and Finite Sample Correction in Parametric Stochastic Frontier Models

Qu Feng, William C. Horrace, and Guiying
Laura Wu

Paper No. 154
(Revised) October 2015

ISSN: 1252-3066

426 Eggers Hall
Syracuse University
Syracuse, NY 13244-1020
(315) 443-3114/email: ctrpol@syr.edu

http://www.maxwell.syr.edu/CPR_Working_Papers.aspx

CENTER FOR POLICY RESEARCH –Fall 2015

Leonard M. Lopoo, Director
Associate Professor of Public Administration and International Affairs (PAIA)

Associate Directors

Margaret Austin
Associate Director
Budget and Administration

John Yinger
Trustee Professor of Economics and PAIA
Associate Director, Metropolitan Studies Program

SENIOR RESEARCH ASSOCIATES

Badi H. Baltagi.....	Economics	Yingyi Ma	Sociology
Robert Bifulco	PAIA	Jerry Miner	Economics
Leonard Burman	PAIA	Cynthia Morrow	PAIA
Thomas Dennison	PAIA	Jan Ondrich.....	Economics
Alfonso Flores-Lagunes	Economics	John Palmer.....	PAIA
Sarah Hamersma	PAIA	David Popp	PAIA
William C. Horrace	Economics	Stuart Rosenthal	Economics
Yilin Hou	PAIA	Rebecca Schewe	Sociology
Hugo Jales.....	Economics	Amy Ellen Schwartz	PAIA/Economics
Duke Kao.....	Economics	Perry Singleton.....	Economics
Jeffrey Kubik.....	Economics	Michael Wasylenko.....	Economics
Yoonseok Lee	Economics	Peter Wilcoxon.....	PAIA
Amy Lutz.....	Sociology		

GRADUATE ASSOCIATES

Emily Cardon	PAIA	Laura Rodriguez-Ortiz	PAIA
Brianna Carrier	PAIA	Fabio Rueda De Vivero	Economics
John T. Creedon.....	PAIA	Max Ruppensburg	PAIA
Carlos Diaz.....	Economics	Iuliia Shybalkina.....	PAIA
Alex Falevich	Economics	Kelly Stevens	PAIA
Wancong Fu	Economics	Mary Stottele.....	PAIA
Boqian Jiang	Economics	Anna Swanson.....	PAIA
Yusun Kim	PAIA	Tian Tang.....	PAIA
Ling Li	Economics	Saied Toossi	PAIA
Michelle Lofton	PAIA	Rebecca Wang	Sociology
Judson Murchie	PAIA	Nicole Watson.....	Social Science
Brian Ohl.....	PAIA	Katie Wood	PAIA
Jindong Pang	Economics	Jinqi Ye	Economics
Malcolm Philogene	PAIA	Pengju Zhang	PAIA
William Reed	PAIA	Xirui Zhang	Economics

STAFF

Kelly Bogart.....	Administrative Specialist	Candi Patterson.....	Computer Consultant
Karen Cimilluca.....	Office Coordinator	Mary Santy.....	Administrative Assistant
Kathleen Nasto.....	Administrative Assistant	Katrina Wingle.....	Administrative Assistant

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 re-specification or pulling a new sample is often prescribed. Since wrong skewness is considered a finite sample problem, this paper proposes a finite sample adjustment to existing estimators to obtain the desired direction of residual skewness. This provides another empirical approach to dealing with the so-called wrong skewness problem.

JEL No. C13, C23, D24

Keywords: Stochastic Frontier Model, Skewness, MLE, Constrained Estimators, BIC

*We thank William Greene for providing the airlines dataset. The comments of Peter Schmidt, Robin Sickles, Daniel Henderson and the participants of the 2011 Conference in Honor of Peter Schmidt, Houston TX, and Singapore Economic Review Conference 2015 are appreciated. Financial support from the MOE AcRF Tier 1 research grant at Nanyang Technological University is gratefully acknowledged.

Qu Feng - Email: qfeng@ntu.edu.sg, Tel: +65 6592 1543. Division of Economics, School of Humanities and Social Sciences, 14 Nanyang Drive, Singapore 637332

William C. Horrace - Email: whorrace@maxwell.syr.edu, Tel: 315-443-9061. Center for Policy Research, 426 Eggers Hall, Syracuse, NY 13244-1020

Guiying Laura Wu - Email: guiying.wu@ntu.edu.sg, Tel: +65 6592 1553. Division of Economics, School of Humanities and Social Sciences, 14 Nanyang Drive, Singapore 637332

1 Introduction

In parametric stochastic frontier models, the error term is composed as the sum of a two-sided noise component and a one-sided inefficiency component. For cross-sectional models, the noise distribution is assumed normal, while the inefficiency distribution is usually assumed to be half-normal (Aigner, Lovell and Schmidt, 1977), exponential (Meeusen and van den Broeck, 1977; Aigner, Lovell and Schmidt, 1977), or truncated normal (Stevenson, 1980). It is sometimes gamma (Stevenson, 1980; Greene, 1980). For surveys, see Greene (2007) and Kumbhakar and Lovell (2000).

In the widely used normal-half normal specification of the stochastic frontier production function model, 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 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 or respecify the model (Li, 1996; Carree, 2002; Almanidis and Sickles, 2011; Almanidis, Qian and Sickles, 2014; Hafner, Manner and Simar, 2013). Simar and Wilson (2010) argue that "wrong skewness" is not an estimation or modelling failure, but a finite sample problem that is most likely to occur when the signal-to-noise ratio (the variance ratio of the inefficiency component to the composite error) is small. That is, wrong skewness may not be an indication that the 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. The sample under

¹The 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).

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

³For example, estimating the variance parameters in COLS is invalid in this case. However, as emphasized by Greene (2007, note 29), this problem does not carry over to other model specifications. Several sources can lead to wrong skewness.

study can still be used to infer the model parameters.

We follow Simar and Wilson's (2010) view that wrong skewness is a consequence of a small signal-to-noise ratio in finite samples. 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 signal-to-noise 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 signal-to-noise ratio used in the constraint. A shortcoming of the approach is that in finite samples the linear approximation may not be accurate enough to guarantee a negative sign of residual skewness. In this case, additional finite sample adjustment is required. Monte Carlo experiments suggest that our correction becomes more reliable when the true signal-to-noise ratio increases. The possible failure of correcting the sign of 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 convenience of our approach, shown in the Monte Carlo experiments and empirical example below, would be lost.

The proposed finite sample adjustment provides a point estimate with a correct sign of residual skewness that can be used in applied research. Since wrong skewness can occur fairly regularly (even when efficiency may exist in the population under study), the finite sample adjustment is attractive particularly in cases where the inefficiency distribution is half-normal. It is worthwhile to note 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 signal-to-noise ratio 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. 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. The last section concludes.

2 Wrong Skewness Issue

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

$$y_i = x_i' \beta + \varepsilon_i, \quad i = 1, \dots, N, \quad (1)$$

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 maximizing the likelihood function:

$$\begin{aligned} \ln L(\beta, \lambda, \sigma^2 | (y_i, x_i), i = 1, \dots, N) & \quad (2) \\ &= \frac{N}{2} \ln\left(\frac{2}{\pi}\right) - \frac{N}{2} \ln \sigma^2 + \sum_{i=1}^N \ln \left[1 - \Phi \left(\frac{\lambda}{\sqrt{\sigma^2}} (y_i - x_i' \beta) \right) \right] \\ & \quad - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - x_i' \beta)^2, \end{aligned}$$

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:

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

where $\hat{\mu}'_2$ and $\hat{\mu}'_3$ are the second and third sample central 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.⁴ 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 $\hat{\mu}'_3 > 0$.

The Monte Carlo studies in Simar and Wilson (2010) show that the wrong skewness issue is not rare, even when the signal-to-noise 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

⁴Waldman (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).

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.⁵ 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.⁶ 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 signal-to-noise ratio $Var(u_i)/Var(\varepsilon_i)$ is small.⁷ Since the OLS residuals of a production function regression with $u_i = 0$ display skewness in either direction with probability of 50%, a sample drawn from an SPF model with small signal-to-noise ratio could generate positively skewed residuals with high probability.⁸

⁵Kumbhakar, Parmeter and Tsionas (2013) propose a stochastic frontier model to accommodate the presence of both efficient and inefficient firms in the sample.

⁶Waldman (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."

⁷Badunenko, 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.

⁸As 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 signal-to-noise 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:

$$\begin{aligned} & \max \ln L(\beta, \lambda, \sigma^2 | (y_i, x_i), i = 1, \dots, N) \\ \text{s.t. } & \frac{1}{N} \sum_{i=1}^N \left[\frac{y_i - \bar{y} - x_i' \beta - \bar{x}' \beta}{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - x_i' \beta - \bar{y} - \bar{x}' \beta)^2}} \right]^3 \leq 0, \end{aligned} \quad (4)$$

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.⁹

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 can not be too small, relative to that of the composite error ε_i . Denote the signal-to-noise ratio by

$$k = \text{Var}(u_i) / \text{Var}(\varepsilon_i),$$

⁹This 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 constraint 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.

instead of $\lambda = \sigma_u/\sigma_v$.¹⁰ That is, a lower bound on the signal-to-noise ratio is implicitly imposed, $k \geq k_0$. To develop the relationship between the upper bound of skewness and the lower bound of the signal-to-noise ratio, consider the second and third moment of ε_i . Under the normal-half normal specification, Olson, Schmidt and Waldman (1980) show that

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

and

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

It follows that the skewness of ε_i is

$$E \left[\frac{\varepsilon_i - E(\varepsilon_i)}{\sqrt{\text{Var}(\varepsilon_i)}} \right]^3 = \left[\frac{\sigma_u^2}{\text{Var}(\varepsilon_i)} \right]^{3/2} \sqrt{2/\pi} [(\pi - 4)/\pi] = \left[\frac{\text{Var}(u_i)}{\text{Var}(\varepsilon_i)} \right]^{3/2} \frac{\pi - 4}{\pi - 2} \sqrt{\frac{2}{\pi - 2}},$$

where σ_u^2 is replaced with $\frac{\pi}{\pi-2} \text{Var}(u_i)$. Reparameterizing the skewness in terms of the signal-to-noise ratio, we have $g(k) = k^{3/2} \frac{\pi-4}{\pi-2} \sqrt{\frac{2}{\pi-2}} = -\Pi k^{3/2}$, with a constant $\Pi = \frac{4-\pi}{\pi-2} \sqrt{\frac{2}{\pi-2}} \simeq 0.9953$. Since $\Pi > 0$, $g(k) < 0$ (e.g., $g(0.1) \approx -0.0315$, $g(0.2) \approx -0.0890$ and $g(0.3) \approx -0.1635$) and $g'(k) = -\frac{3}{2} \Pi k^{1/2} < 0$. 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, $g(k) \leq g_0$, is equivalent to a lower bound, denoted by k_0 , of the signal-to-noise ratio, $k \geq k_0$, 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

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

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

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

¹⁰Coelli (1995) also uses this signal-to-noise ratio measure, denoted by γ^* , in his Monte Carlo experiments.

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 to 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, $\hat{\mu}'_2$. 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:

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

where $\frac{\partial f(\beta)}{\partial \beta} \Big|_{\hat{\beta}_{OLS}}$ 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

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

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

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

or

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

Letting the $N \times 1$ vector \tilde{e} be the squared OLS residual vector $(e_1^2, \dots, e_N^2)'$, 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 $\iota = (1, \dots, 1)'$. Thus, the linear constraint above can be written as

$$R\beta = q(k_0) \quad (10)$$

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).

¹¹It 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.

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 signal-to-noise ratio is very small.

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 signal-to-noise 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 signal-to-noise ratio increases.¹³ In the spirit of this trade-off we develop 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.

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

¹³Table 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.

For the 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 $BIC(k_0)$:

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

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.

¹⁴The 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,

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

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.

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. The wrong skewness is a finite sample issue under the true specification. As the sample size increases, wrong skewness is less likely to appear, so the proposed finite sample adjustment becomes unnecessary. Thus, asymptotics are less of a

concern here. 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.¹⁵

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

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

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:

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

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

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

¹⁵In this sense, our approach is different from the literature on models with moment conditions, e.g., Moon and Schorfheide (2009).

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 reasonable sample size N , there exists a solution for \hat{k}_0 such that \hat{k}_0 minimizes $C(k_0)$; (iii) $\frac{d^2C(k_0)}{dk_0^2} > 0$, implying that \hat{k}_0 is the unique solution.*

The proof is in Appendix.

Since $\frac{\ln N}{N} \rightarrow 0$, when $N \rightarrow \infty$, 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 $\hat{k}_0 \rightarrow 0$ as $N \rightarrow \infty$. 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 avoid wrong skewness, a small positive value, say, 0.05, is suggested.

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 .

4.2 Constrained MLE

For a given k_0 , the constrained MLE $(\hat{\beta}_{CMLE}(k_0), \hat{\lambda}_{CMLE}(k_0), \hat{\sigma}_{CMLE}^2(k_0))$ 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, $(\hat{\beta}_{CMLE}, \hat{\lambda}_{CMLE}, \hat{\sigma}_{CMLE}^2)$ is defined as $(\hat{\beta}_{CMLE}(\tilde{k}_0), \hat{\lambda}_{CMLE}(\tilde{k}_0), \hat{\sigma}_{CMLE}^2(\tilde{k}_0))$.

It can also be written as a penalized maximum likelihood estimator with a constraint,

$$\begin{aligned} \min_{\beta, \lambda, \sigma^2, k_0} \quad & -2l(\beta, \lambda, \sigma^2) - k_0 \ln N \\ \text{s.t.} \quad & R\beta = q(k_0), \end{aligned}$$

where $l(\beta, \lambda, \sigma^2) = \frac{N}{2} \ln\left(\frac{2}{\pi}\right) - \frac{N}{2} \ln \sigma^2 + \sum_{i=1}^N \ln \left[1 - \Phi\left(\frac{\lambda}{\sqrt{\sigma^2}}(y_i - x'_i\beta)\right) \right] - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - x'_i\beta)^2$ 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 $N \rightarrow \infty$, implying that \tilde{k}_0 tends to 0 as $N \rightarrow \infty$. Since \tilde{k}_0 is small when N is large, the proposed constrained MLE is expected to 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 avoid wrong skewness.

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:

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

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

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

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

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

To show how restricting λ affects the estimation result and how the constrained MLE of β is different from the OLS, consider equation (13).¹⁶ 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

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

That is,

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

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

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

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

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

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).

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

¹⁶Strictly 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.

of λ .¹⁷ 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. Different from the constrained MLE, the constrained COLS has an analytical solution and established results are in the previous section. Thus, it is computationally convenient to use the constrained COLS in Monte Carlo experiments, and the main results can be carried over to the constrained MLE.

We consider a specification

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

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 $u_i \sim |N(0, \sigma_u^2)|$. $k = \text{Var}(u_i)/\text{Var}(\varepsilon_i)$ is the signal-to-noise ratio.¹⁸ $\sigma_u^2 = \frac{\pi}{\pi-2} \text{Var}(u_i) = \frac{\pi-2}{\pi} k \text{Var}(\varepsilon_i)$ and $\sigma_v^2 = (1-k)\text{Var}(\varepsilon_i)$. We set $\text{Var}(\varepsilon_i) = \sigma_v^2 + \text{Var}(u_i) = 0.026$, so the variance of x_{1i} and $\text{Var}(\varepsilon_i)$ are comparable to those in the empirical example below.

Since the focus is the proposed correction for samples with wrong residual skewness, we drop the samples with correct skewness. The number of replications is 4000 after dropping the samples with correct skewness. We conduct experiments with $k = 0.1, 0.2, 0.3, 0.5, 0.7$ and $N = 50, 100$,

¹⁷For a small value of k_0 , e.g., $k_0 \in [0.1, 0.3]$, λ lies in the interval $[0.5530, 1.0860]$.

¹⁸Coelli (1995) also uses this signal-to-noise ratio measure, denoted by γ^* , in his Monte Carlo experiments.

200.¹⁹

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 signal-to-noise 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 (5), (7) (and (9), (11)) 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 (13).

Column (14) of Table 1 shows that k is generally underestimated. This is due to the fact that a relatively small value of \hat{k}_0 is often chosen when N is large, and that the estimated k is implicitly determined by \hat{k}_0 suggested by the linear constraint (10).²⁰

¹⁹Our Monte carlo experiments show that it is very unlikely to have wrong skewness when $k \geq 0.7$ and $N \geq 100$. Results are available upon request.

²⁰But this is not a big concern since k is not a parameter of interest in this model.

Finally, column (15) reports the bias 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 (15) shows that the mean technical efficiency estimator using the proposed correction could be unbiased 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 $k > 0.2$.

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.²¹ Column (2) of Table 2 presents the OLS estimates along with standard errors (column 3). Except for the 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 2. This MLE yields a small positive residual skewness 0.0093.²² Usually, in the presence of "wrong" skewness, researchers are advised to obtain a new sample or respecify the model.

Instead, we use the constrained MLE (and constrained COLS), a finite sample adjustment to

²¹With 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.

²²Inconsistent 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."

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 (6)-(13) of Table 2 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. 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 10), while its unconstrained counterpart is 0.3836 (in column 4) and OLS coefficient is 0.3828 (in column 2).

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 magnitude 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 12 of Table 2), compared with the OLS 0.3828 and the unconstrained MLE 0.3836 (in columns 2 and 4 of Table 2). 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 2 is that the value of the likelihood decreases with k_0 .²³

The results of constrained COLS are reported in columns (6)-(13) of Table 3 and are very close

²³This property can be obtained by the equation (3) in Waldman (1982, p.278):

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

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 of wrong skewness ($\sum_{i=1}^N e_i^3 > 0$), the log-likelihood decreases with the imposed value of λ (and k_0).

to their constrained MLE counterparts for given values of $k_0 = 0.05, 0.1, 0.15, \text{ and } 0.2$.²⁴ 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 3 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 (8).

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 .

Since the proposed finite sample adjustment restricts the signal-to-noise 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 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.

7 Conclusions

This paper studies the wrong skewness issue in parametric stochastic frontier models. Following Simar and Wilson's (2010) opinion, we consider wrong skewness to be a consequence of estimation in finite samples when the signal-to-noise ratio is small. 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 signal-to-noise ratio could be fairly large. Thus, the resulting residuals could

²⁴The 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).

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 signal-to-noise 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 signal-to-noise ratio increases. 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., 1995, Estimators and Hypothesis Tests for a Stochastic Frontier Function: A Monte Carlo Analysis, *Journal of Productivity Analysis*, 6, 247-268.
- [9] 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, Armidale, NSW 2351, Australia.
- [10] Feng, Q. and W. C. Horrace, 2012, Alternative Technical Efficiency Measures: Skew, Bias and Scale, *Journal of Applied Econometrics* 27, 253-268.
- [11] Green, A. and Mayes, D., 1991, Technical Inefficiency in Manufacturing Industries, *Economic Journal* 101, 523-538.
- [12] Greene, W., 1980, On the Estimation of a Flexible Frontier Production Model, *Journal of Econometrics*, 3, 101-115.
- [13] Greene, W., 1995, LIMDEP Version 7.0 User's Manual, New York: Econometric Software, Inc.

- [14] Greene, W., 2007, The Econometric Approach to Efficiency Analysis, in H.O. Fried, C. A. K. Lovell and S. S. Schmidt, eds., *The Measurement of Productive Efficiency: Techniques and Applications*. New York: Oxford University Press.
- [15] Greene, W., 2012, *Econometric Analysis*, 7th edition, Pearson.
- [16] Hafner, C., H. Manner and L. Simar, 2013, The “Wrong Skewness” Problem in Stochastic Frontier Models: A New Approach, working paper.
- [17] 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.
- [18] Kumbhakar, S. and K. Lovell, 2000, *Stochastic Frontier Analysis*, Cambridge University Press, Cambridge, UK.
- [19] Q. Li, 1996, Estimating a Stochastic Production Frontier When the Adjusted Error is Symmetric, *Economics Letters*, 52, 221-228.
- [20] Kumbhakar, S., C. Parmeter and E. Tsionas, 2013, A Zero Inefficient Stochastic Frontier Model, *Journal of Econometrics*, 172, 66-76.
- [21] Meeusen, W., and J. van den Broeck, 1977, Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error, *International Economic Review*, 18, 435-444.
- [22] Mester, L.J., 1997, Measuring Efficiency at US Banks: Accounting for Heterogeneity is Important, *European Journal of Operational Research*, 98, 230-242.
- [23] Moon, H. R., and F. Schorfheide, 2009, Estimation with Overidentifying Inequality Moment Conditions, *Journal of Econometrics*, 153, 136-154.
- [24] 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.
- [25] 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.
- [26] Simar, L. and P. W. Wilson (2010), Inferences from Cross-Sectional Stochastic Frontier Models, *Econometric Reviews*, 29, 62-98.

- [27] Stevenson, R. 1980, Likelihood Functions for Generalized Stochastic Frontier Estimation, *Journal of Econometrics*, 13, 58-66.
- [28] Waldman, D. (1982), A Stationary Point for the Stochastic Frontier Likelihood, *Journal of Econometrics* 18, 275-279.
- [29] 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

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

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

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

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., $\hat{e} = \tilde{e}/\hat{\mu}'_2$, such that $\hat{e}' \iota = \tilde{e}' \iota / \hat{\mu}'_2 = N$. Thus,

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

Since for a large N ,

$$\hat{e}' M_0 X (X'X)^{-1} X' M_0 \hat{e} = \frac{1}{N} \hat{e}' M_0 X \left(\frac{X'X}{N} \right)^{-1} X' M_0 \hat{e} = O_p(1),$$

we obtain,

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

For a relative 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

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

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				β_1				σ_u^2	σ_v^2	k	mean efficiency
			bias	OLS bias	rmse	OLS rmse	bias	OLS bias	rmse	OLS rmse	bias	bias	bias	bias
(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	0.083	-0.025
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	0.072	-0.023
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	0.079	-0.024
$k=0.2$														
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	-0.026	0.006
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	-0.036	0.007
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	-0.045	0.006
$k=0.3$														
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	-0.128	0.027
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	-0.161	0.030
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	-0.164	0.036
$k=0.5$														
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	-0.352	0.061
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	-0.380	0.070
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	-0.400	0.073
$k=0.7$														
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	-0.571	0.092
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	-0.606	0.099
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	-0.624	0.105

Note: 1. The frequency of wrong skewness after the proposed finite sample correction.

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

Table 2: Estimates of Airlines Example: Constrained MLE

Dependent variable: Log output

Variables	<u>OLS</u>		<u>MLE</u>		$k_0=0.05$		$k_0=0.1$		$k_0=0.15$		$k_0=0.2$	
	estimates	se	estimates	se	estimates	se	estimates	se	estimates	se	estimates	se
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log fuel	0.3828	0.0712	0.3836	0.0707	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	0.7097	0.0685	0.7028	0.0685	0.6940	0.0685	0.6836	0.0685
Log equipment	0.2192	0.0739	0.2196	0.0731	0.2201	0.0730	0.2206	0.0729	0.2210	0.0729	0.2215	0.0729
Log labor	-0.4101	0.0645	-0.4114	0.0648	-0.4146	0.0618	-0.4178	0.0618	-0.4219	0.0618	-0.4267	0.0618
Log property	0.1880	0.0298	0.1897	0.0336	0.1945	0.0184	0.1992	0.0184	0.2053	0.0184	0.2126	0.0185
constant	-0.9105		-0.8562	0.1835	-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	0.0301	0.0116	0.0311	0.0103	0.0324	0.0095	0.0337	0.0089
λ	0.0000		0.4351	0.2428	0.5371	0.1288	0.6082	0.1032	0.6829	0.0858	0.7587	0.0740
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		105.047		105.002		104.898		104.707	
BIC(k_0)			-210.123		-210.371		-210.558		-210.627		-210.524	

Table 3: Estimates of Airlines Example: Constrained COLS

Dependent variable: Log output

Variables	OLS		Constrained MLE		$k_0=0.05$		$k_0=0.1$		$k_0=0.15$		$k_0=0.2$	
	estimates	se	estimates	se	estimates	se	estimates	se	estimates	se	estimates	se
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log fuel	0.3828	0.0712	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	0.7028	0.0685	0.7098	0.0692	0.7029	0.0692	0.6940	0.0692	0.6835	0.0693
Log equipment	0.2192	0.0739	0.2206	0.0729	0.2197	0.0737	0.2200	0.0738	0.2204	0.0738	0.2210	0.0739
Log labor	-0.4101	0.0645	-0.4178	0.0618	-0.4145	0.0625	-0.4177	0.0625	-0.4218	0.0625	-0.4266	0.0626
Log property	0.1880	0.0298	0.1992	0.0184	0.1944	0.0185	0.1991	0.0186	0.2052	0.0186	0.2123	0.0186
constant	-0.9105		-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		0.0311	0.0103	0.0282		0.0305		0.0329		0.0353	
λ	0.0000		0.6082	0.1032	0.3854		0.5593		0.7064		0.8436	
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				6.598		6.602		6.611		6.625	
$N * C(k_0)$	6.596				6.591		6.588		6.589		6.597	