RUHR
ECONOMIC PAPERS

Mark Andor
Christopher Parmeter

Pseudolikelihood Estimation of the Stochastic Frontier Model

rwi

#693

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

RWI Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #693

Responsible Editor: Manuel Frondel

ISSN 1864-4872 (online) – ISBN 978-3-86788-804-2

Ruhr Economic Papers #693

Mark Andor and Christopher Parmeter

Pseudolikelihood Estimation of the Stochastic Frontier Model

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

Das RWI wird vom Bund und vom Land Nordrhein-Westfalen gefördert.

http://dx.doi.org/10.4419/86788804
ISSN 1864-4872 (online)
ISBN 978-3-86788-804-2

Mark Andor and Christopher Parmeter[1]

Pseudolikelihood Estimation of the Stochastic Frontier Model

Abstract

Stochastic frontier analysis is a popular tool to assess firm performance. Almost universally it has been applied using maximum likelihood estimation. An alternative approach, pseudolikelihood estimation, decouples estimation of the error component structure and the production frontier, has been adopted in both the nonparametric and panel data settings. To date, no formal comparison has yet to be conducted comparing these methods in a standard, parametric cross sectional framework. We produce a comparison of these two competing methods using Monte Carlo simulations. Our results indicate that pseudolikelihood estimation enjoys almost identical performance to maximum likelihood estimation across a range of scenarios and performance metrics, and for certain metrics outperforms maximum likelihood estimation when the distribution of inefficiency is incorrectly specied.

JEL Classification: C1, C5, D2

Keywords: Stochastic frontier analysis; maximum likelihood; production function; Monte Carlo simulation

May 2017

---

1 Mark Andor, RWI; Christopher Parameter; University of Miami, USA. *– We are deeply indebted to the participants of the 14th European Workshop on Efficiency and Productivity Analysis in Helsinki, Finland for providing valuable comments. Furthermore, we would like to thank Artem Prokhorov for helpful comments and suggestions. The authors are responsible for all errors and omissions. We gratefully acknowledge that this work has been partly supported by the Collaborative Research Center "Statistical Modeling of Nonlinear Dynamic Processes" (SFB 823) of the German Research Foundation (DFG), within Project A3, "Dynamic Technology Modeling".**– All correspondence to: Mark Andor, RWI, Hohenzollernstr. 1-3, 45128 Essen,Germany; e-mail: andor@rwi-essen.de*

1. Introduction

The study of firm performance has a long history in economics. Accounting for the presence of inefficiency was a vexing econometric issue until a composed error approach was proposed by Aigner, Lovell & Schmidt (1977) and Meeusen & van den Broeck (1977). This approach, dubbed stochastic frontier analysis (SFA), treats the error term in a standard regression model as stemming from two sources, noise/measurement error and firm level inefficiency. These two separate components can be identified given that inefficiency operates in one direction; for example, in a production context, it lowers output. SFA is almost universally implemented using maximum likelihood (ML). However, an alternative method of moments (MoM) approach was also suggested (e.g. in Aigner et al. 1977, Olson, Schmidt & Waldman 1980) which decouples estimation of the frontier and the unknown parameters of the noise and inefficiency distribution. To date, several simulation based papers exist which have compared the relative performance of ML and MoM estimation of the stochastic frontier model (Olson et al. 1980, Coelli 1995, Behr & Tente 2008).

In this paper we compare an alternative stochastic frontier estimator, based on pseudo-likelihood (PL) estimation, to both ML and MoM. Similar to MoM, PL estimation proceeds by decoupling estimation of the frontier and the parameters of the error component. The PL approach was first suggested by Fan, Li & Weersink (1996) owing to the fact that they wanted to estimate the production frontier in a nonparametric fashion, but doing this in a pure ML framework was unknown at the time. By separating estimation of the frontier and the parameters of the distribution of the composed error term, a range of alternative estimation techniques became available. Also in this vein, Kuosmanen & Kortelainen (2012) used the PL estimator for their model of production and efficiency. Since the frontier in their model needs to be estimated using constrained nonparametric methods, the PL estimator was an obvious choice.

What is interesting about the PL framework is that while it was introduced in a nonparametric context, it is equally applicable in a parametric setting. Fan et al. (1996, pg. 466) even acknowledge the applicability of the PL estimator in the parametric context, noting "...if $g(x_i)$ is linear then $\widehat{E}[y_i|x_i]$ ...can be replaced by the least squares prediction of y_i given x_i.", yet it has rarely been used in applied settings. Furthermore, its performance has never been properly adjudicated against ML. To date, studies that provide Monte Carlo simulation evidence regarding the PL estimator are Andor & Hesse (2014), Badunenko, Henderson & Kumbhakar (2012) and Fan et al. (1996), which all focus on the nonparametric setting.

While the PL estimator has been studied in the nonparametric setting, there are a variety of reasons to use parametric methods. In particular, many nonparametric methods carry finite sample bias, especially for sample sizes that are common in agricultural or energy studies. These biases can be further exacerbated by the number of inputs that are commonly modeled. Another important reason to study the PL estimator is that, recently, the PL estimator was used in a three step procedure to recover both persistent and time varying inefficiency in a parametric panel data stochastic frontier model (Kumbhakar, Lien & Hardaker 2014, Kumbhakar, Wang & Horncastle 2015), which would otherwise require optimization of a complicated maximum likelihood function. To sum up, the PL estimator has rarely see empirical adoption and its performance in regard to ML, as well as MoM, is unknown for the parametric setting, although there are some indications that the PL estimator might have some comparative advantages.

Theoretically, a key advantage of PL – and analogously MoM – is that when either of the distributions of the error component are misspecified, consistent estimators of the shape of the production frontier should still be produced.[1] As Kumbhakar & Lovell (2003, pg. 93) note, referencing MoM, two-stage methods "…use distributional assumptions only in the second step, and so the first-step estimators are robust to distributional assumptions on v_i and u_i". Under distributional misspecification, ML estimation may potentially produce biased and/or inconsistent estimators. Another practical advantage of PL in comparison to ML is that it potentially lessens the (numerical) maximization complexity due to the fact that it reduces the number of variables over which to perform the optimization.

A further contribution of the paper is that it sheds more light on the relative comparison of ML and MoM. To date, papers comparing ML and MoM (Olson et al. 1980, Coelli 1995, Behr & Tente 2008) have focused on estimation of the parameters of the model (slope coefficients and variance parameters of the composed error distribution). All three studies show that both estimators have their strengths and weaknesses. Yet, the three studies come to somewhat diverging conclusions. While Olson et al. (1980, pg.80) reason that "[f]or all sample sizes below 400 and for λ less than 3.16, [MoM] is preferred. But, even for higher sample sizes and variance ratios, the additional efficiency of the [ML] may not be worth the extra

[1]The intercept will still be biased as it depends on the unknown, nonzero mean of the error component.

trouble required to compute it.", Coelli (1995, pg. 264) concludes that "[o]verall, these results suggest the ML estimator should be preferred to the [MoM] estimator …". Lastly, based on their simulation results, Behr & Tente (2008, pg. 16) suggest that "…method of moment estimation should be considered an alternative to maximum likelihood estimation …". However, it should be noted that all three studies draw their conclusions on a generic data generation procedure deliberately for cases with no covariates, arguing that covariates should not matter for relative performance.[2] An additional focus of our Monte Carlo simulation is to learn if the presence of covariates has any meaningful effect on the earlier claims of Olson et al. (1980), Coelli (1995) and Behr & Tente (2008).

Naturally, no Monte Carlo investigation is above reproach and sampling and specification issues naturally preclude making definitive recommendations. However, ignoring covariates in the comparison between the one-step ML approach and the two-stage procedures, MoM and PL, eliminates the key advantage that the latter two methods may possess over ML. Given that earlier papers comparing ML and MoM did not extensively model the frontier, we believe that our simulation results here are instructive. Our main results are twofold: First, when the distribution of the inefficiency term is correctly specified, all three methods have relatively similar performance when estimating returns to scale, inefficiency levels, and firm output; second, when the distribution of inefficiency is misspecified, PL appears to be the dominant method for the majority of sample size/signal-to-noise ratio scenarios considered. In combination, this may suggest use of the PL estimator in settings where there is uncertainty regarding the correct distribution of inefficiency, which is the case in most real world applications.

2. Estimation of the Stochastic Frontier Model

The stochastic production frontier of Aigner et al. (1977) and Meeusen & van den Broeck (1977), across n firms, is

$$y_i = m(\boldsymbol{x}_i; \boldsymbol{\beta}) - u_i + v_i = m(\boldsymbol{x}_i; \boldsymbol{\beta}) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where u_i captures inefficiency (shortfall from maximal output), v_i captures outside influences beyond the control of the producer and the production frontier is $m(\boldsymbol{x}_i; \boldsymbol{\beta})$. We assume that u_i and v_i are independent of one another as well as the covariates $\boldsymbol{x}_i$. Due to the fact that

[2]To be precise, the study of Olson et al. (1980) presents one experiment based on real data with four covariates. However, this experiment is limited from several perspectives (number of replications, maintained assumptions, etc.).

inefficiency can only affect output in one direction, we have that $E(u_i) = \mu > 0$. In contrast, v_i, which can positively or negatively impact output, is assumed to have $E(v_i) = 0$; thus, $E(\varepsilon_i) \neq 0$. The most basic formulation of the stochastic frontier model is to assume that $v_i \sim N(0, \sigma_v^2)$ and $u_i \sim N_+(0, \sigma_u^2)$. In this case, ML estimation proceeds by optimizing $\mathcal{L} = \prod_{i=1}^{n} f(\varepsilon_i)$, where $\varepsilon_i = y_i - m(\boldsymbol{x}_i; \boldsymbol{\beta})$, yielding

$$\ln \mathcal{L}(\beta, \lambda, \sigma) = -n \ln \sigma + \sum_{i=1}^{n} \ln \Phi(-\varepsilon_i \lambda/\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \varepsilon_i^2, \tag{2}$$

with $\Phi(\cdot)$ representing the cumulative distribution function of a standard normal random variable and the parameterization $\sigma = \sigma_u + \sigma_v$ and $\lambda = \frac{\sigma_u}{\sigma_v}$ has been used.

Alternatively, MoM proceeds by estimating the model in (1) using ordinary least squares (OLS). From there, MoM uses the second and third moment conditions for ε to estimate σ_v^2 and σ_u^2. With an estimate for σ_u^2, the intercept of $m(\boldsymbol{x}_i; \boldsymbol{\beta})$ can be shifted up to account for the non-zero mean of the composed error. See Kumbhakar & Lovell (2003) or Greene (2008) for a more detailed account of the exact moment conditions.

PL estimation of the stochastic frontier model proceeds – analogously to MoM – by estimating the model in (1) using OLS. Then, the variance parameters are estimated by maximizing[3]

$$\ln \mathcal{L}(\lambda) = -n \ln \hat{\sigma} + \sum_{i=1}^{n} \ln \Phi \left[\frac{-\hat{\varepsilon}_i \lambda}{\hat{\sigma}} \right] - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2, \tag{3}$$

where $\hat{\varepsilon}_i = \hat{\varepsilon}_{i,OLS} - \frac{\sqrt{2}\lambda\hat{\sigma}}{\sqrt{\pi(1+\lambda^2)}}$ and $\hat{\varepsilon}_{i,OLS}$ are the residuals from OLS estimation of (1) and

$$\hat{\sigma} = \sqrt{\frac{\frac{1}{n}\sum_{j=1}^{n} \hat{\varepsilon}_{j,OLS}^2}{1 - \frac{2\lambda^2}{\pi(1+\lambda)}}}.$$

Subsequently, a consistent estimator for the intercept of the production frontier is given by:

$$\hat{\beta}_0 = \hat{\beta}_{0,OLS} + \hat{E}(u_j) = \hat{\beta}_{0,OLS} + \sqrt{\frac{2}{\pi}} \hat{\sigma}_u.$$

After shifting the OLS frontier upwards by the expected value of the inefficiency term, all of the estimators are unbiased and consistent (see Aigner et al. 1977, Kumbhakar &

[3]Note that this optimization is over the single unknown parameter λ as, from Fan et al. (1996), σ can be concentrated out with the normal-half normal distributional assumptions.

Lovell 2003, Greene 2008). For any of the ML, MoM, or PL estimators, an estimator for expected firm level inefficiency can be obtained through the conditional expectation of u given ε following the approach of Jondrow et al. (1982). Measurement of technical efficiency (Battese & Coelli 1988) follows from

$$\widehat{TE}_i = \hat{E}(e^{-u_i}|\hat{\varepsilon}_j) = \frac{\Phi(\hat{\mu}_{*j}/\hat{\sigma}_* - \hat{\sigma}_*)}{\Phi(\hat{\mu}_{*j}/\hat{\sigma}_*)} \cdot e^{\left(\frac{1}{2}\hat{\sigma}_*^2 - \hat{\mu}_{*j}\right)}, \tag{4}$$

where $\hat{\mu}_* = -\hat{\varepsilon}\hat{\sigma}_u^2/\hat{\sigma}^2$ and $\hat{\sigma}_*^2 = \hat{\sigma}_u^2\hat{\sigma}_v^2/\hat{\sigma}^2$.

In this paper, we focus on the case of a constant variance for firm level inefficiency, u_i, i.e. we assume homoscedasticity. It is important to understand the consequences of this assumption because – independent of the first stage estimator – in the setting where the variance of u_i varies across a set of covariates, the ignorance of the error structure can produce biased and inconsistent estimators of the production frontier parameters (Parmeter & Kumbhakar 2014). Therefore, standard tests of heteroscedasticity should be applied after the first stage OLS estimation (Kuosmanen & Fosgerau 2009). If heteroscedasticity is present, this has to be taken into account. While ML estimation allows direct modeling of heteroscedasticity, PL and MoM approaches which can handle this level are embryonic. A first endeavor in this direction can be found in Kuosmanen, Johnson & Saastamoinen (2015, sect. 7.8). They show that under specific assumptions it is possible to estimate a heteroscedastic frontier model using PL.

3. Monte Carlo Simulation

3.1. **Data generating process and performance criteria.** To assess the performance of the PL approach to ML and MoM, we turn to Monte Carlo experiments. Rather than generate data from a generic production function, we instead base our simulations around a real world dataset. Specifically, we use the Philippines rice dataset which has become a benchmark example in applied efficiency analysis, serving as the dominant heuristic illustration in Coelli et al. (2005) and also appearing recently in Rho & Schmidt (2015). The data are composed of 43 farmers observed annually for eight years. Even though the data constitutes a panel, we will ignore this aspect for our purposes. The output variable is tonnes of freshly threshed rice with the main input variables being area of planted rice (hectares), total labor used (man-days of family and hired-labor) and fertilizer used (kilograms). There is also a fourth input, other inputs, which is measured relative to farm 17 in the data via the Laspeyres index for 1991.[4]

Due to its popularity in applied efficiency analysis, we assume a translog production function in the data generating process. While it is commonly known that the translog functional form can violate axioms of production, its use in applied efficiency studies is ubiquitous. Further, Parmeter et al. (2014) enforce axioms of production on a translog function and find, for their application, minor differences in estimates of returns to scale between the unrestricted and restricted models. Additionally, the stochastic frontier methodology has been used in a variety of contexts outside of the production environment where axioms of production do not necessarily carry over and so focusing on these axioms limits the scope of what our results here can be used for. Lastly, as a robustness check, we also consider the assumption of a Cobb-Douglas production function, the results of which are included in Appendix C.

To allow for various sample sizes, we first estimate the translog production function based on the actual data and set these estimates as the true parameter values. We then take smooth samples from the four main inputs following the approach of Silverman (1986). We vary λ from $\{0.562, 1.000, 1.778\}$ so that signal to noise is equal to, greater than, and less than 1. We set $\sigma = 1$ across all scenarios given the invariance of the methods as noted in Olson et al. (1980) (i.e. doubling σ should lead to a quadrupling of the mean square errors). For the noise term, we assume a normal distribution. As the assumption about the inefficiency distribution is of special interest, we analyze two cases. First we focus on the performance of the estimators when the distributional assumption is correctly specified. Specifically, inefficiency is generated from a half normal distribution and we assume that it stems from a half normal distribution for the estimators. Alternatively, we assess the three methods when inefficiency is generated from an exponential distribution but we still assume that it stems from a half normal distribution. Each case considers 12 (3 values for λ by 4 sample sizes) scenarios for a total of 24 scenarios. Each scenario is replicated $R = 10,000$ times.

Due to the fact that productivity and efficiency analysis is generally applied to estimate returns to scale (RTS) (also referred to as scale elasticity), predict expected firm output (the estimation of the conditional mean of output) or measure individual technical efficiency ($\widehat{TE}_i$), we evaluate the performance of the methods based on these measures.[5] Furthermore,

[4]See Coelli et al. (2005, Appendix 2) for a more detailed description of the data.

[5]The last two measures are very common in Monte Carlo simulations that compare frontier estimation models, see, for instance, Andor & Hesse (2014) and Kuosmanen & Kortelainen (2012). We choose to

the methods are constructed so that the mean square error is minimized. Accordingly, for each measure, our main performance criterion is the mean square error (MSE) between the estimated and true value of the object of interest:

$$MSE_r = \frac{1}{n}\sum_{i=1}^{n}(\widehat{M}_{i,r} - M_{i,r})^2,$$

where $M_{i,r}$ is the true value of our measure for the ith firm in the rth simulation (M being RTS, expected output, or technical efficiency) and $\widehat{M}_{i,r}$ the estimated value. Given that each simulation will provide a MSE, to easily distill this information we present the median MSE, defined as $\underset{r=1,\ldots,R}{\text{median}}\, MSE_r$, across the $R = 10,000$ simulations. The use of the median MSE is to avoid a small percentage of poor simulations driving the results for a given scenario (by producing large MSEs), which would appear if we were to instead use the mean MSE.

In order to gain additional insights on the performance of the three estimators, we also calculate the mean absolute deviation (MAD) for each of the R simulations:

$$MAD_r = \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{M}_{i,r} - M_{i,r}\right|,$$

and the mean deviation (MD)

$$MD_r = \frac{1}{n}\sum_{i=1}^{n}\widehat{M}_{i,r} - M_{i,r}.$$

Again, we present the median of MAD and MD, defined as $\underset{r=1,\ldots,R}{\text{median}}\, MAD_r$, and $\underset{r=1,\ldots,R}{\text{median}}\, MD_r$, respectively, across the 10,000 simulations in Appendices A and B. Lastly, we also present the mean MSE ($R^{-1}\sum_{r=1}^{R} MSE_r$) in the same appendices as a comparison with the median MSE to determine if any type of tail behavior across the 10,000 simulations that we perform could be impacting what we learn about the three estimators. The Monte Carlo experiments are conducted in `R` (version 3.2.3) and all code is available upon request.

3.2. **Correctly Specified Distribution.** Table 1 shows the results for the setting where the distribution of u_i is correctly specified. The results show in general that each method is superior in a particular scenario. PL and MoM perform relatively better than ML when the sample size is small and/or when λ is low. With increasing λ and increasing sample size

consider additionally the estimation of RTS because many studies investigate scale elasticity to focus on optimal firm size, in particular in the banking sector, e.g. Henderson et al. (2015).

the relative performance of ML improves. However, consistent with Olson et al. (1980), even when ML performs better the gains are not substantial. For the comparison between PL and MoM, the results suggest that MoM has performance advantages when both the sample size and λ are small.

Regarding the estimation of RTS, the results show that PL and MoM (note that these estimates are the first step OLS estimates) estimate RTS more accurately when the sample size is relatively small (i.e. n=100). For all other scenarios, the estimates of RTS across all three methods are nearly identical. For the estimation of predicted output, across all sample sizes, the MoM estimator is the most accurate when λ is small. When λ is 1.778, MoM is the worst method, regardless of sample size. PL is the best method when the sample size is small and is as good as ML when the sample size is large (i.e. n=800). With respect to the estimation of individual efficiency, MoM is the best method, except for one scenario. However, there is only one scenario (n=100 and λ=0.562) where any noticeable difference between MoM and PL arises. For larger values of λ or n the results for ML are also indistinguishable.

Table 1. Performance of PL, ML and MoM for the correctly specified distribution, 10,000 Simulations.

		Returns to Scale (MMSE)		Production Value (MMSE)			Efficiency (MMSE)		
λ	Sample Size	PL/MoM	ML	PL	ML	MoM	PL	ML	MoM
0.562	$n = 100$	0.093	0.098	0.235	0.268	0.208	0.066	0.083	0.054
0.562	$n = 200$	0.041	0.041	0.165	0.173	0.141	0.044	0.050	0.041
0.562	$n = 400$	0.019	0.019	0.120	0.119	0.102	0.028	0.028	0.026
0.562	$n = 800$	0.009	0.009	0.067	0.068	0.065	0.016	0.017	0.016
1.000	$n = 100$	0.075	0.081	0.171	0.252	0.166	0.017	0.042	0.016
1.000	$n = 200$	0.033	0.033	0.083	0.096	0.081	0.009	0.012	0.008
1.000	$n = 400$	0.015	0.015	0.041	0.043	0.041	0.004	0.005	0.004
1.000	$n = 800$	0.008	0.008	0.021	0.021	0.021	0.002	0.002	0.002
1.778	$n = 100$	0.055	0.061	0.101	0.122	0.105	0.009	0.016	0.010
1.778	$n = 200$	0.025	0.024	0.048	0.047	0.050	0.004	0.004	0.004
1.778	$n = 400$	0.012	0.011	0.023	0.022	0.024	0.002	0.002	0.002
1.778	$n = 800$	0.006	0.005	0.011	0.011	0.012	0.001	0.001	0.001

MMSE: median of the mean square error between the estimated and the true value over all replications. The lowest MMSE for each scenario and for each performance criterion are indicated in grey.

The additional performance criteria support these results (see Appendix A). Furthermore, the results suggest that all methods overestimate RTS, the production value and individual efficiency. In the correctly specified case, ML has the lowest overestimation of the production value and individual efficiency on average in all scenarios. Lastly, assuming a Cobb-Douglas production function in the data generating process has no impact on the main findings (Appendix C) and thus we conclude that the results are robust to standard variations of the functional form of the production function.

3.3. **Misspecified Distribution.** As noted earlier, we expect the PL and MoM first stage estimators to be less affected by misspecified distributional assumptions pertaining to inefficiency than ML. To assess this presumption, we conducted the same scenarios as above but used an incorrect distributional assumption for the inefficiency term to estimate the model. Table 2 shows the results for the same 12 scenarios discussed above but by generating firm level inefficiency from an exponential distribution.

Table 2. Performance of PL, ML and MoM for the misspecified distribution, 10,000 Simulations.

		Returns to Scale (MMSE)		Production Value (MMSE)			Efficiency (MMSE)		
λ	Sample Size	PL/MoM	ML	PL	ML	MoM	PL	ML	MoM
0.562	$n = 100$	0.110	0.121	0.271	0.345	0.260	0.051	0.081	0.051
0.562	$n = 200$	0.048	0.048	0.177	0.210	0.179	0.040	0.050	0.043
0.562	$n = 400$	0.023	0.022	0.120	0.134	0.131	0.035	0.039	0.037
0.562	$n = 800$	0.011	0.011	0.099	0.105	0.112	0.033	0.035	0.037
1.000	$n = 100$	0.108	0.121	0.204	0.311	0.233	0.021	0.042	0.026
1.000	$n = 200$	0.049	0.045	0.143	0.176	0.189	0.022	0.029	0.031
1.000	$n = 400$	0.023	0.020	0.124	0.138	0.195	0.023	0.027	0.036
1.000	$n = 800$	0.011	0.010	0.118	0.123	0.207	0.024	0.026	0.040
1.778	$n = 100$	0.107	0.109	0.173	0.235	0.236	0.015	0.035	0.031
1.778	$n = 200$	0.048	0.038	0.124	0.140	0.248	0.015	0.019	0.051
1.778	$n = 400$	0.023	0.016	0.103	0.103	0.282	0.014	0.016	0.056
1.778	$n = 800$	0.011	0.007	0.094	0.090	0.305	0.014	0.014	0.057

MMSE: median of the mean square error between the estimated and the true value over all replications. The lowest MMSE for each scenario and for each performance criterion are indicated in grey.

Several key differences with the correctly specified setting emerge. Regarding the estimation of RTS, ML is the superior method except for scenarios with small sample sizes (i.e. n=100). These results are somewhat surprising as one might expect that the two step estimators would estimate RTS more accurately (due to the fact that the estimated shape of the production function is unbiased). We have two explanations which might explain this result. First, RTS does not exactly directly link to the estimation of a single β, but represents an index of the entire β vector. Second, it could be the case that even when the econometrician misspecifies the convoluted error term's distribution, that it is 'close' to the convoluted error term's true distribution, i.e. while the exponential distribution looks different than the half normal, the convoluted densities look nearly identical. Even with this unexpected performance of the ML estimator in the misspecified distribution setting, the remaining objects of interest (predicted output and efficiency), show that the PL estimator is almost always the dominant method. In addition, in contrast to the correctly specified distribution case, the performance differences are more distinct.

Again, the additional performance criteria support these results (see Appendix B). Almost uniformly we see the same performance as with median MSE. However, in contrast to the correctly specified distribution case, all methods underestimate the production value as well as individual efficiency and PL show the lowest underestimation on average in almost all scenarios.

No Monte Carlo investigation can completely cover all realistic settings and it could be that ML outperforms PL when λ is higher than 1.778 or for an entirely different set of parameter values or production frontier structures. However, the range of λ from 0.562 to 1.778 covers estimates from many practical settings, and the translog functional form is a common modeling approach. Hence, we believe our results suggest that the PL estimator holds practical value for applied researchers.

4. Conclusions

In this paper we investigated the PL estimator's ability to estimate parameters from the stochastic frontier model. The PL approach decouples estimation of the production frontier and the parameters of the error components. A commonly held notion is that under distributional misspecification this decoupling can provide a consistent estimator of the production structure. Using a Monte Carlo investigation centered around a publicly available dataset we compared the performance of ML and PL under correct specification of the distribution of the inefficiency term, as well as the more practically relevant setting (because it is unknown in reality) where this distribution is misspecified. For measures of returns to scale, expected output and firm level technical efficiency, the PL estimator is seen as holding its ground, or even outperforming ML, across all the scenarios considered.

A limitation of both MoM and PL in current practice is the inability to include so-called $\boldsymbol{z}$-variables, also referred to as determinants of inefficiency or inefficiency explanatory variables (see, for example, Reifschneider & Stevenson (1991)). Such $\boldsymbol{z}$-variables are any non-standard inputs or environmental variables that are likely to influence production or cost, but are not direct inputs into the production process. This includes, for instance, a variable delineating if a bank is publicly or privately owned, the age of a farmer, or the gender of a plant manager. Therefore, $\boldsymbol{z}$-variables can influence the parameters of the inefficiency distribution. As the literature has shown, two-step procedures to model the determinants of inefficiency are generally biased (see, among others, Caudill & Ford (1993), Wang & Schmidt (2002) and the discussion in Parmeter & Kumbhakar (2014)). Hence, given the optimistic performance of PL, a potential future avenue for research would include development of approaches enabling researchers and practitioners to consider $\boldsymbol{z}$-variables using PL estimation. A useful starting point in this endeavor is Kuosmanen et al. (2015, sect. 7.8).

References

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

Andor, M. & Hesse, F. (2014), 'The stoned age: The departure into a new era of efficiency analysis? a monte carlo comparison of stoned and the oldies (sfa and dea).', *Journal of Productivity Analysis* **41**(1), 85–109.

Badunenko, O., Henderson, D. J. & Kumbhakar, S. C. (2012), 'When, where and how to perform efficiency estimation', *Journal of the Royal Statistical Society, Series A* **175**(Part 4), 863–892.

Battese, G. E. & Coelli, T. J. (1988), 'Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data', *Journal of Econometrics* **38**(3), 387–399.

Behr, A. & Tente, S. (2008), Stochastic frontier analysis by means of maximum likelihood and the method of moments. Discussion Paper Series 2: Banking and financial studies.

Caudill, S. B. & Ford, J. M. (1993), 'Biases in frontier estimation due to heteroscedasticity', *Economics Letters* **41**(1), 17–20.

Coelli, T. J. (1995), 'Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis', *Journal of Productivity Analysis* **6**(4), 247–268.

Coelli, T. J., Rao, D. S. P., O' Donnell, C. J. & Battese, G. E. (2005), *An Introduction to Efficiency and Productivity Analysis*, Springer, Berlin, Heidelberg.

Fan, Y., Li, Q. & Weersink, A. (1996), 'Semiparametric estimation of stochastic production frontier models', *Journal of Business & Economic Statistics* **14**(4), 460–468.

Greene, W. H. (2008), The Econometric Approach to Efficiency Analysis, *in* H. Fried, C. A. K. Lovell & S. Schmidt, eds, 'The Measurement of Productive Efficiency and Productivity Growth', Oxford University Press, New York. pp. 92-250.

Henderson, D. J., Kumbhakar, S. C., Li, Q. & Parmeter, C. F. (2015), 'Smooth coefficient estimation of a seemingly unrelated regression', *Journal of Econometrics* **189**(1), 148–162.

Jondrow, J., Lovell, C. A. K., Materov, I. S. & Schmidt, P. (1982), 'On the estimation of technical inefficiency in the stochastic frontier production function model', *Journal of Econometrics* **19**(3), 233–238.

Kumbhakar, S. C., Lien, G. & Hardaker, J. B. (2014), 'Technical efficiency in competing panel data models: a study of Norwegian grain farming', *Journal of Productivity Analysis* **41**(2), 321–337.

Kumbhakar, S. C. & Lovell, C. A. K. (2003), *Stochastic Frontier Analysis*, Cambridge University Press, Cambridge.

Kumbhakar, S. C., Wang, H.-J. & Horncastle, A. (2015), *A Practitioner's Guide to Stochastic Frontier Analysis Using Stata*, Cambridge University Press.

Kuosmanen, T. & Fosgerau, M. (2009), 'Neoclassical versus frontier production models? testing for the skewness of regression residuals', *The Scandinavian Journal of Economics* **111**(2), 351–367.

Kuosmanen, T., Johnson, A. & Saastamoinen, A. (2015), 'Stochastic nonparametric approach to efficiency analysis: A unified framework', pp. 191–244.

Kuosmanen, T. & Kortelainen, M. (2012), 'Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints', *Journal of Productivity Analysis* pp. 1–18.

Meeusen, W. & van den Broeck, J. (1977), 'Efficiency estimation from Cobb-Douglas production functions with composed error', *International Economic Review* **18**(2), 435–444.

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

Parmeter, C. F. & Kumbhakar, S. C. (2014), 'Efficiency Analysis: A Primer on Recent Advances.', *Foundations and Trends in Econometrics* **7(3-4)**(1), 191–385.

Parmeter, C. F., Sun, K., Henderson, D. J. & Kumbhakar, S. C. (2014), 'Estimation and inference under economic restrictions', *Journal of productivity analysis* **41**(1), 111–129.

Reifschneider, D. & Stevenson, R. (1991), 'Systematic departures from the frontier: a framework for the analysis of firm inefficiency', *International Economic Review* pp. 715–723.

Rho, S. & Schmidt, P. (2015), 'Are all firms inefficient?', *Journal of Productivity Analysis* **43**(3), 327–349.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Vol. 26, CRC press.

Wang, H.-J. & Schmidt, P. (2002), 'One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels', *journal of Productivity Analysis* **18**(2), 129–144.

Appendix A. Alternative Performance Metrics With the Correctly Specified Distribution

Tables A1 - A3 compare ML, PL and MoM using different performance metrics, MAD, MMD and mean MSE, respectively. Regardless of the performance metric, the same overall insights as described in the body of the paper hold true. For λ small, MoM outperforms ML/PL and as λ or n increases, these advantages disappear.

Table A1. Performance of PL, ML and MoM for the correctly specified distribution, 10,000 Simulations, median of the mean absolute deviation (MAD).

		Returns to Scale (MAD)		Production Value (MAD)			Efficiency (MAD)		
λ	Sample Size	PL/MoM	ML	PL	ML	MoM	PL	ML	MoM
0.562	$n = 100$	0.247	0.255	0.402	0.434	0.371	0.221	0.278	0.211
0.562	$n = 200$	0.164	0.164	0.351	0.361	0.318	0.183	0.193	0.176
0.562	$n = 400$	0.112	0.112	0.312	0.312	0.283	0.148	0.152	0.144
0.562	$n = 800$	0.078	0.078	0.234	0.235	0.230	0.115	0.118	0.113
1.000	$n = 100$	0.222	0.231	0.332	0.431	0.324	0.109	0.173	0.105
1.000	$n = 200$	0.147	0.148	0.230	0.253	0.226	0.077	0.090	0.074
1.000	$n = 400$	0.101	0.101	0.161	0.166	0.160	0.054	0.058	0.054
1.000	$n = 800$	0.071	0.071	0.113	0.115	0.113	0.039	0.040	0.039
1.778	$n = 100$	0.192	0.202	0.246	0.279	0.251	0.075	0.105	0.080
1.778	$n = 200$	0.128	0.127	0.166	0.167	0.171	0.049	0.053	0.052
1.778	$n = 400$	0.088	0.085	0.116	0.114	0.119	0.034	0.034	0.036
1.778	$n = 800$	0.061	0.059	0.081	0.079	0.083	0.023	0.023	0.025

MAD: median of the mean absolute deviation between the estimated and the true value over all replications. The lowest median MAD for each scenario and for each performance criterion are indicated in grey.

The results for MMD (Table A2) show that all methods overestimate RTS, the production value and individual efficiency in the correctly specified distribution case. Thereby, ML has the lowest overestimation of the production value and individual efficiency on average in all scenarios.

The results in Table A3 also suggest that few, if any, simulations produces poor estimates, leading to exorbitant differences between the mean and the median. Overall, both measures generate almost identical insights across MSE and the three different objects estimated with the estimates, RTS, predicted output and technical efficiency.

Table A2. Performance of PL, ML and MoM for the correctly specified distribution, 10,000 Simulations, median of the mean deviation (MMD).

		Returns to Scale (MMD)		Production Value (MMD)			Efficiency (MMD)		
λ	Sample Size	PL/MoM	ML	PL	ML	MoM	PL	ML	MoM
0.562	$n = 100$	0.002	0.003	0.130	0.072	0.098	0.070	0.039	0.049
0.562	$n = 200$	0.000	0.002	0.057	0.029	0.039	0.025	0.008	0.015
0.562	$n = 400$	0.000	0.000	0.030	0.017	0.024	0.016	0.008	0.009
0.562	$n = 800$	0.000	0.000	0.016	0.009	0.012	0.007	0.003	0.004
1.000	$n = 100$	0.001	0.002	0.116	0.019	0.118	0.051	0.003	0.052
1.000	$n = 200$	0.000	0.001	0.056	0.013	0.060	0.024	0.004	0.026
1.000	$n = 400$	0.000	0.000	0.027	0.006	0.031	0.012	0.002	0.012
1.000	$n = 800$	0.000	0.000	0.014	0.003	0.014	0.006	0.001	0.006
1.778	$n = 100$	−0.001	−0.002	0.114	−0.018	0.128	0.046	−0.006	0.053
1.778	$n = 200$	−0.001	0.000	0.054	0.000	0.064	0.021	0.000	0.026
1.778	$n = 400$	0.000	0.000	0.027	0.001	0.032	0.011	0.000	0.013
1.778	$n = 800$	0.000	0.000	0.012	0.000	0.015	0.005	0.000	0.006

MMD: median of the mean deviation between the estimated and the true value over all replications. The MMD, which is closest to zero, for each scenario and for each performance criterion are indicated in grey.

Table A3. Performance of PL, ML and MoM for the correctly specified distribution, 10,000 Simulations, mean of the mean square error (MSE).

		Returns to Scale (MSE)		Production Value (MSE)			Efficiency (MSE)		
λ	Sample Size	PL/MoM	ML	PL	ML	MoM	PL	ML	MoM
0.562	$n = 100$	0.115	0.126	0.248	0.331	0.223	0.056	0.068	0.041
0.562	$n = 200$	0.050	0.051	0.168	0.182	0.147	0.050	0.051	0.036
0.562	$n = 400$	0.023	0.024	0.120	0.120	0.102	0.044	0.042	0.032
0.562	$n = 800$	0.011	0.011	0.089	0.085	0.074	0.038	0.035	0.027
1.000	$n = 100$	0.092	0.104	0.235	0.282	0.210	0.063	0.069	0.048
1.000	$n = 200$	0.040	0.041	0.151	0.155	0.134	0.047	0.045	0.036
1.000	$n = 400$	0.019	0.019	0.091	0.089	0.080	0.030	0.028	0.024
1.000	$n = 800$	0.009	0.009	0.047	0.046	0.043	0.015	0.014	0.012
1.778	$n = 100$	0.069	0.081	0.165	0.183	0.157	0.040	0.043	0.034
1.778	$n = 200$	0.031	0.030	0.071	0.070	0.070	0.014	0.014	0.013
1.778	$n = 400$	0.014	0.013	0.029	0.027	0.030	0.004	0.004	0.004
1.778	$n = 800$	0.007	0.007	0.013	0.013	0.014	0.001	0.001	0.002

MSE: mean of the mean square error between the estimated and the true value over all replications. The lowest mean MSE for each scenario and for each performance criterion are indicated in grey.

Appendix B. Alternative Performance Metrics With the Misspecified Distribution

Tables B1 and B2 compare ML, PL and MoM using MAD and MMD, respectively when the true distribution of inefficiency is exponential, but the econometrician erroneously assumes that inefficiency stems from a half normal distribution. Regardless of the performance metric, the same overall insights as described in the body of the paper hold true. PL is the preferred method for estimating output or technical efficiency, regardless of either the sample size or λ. Similarly, ML, for $n > 100$ estimates RTS better than the PL/MoM estimator, across all levels of λ considered. Again, this is surprising on the surface, but not necessarily when one considers that the convoluted densities (either for exponential or half normal) are quite similar.

Table B1. Performance of PL, ML and MoM for the misspecified distribution, 10,000 Simulations, median of the mean absolute deviation (MAD).

		Returns to Scale (MAD)		Production Value (MAD)			Efficiency (MAD)		
λ	Sample Size	PL/MoM	ML	PL	ML	MoM	PL	ML	MoM
0.562	$n = 100$	0.270	0.283	0.430	0.500	0.416	0.196	0.247	0.197
0.562	$n = 200$	0.178	0.178	0.358	0.401	0.362	0.179	0.198	0.183
0.562	$n = 400$	0.122	0.122	0.306	0.328	0.322	0.168	0.176	0.174
0.562	$n = 800$	0.085	0.084	0.290	0.301	0.312	0.165	0.169	0.174
1.000	$n = 100$	0.267	0.283	0.356	0.474	0.386	0.121	0.174	0.138
1.000	$n = 200$	0.179	0.173	0.314	0.365	0.375	0.128	0.148	0.152
1.000	$n = 400$	0.122	0.115	0.312	0.337	0.407	0.135	0.143	0.167
1.000	$n = 800$	0.085	0.080	0.320	0.331	0.437	0.140	0.143	0.177
1.778	$n = 100$	0.266	0.269	0.323	0.400	0.391	0.097	0.160	0.147
1.778	$n = 200$	0.177	0.158	0.288	0.326	0.443	0.100	0.119	0.194
1.778	$n = 400$	0.122	0.102	0.280	0.291	0.501	0.103	0.109	0.208
1.778	$n = 800$	0.085	0.069	0.283	0.282	0.537	0.105	0.106	0.213

MAD: median of the mean absolute deviation between the estimated and the true value over all replications. The lowest median of MAD for each scenario and for each performance criterion are indicated in grey.

In contrast to Appendix A, all methods underestimate the production value and individual efficiency in the misspecified distribution case (Table B2). Thereby, PL has the lowest underestimation of the production value and individual efficiency on average in almost all scenarios.

Similarly to Appendix A, the results in Table B3 suggest that almost none of the simulations produced poor estimates, leading to the mean and the median results being in

Table B2. Performance of PL, ML and MoM for the misspecified distribution, 10,000 Simulations, median of the mean deviation (MMD).

		Returns to Scale (MMD)		Production Value (MMD)			Efficiency (MMD)		
λ	Sample Size	PL/MoM	ML	PL	ML	MoM	PL	ML	MoM
0.562	$n = 100$	−0.001	0.001	−0.160	−0.291	−0.159	−0.122	−0.170	−0.121
0.562	$n = 200$	−0.001	0.000	−0.232	−0.288	−0.236	−0.147	−0.167	−0.149
0.562	$n = 400$	0.000	0.001	−0.258	−0.286	−0.277	−0.158	−0.167	−0.163
0.562	$n = 800$	−0.001	0.000	−0.280	−0.293	−0.303	−0.164	−0.168	−0.172
1.000	$n = 100$	−0.001	0.000	−0.188	−0.360	−0.233	−0.103	−0.149	−0.117
1.000	$n = 200$	0.000	0.001	−0.262	−0.330	−0.337	−0.125	−0.143	−0.147
1.000	$n = 400$	−0.001	0.000	−0.298	−0.326	−0.400	−0.135	−0.142	−0.164
1.000	$n = 800$	0.000	0.000	−0.316	−0.328	−0.436	−0.139	−0.143	−0.174
1.778	$n = 100$	0.000	0.003	−0.158	−0.295	−0.285	−0.073	−0.099	−0.111
1.778	$n = 200$	−0.001	0.000	−0.230	−0.294	−0.420	−0.092	−0.107	−0.142
1.778	$n = 400$	0.000	0.001	−0.261	−0.281	−0.497	−0.099	−0.104	−0.165
1.778	$n = 800$	−0.001	0.000	−0.277	−0.280	−0.536	−0.103	−0.104	−0.178

MMD: median of the mean deviation between the estimated and the true value over all replications. The MMD, which is closest to zero, for each scenario and for each performance criterion are indicated in grey.

near unison. Overall, both measures generate almost identical insights across MSE and the three different objects estimated with the estimates, RTS, predicted output and technical efficiency.

TABLE B3. Performance of PL, ML and MoM for the misspecified distribution, 10,000 Simulations, mean of the mean square error (MSE).

		Returns to Scale (MSE)		Production Value (MSE)			Efficiency (MSE)		
λ	Sample Size	PL/MoM	ML	PL	ML	MoM	PL	ML	MoM
0.562	$n = 100$	0.137	0.157	0.294	0.430	0.288	0.054	0.076	0.047
0.562	$n = 200$	0.059	0.059	0.197	0.232	0.204	0.045	0.051	0.043
0.562	$n = 400$	0.028	0.028	0.137	0.150	0.152	0.037	0.040	0.039
0.562	$n = 800$	0.013	0.013	0.107	0.114	0.127	0.033	0.035	0.037
1.000	$n = 100$	0.136	0.159	0.238	0.372	0.297	0.030	0.052	0.041
1.000	$n = 200$	0.060	0.056	0.159	0.199	0.238	0.023	0.031	0.040
1.000	$n = 400$	0.027	0.025	0.130	0.144	0.225	0.023	0.026	0.044
1.000	$n = 800$	0.013	0.012	0.121	0.126	0.225	0.024	0.026	0.046
1.778	$n = 100$	0.137	0.144	0.188	0.251	0.313	0.017	0.033	0.041
1.778	$n = 200$	0.059	0.050	0.131	0.164	0.310	0.015	0.022	0.048
1.778	$n = 400$	0.028	0.019	0.107	0.108	0.321	0.015	0.016	0.054
1.778	$n = 800$	0.013	0.009	0.096	0.092	0.328	0.014	0.015	0.057

MSE: mean of the mean square error between the estimated and the true value over all replications. The lowest mean of the MSE for each scenario and for each performance criterion are indicated in grey.

Appendix C. Cobb-Douglas Production Function

When we use a Cobb-Douglas production function, as opposed to translog, we see almost similar performance between the three estimators. The estimator of RTS is identical between PL/MoM and ML regardless of the sample size and the value of λ. In the Cobb-Douglas setting RTS is constant across all firms and is only dependent upon the four βs for the inputs. In this case it would make sense that all three methods are essentially identical as the estimators of β, excluding the intercept, are all consistent. When it comes to prediction, MoM dominates for small values of λ, but PL and ML perform at least as well as MOM for $\lambda = 1.778$ across all sample sizes; notice that while there are performance gains, they are quite small. Finally, for the estimation of firm level inefficiency, all three of the methods are virtually identical across sample sizes for $\lambda = 1.778$, with only minor performance differences arising. However, for $\lambda = 0.562$, MoM, is the dominant estimator of firm level inefficiency with the relative performance gains, as expected, decreasing as the sample size increased. In sum, the use of a different production function did not add much additional insight into which of the three different estimators was preferred.

Table C1. Performance of PL, ML and MoM for a Cobb-Douglas Production Frontier, 10,000 Simulations

		Returns to Scale (MMSE)		Production Value (MMSE)			Efficiency (MMSE)		
λ	Sample Size	PL/MoM	ML	PL	ML	MoM	PL	ML	MoM
0.562	$n = 100$	0.006	0.006	0.159	0.165	0.127	0.073	0.079	0.055
0.562	$n = 200$	0.003	0.003	0.126	0.127	0.101	0.044	0.047	0.041
0.562	$n = 400$	0.001	0.001	0.104	0.103	0.084	0.029	0.030	0.028
0.562	$n = 800$	0.001	0.001	0.055	0.056	0.054	0.016	0.017	0.016
1.000	$n = 100$	0.005	0.005	0.096	0.111	0.084	0.015	0.019	0.013
1.000	$n = 200$	0.002	0.002	0.047	0.049	0.044	0.008	0.009	0.007
1.000	$n = 400$	0.001	0.001	0.023	0.023	0.022	0.004	0.004	0.004
1.000	$n = 800$	0.001	0.001	0.011	0.011	0.011	0.002	0.002	0.002
1.778	$n = 100$	0.004	0.004	0.039	0.041	0.042	0.005	0.005	0.005
1.778	$n = 200$	0.002	0.002	0.019	0.019	0.020	0.002	0.002	0.003
1.778	$n = 400$	0.001	0.001	0.009	0.009	0.010	0.001	0.001	0.001
1.778	$n = 800$	0.000	0.000	0.005	0.004	0.005	0.001	0.001	0.001

MMSE: median of the mean square error between the estimated and the true value over all replications. The lowest median of the MSE for each scenario and for each performance criterion are indicated in grey.