Robust Stochastic Frontier Analysis: a Minimum Density Power Divergence Approach

Federico Belotti*
Giuseppe Ilardi°

* CEIS, University of Rome Tor Vergata
° Bank of Italy

Workshop on the Econometrics and Statistics of Efficiency Analysis: Recent Developments and Perspectives
Lecce, June 21, 2015
Motivation

- Maximum Likelihood (ML) remains the most widely used approach to estimate SF models **parametrically**.

- Computational simplicity and asymptotic efficiency but **very poor robustness properties**.

- Only **Nonparametric literature** on both deterministic and stochastic frontier models estimation deals with the **robustness issue** (See Wilson, 1993; Cazals et al., 2002; Florens and Simar, 2014, among others).
Fragile behavior of ML in presence of extreme values
We propose to exploit a density-based minimum distance approach, the so-called Minimum Density Power Divergence (MDPD, Basu et al., 1998) to “robustify” the parametric estimation of Stochastic Frontier (SF) models.

We investigate the small samples properties of MDPD focusing on the normal-half normal SF model.

We also study the finite sample properties of a simple Hausman-like test for the presence of outliers.
The Density Power Divergences

Consider a parametric family of models $\mathcal{F}_\theta$, indexed by the unknown parameter $\theta \in \Theta \subset \mathbb{R}^k$, possessing densities $f_\theta$; let $\mathcal{G}$ be the class of all distributions $G$ having densities $g$.

The family of divergences by Basu et al. (1998), as a function of $\alpha$, is

\[
d_\alpha(g, f) = \int \left[ f^{1+\alpha}_\theta(y) - (1 + \frac{1}{\alpha})g(y)f^\alpha_\theta(y) + \frac{1}{\alpha}g^{1+\alpha}(y) \right] dy
\]  

When $\alpha = 0$, the integral in (1) is undefined

\[
d_0(g, f) = \lim_{\alpha \to 0} d_\alpha(g, f) = \int g(y) \log \left[ \frac{g(y)}{f(y)} \right] dy
\]  

is the Kullback-Leibler (KL) divergence.
Link between ML and MDPD estimators

- When $\alpha = 0$, define $T_0(G)$ the minimum KL divergence functional at $G$ and assume that it exists and is unique.

- Then, given a random sample $y_1, \ldots, y_N$ from $G$, $T_0(G_N)$ maximizes $\int \log f_\theta(y) dG_N(y)$, and is therefore the ML estimate of $\theta$, where $G_N$ is the empirical distribution function.

- In the same way, define $T_\alpha(G)$ the minimum density power divergence functional at $G$. Then, given the data, for each $\alpha$

$$
\hat{\theta}_{MDPD} = \arg \min_{\theta \in \Theta} \int f_\theta^{1+\alpha}(y) dy - (1 - \frac{1}{\alpha}) n^{-1} \sum_{i=1}^{N} f_\theta^{\alpha}(y_i) \quad (3)
$$

is the MDPD estimator of $\theta$. 
**MDPD robusteness** \((\alpha = 1)\)
Properties of the MDPD estimators

- The parameter $\alpha$ controls the trade-off between robustness and efficiency.

- Asymptotic properties follows immediately from existing theory since the MDPD family of estimators are M-estimators (Huber, 1981), i.e. solve an equation of the form $\sum_{i=1}^{N} \psi(y_i, \theta)$ with

$$\psi(y_i, \theta) = u_\theta(y)f_\theta^\alpha(y) - \int u_\theta(z)f_\theta^{1+\alpha}(y)dy$$

and where $u_\theta(y) = \partial \log f_\theta(y) / \partial \theta$ is the maximum likelihood score function.

- Under mild regularity conditions, there exists $\hat{\theta}_{MDPD}$ such that, as $N \to \infty$
  1. $\hat{\theta}_{MDPD}$ is consistent for $\theta$, and
  2. $N^{\frac{1}{2}}(\hat{\theta}_{MDPD} - \theta) \sim \mathcal{N}(0, A^{-1}BA^{-1})$. 
Why the Density Power Divergences approach?

- Other divergences, like the Hellinger divergence (Beran, 1977), show strong robustness retaining first-order efficiency, but they force to use some form of non-parametric smoother to produce an estimate of the true density $g$.

- Among the huge class of density-based minimum distance methods, it has been shown that MDPD has strong robustness properties with a negligible loss in terms of asymptotic efficiency relative to ML (Basu et al., 1998; Lee and Sriram, 2013; Lee and Song, 2013; Kang and Lee, 2014).

- Can be easily applied to a wide range of SF models.
The SF model

- We consider the Aigner et al. (1977) model

\[ y_i = x_i \beta + v_i - u_i \]  

where \( v_i \sim \mathcal{N}(0, \psi^2) \) and \( u_i \sim \mathcal{N}^+(0, \sigma^2) \). Thus, \( \varepsilon = v_i - u_i \sim SN(0, \sigma, \lambda) \), with \( \lambda = \sigma / \psi \) and \( \theta = (\beta, \sigma, \lambda) \).

- In this case the MDPD estimator is easily obtained by

\[ \hat{\theta}_{MDPD} = \arg \min_{\theta \in \Theta} \int f(\varepsilon)^{1+\alpha} d\varepsilon - (1 + \frac{1}{\alpha}) n^{-1} \sum_{i=1}^{N} f(\varepsilon_i)^{\alpha} \]  

where the first integral has been numerically approximated using Gauss-Hermite quadrature.
Monte Carlo Design

- $\beta = 0.5, \sigma = 0.5, \psi = 0.25$

- $\phi = 0, 0.01, 0.02, 0.03, 0.04, 0.05$

- Measurement error in $y$

$$y_i = x_i \beta + v_i - u_i$$
$$v_i = \phi v_{2i} + (1 - \phi)v_{1i}$$
$$v_{1i} \sim \mathcal{N}(0, \psi)$$
$$v_{2i} \sim \mathcal{N}(0, 4\psi)$$

- Heterogeneous technology

$$y_i = \phi y_{2i} + (1 - \phi)y_{1i}$$
$$y_{1i} = x_i \beta + v_i - u_i$$
$$y_{2i} = x_i (4\beta) + v_i - u_i$$
Results: \( \phi = 0 \)

| \( \alpha \) | \( \beta \) | \( \sigma \) | \( \psi \) | \( E(u|\varepsilon) \) |
|---|---|---|---|---|
| \( \alpha = 0 \) | 0.0022 | 0.0011 | 0.0005 | 0.0352 |
| \( \alpha = 0.25 \) | 0.0022 | 0.0011 | 0.0006 | 0.0353 |
| \( \alpha = 0.5 \) | 0.0025 | 0.0013 | 0.0007 | 0.0355 |
| \( \alpha = 1 \) | 0.0031 | 0.0019 | 0.0010 | 0.0362 |

| \( \alpha \) | \( \beta \) | \( \sigma \) | \( \psi \) | \( E(u|\varepsilon) \) |
|---|---|---|---|---|
| \( \alpha = 0 \) | 0.0010 | 0.0005 | 0.0002 | 0.0347 |
| \( \alpha = 0.25 \) | 0.0010 | 0.0005 | 0.0002 | 0.0347 |
| \( \alpha = 0.5 \) | 0.0011 | 0.0007 | 0.0003 | 0.0348 |
| \( \alpha = 1 \) | 0.0015 | 0.0011 | 0.0005 | 0.0352 |
Results: measurement error in $y - \beta$
Results: measurement error in $y - \sigma$
Results: measurement error in $y - \psi$
Results: measurement error in $y - \hat{y}$
Results: heterogeneous technology - $\beta$
Results: heterogeneous technology - $\sigma$
Results: heterogeneous technology - $\psi$
Results: heterogeneous technology - $\hat{u}$
Testing for the presence of outliers

- Denote with $\hat{\theta}_1$ the ML estimator and with $\hat{\theta}_2$ the MDPD estimator with $\alpha = 1$.

- Because of asymptotic normality of our estimators, an asymptotic test may be based on the test statistic
  \[ \xi = \hat{\Delta}'\hat{V}\hat{\Delta}, \]
  where $\hat{\Delta} = \hat{\theta}_1 - \hat{\theta}_2$, $\hat{V} = D'\hat{W}D$, $D = [I_k, -I_k]$, and
  \[ \hat{W} = H^{-1}SH^{-1} = \begin{bmatrix} H_1^{-1} & 0 \\ 0 & H_2^{-1} \end{bmatrix} \begin{bmatrix} S_1'S_1 & S_1'S_2 \\ S_2'S_1 & S_2'S_2 \end{bmatrix} \begin{bmatrix} H_1^{-1} & 0 \\ 0 & H_2^{-1} \end{bmatrix}. \]

- Under $H_0$
  \[ n\xi \xrightarrow{d} \chi^2_k, \]
  where $k = \text{rank}(V)$.
Small samples properties of the test
Concluding remarks and future directions

- MDPD appears to be quite promising in robustifying the SF model.
- This approach may be used to test for the presence of outliers.
- 1) Panel data extension; 2) Use MDPD in a semiparametric SF model.
References I


