Comparing Bayesian Models for Production Efficiency

Submited version available in my webpage

Ricardo Sandes Ehlers

Stats & Geoinformation Lab

Universidade Federal do Paraná

http://leg.ufpr.br/∼ehlers

ehlers@leg.ufpr.br

Stochastic Production Frontier Models

In stochastic production frontier models it is usually assumed that the error term is composed of a random error (v) capturing statistical noise and a one-sided non-negative error (u). For cross-sectional data on N observed economic agents (generically denoted by "firms"), the model can be expressed as

$$y_i = f(\boldsymbol{x}_i, \boldsymbol{\beta}) + v_i - u_i, \; i = 1, \ldots, N \tag{1}$$

where

y_i is the logarithm of an output,

$\boldsymbol{x}_i$ is a vector of the logarithms of inputs including an intercept and possibly crossproducts,

$\boldsymbol{\beta}$ is the vector of coefficients and

v_i are independent and identically distributed $N(0, \sigma_v^2)$ error terms

u_i, which measures technical inefficiency of the ith firm.

v_i and u_i are assumed to be independent.

Production Functions

For a product Q produced with 2 inputs: capital (K) and Labour (L) let $y = \log(Q)$, $x_1 = \log(K)$ and $x_2 = \log(L)$. Production functions used here:

- Cobb-Douglas: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$,
- Translog $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \frac{1}{2}\beta_3 x_1^2 + \frac{1}{2}\beta_4 x_2^2 + \beta_5 x_1 x_2$,
- Constant Elasticity of Substitution (CES) $y = \beta_0 + v \log(\,[(1-\delta)K^{-\rho} + \delta L^{-\rho}]^{-1/\rho})$, where $0 < \delta < 1$, $-\infty < v < \infty$ and $\rho > -1$,
- Generalized Production Function (GPF) $z = y + \lambda \exp(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$,

The MCMC sampling is done using the package `JAGS` (Just Another Gibbs Sampler, Plummer 2003) which was originally developed as a clone of the classic `BUGS` package. JAGS was written in `C++` and is designed to work closely with the `R` package where all statistical computations and graphics are done. It is open source and freely available and can be downloaded from the website

`http://www-fis.iarc.fr/~ martyn/software/jags/`.

Bayesian Models

For Cobb-Douglas, Translog and CES production functions the likelihood is

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} p_N(y_i|f(x_{i1}, x_{i2}, \boldsymbol{\theta}) - u_i, \sigma^2).$$

For the GPF the likelihood function is given by

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} p_N(z_i|f(x_{i1}, x_{i2}, \boldsymbol{\theta}) - u_i, \sigma^2) \left|\frac{dz_i}{dy_i}\right|. \quad (2)$$

The parameters in $\boldsymbol{\theta}$ are all assumed to be *a priori* independent.

- $\beta_0 \sim N(0, \sigma_\beta^2)$ constrained to be positive as we want to exclude production frontiers in which more inputs lead to less output.
- $\beta_j \sim N(0, \sigma_\beta^2)$ truncated to $\beta_j > 0$, as we want to exclude production frontiers in which more inputs lead to less output
- For a CES production function we assign $v \sim N(0, \sigma_v^2)$, $\rho \sim N(0, \sigma_\rho^2)$ truncated to $\rho > -1$ and $\delta \sim Beta(a, b)$ and $\sigma^{-2} \sim Gamma(a, b)$.

Priors for u_i

1. $u_i \sim Exp(\lambda)$ and $\lambda \sim Exp(-\log r^*)$ implies prior median efficiency equals r^* (van den Broeck et al. 1994).

2. $u_i \sim N(0, \lambda)$ truncated to $u_i > 0$ and $\lambda^{-1} \sim Ga(1, 1/37.5) \Rightarrow$ median$(\exp(-u_i)) \approx 0.875$ with a reasonable spread (van den Broeck et al. 1994).

3. $u_i \sim N(\xi, \lambda)$ truncated to $u_i > 0$. Include a prior $\xi \sim N(0, \sigma_\xi^2)$ and use the same prior for λ as above.

4. $u_i \sim Ga(\phi, \lambda)$, Griffin and Steel (2004) propose a prior on ϕ and λ which extends the informative prior for an exponential inefficiency distribution. $\phi \sim Ga(d_1, d_1 + 1)$ implies that mode(ϕ)=1 (centred around an exponential prior with d_1 controlling the variability). Also $\lambda|\phi \sim Ga(\phi, -\log r^*)$ where r^*=median($\exp(-u_i)$).

5. $u_i \sim LN(0, \psi^2)$ and $\psi^{-2} \sim Ga(a, b)$.

6. $u_i \sim$ Generalized Gamma (c, ϕ, λ) where c, ϕ and λ are specified following the development in Griffin and Steel (2004). We take $\lambda | c, \phi \sim Ga(\phi, (-\log r^*)^c)$ and assume prior independence between c and $\psi = \phi c$ assigning priors

$$\psi^{-1} \sim Ga(d_1, d_1 + 1) \quad \text{and} \quad c^{-1} \sim Ga(d_2, d_2 + 1).$$

   This is again centred over the exponential case (mode at $\phi = c = 1$) but allows considerable deviations from the exponential if we choose d_1 and d_2 not too large. We take $d_1 = d_2 = 3$ here.

7. For $u_i \sim Weibull(c, \lambda)$ we adapt the development in the Generalized Gamma with $\phi = 1$ and assign $\lambda | c \sim Exp((-\log r^*)^c)$ and $c^{-1} \sim Ga(d_2, d_2 + 1)$ which is centred over the exponential case.

The R code implemented allows the user to specify values for the hyperparameters σ^2_β, σ^2_v, σ^2_ρ, a, b, d_1, d_2 and r^*.
All R functions used are available: `http://leg.ufpr.br/~ ehlers/SPF`.

Comparing Models

Compare and select the most appropriate model using the *Deviance Information Criterion* (DIC), where lower values indicate a good model fit relative to the number of parameters in the model (Spiegelhalter et al. 2002).

Denote the competing models by $M_1, M_2, \ldots$ and the vector of parameters under model M_i by ξ_i, then

$$DIC(M_i) = \overline{D}_i + p_i$$

$D_i = -2\log(p(\boldsymbol{y} \mid \xi_i, M_i))$ is the deviance, measuring model fit,

$\overline{D}_i = E(D_i|\boldsymbol{y})$.

$p_i = \overline{D}_i - D(\bar{\xi}_i)$ measures model complexity,

$\bar{\xi}_i = E(\xi_i|\boldsymbol{y})$.

Computing Bayes factors from a MCMC output is not a trivial task and in particular is not easily implemented in an all purpose package like `JAGS` (the same is true for the `WinBUGS package`).
In this work we include an `R` function for the computation of DIC.

DIC Weights

It could be misleading just to report the model with the lowest DIC.

It is difficult to say what would constitute an important difference between DIC values.

DIC is subject to Monte Carlo error since it is a function of simulated quantities. This might cast some doubt whether an improvement in model fit is substantial.

We propose to use DIC weights obtained by subtracting from each DIC the value associated with the "best" model and setting

$$w_i \propto \exp(-\Delta DIC(M_i)/2)$$

where $\Delta DIC(M_i)$ denotes the transformed DIC value for model i. The weights are normalized to sum to 1 over the models under consideration.
This was first suggested in Burnham and Anderson (1998) for the Akaike Information Criterion (the differences are interpreted as the strength of evidence). It can be extended to be used with the DIC which, as pointed out in Spiegelhalter et al. (2002), can be viewed as a Bayesian analogue of AIC.

Example

123 cross-sectional data from the US electric industry in 1970 available at http://econ.queensu.ca/jae/1998-v13.2/zellner-ryu/data.zr.

For the 28 competing models a total of 100 000 iterations were run, and the first half was discarded as burn-in. After burn-in, simulated values of every 5th iteration were kept for posterior analysis.

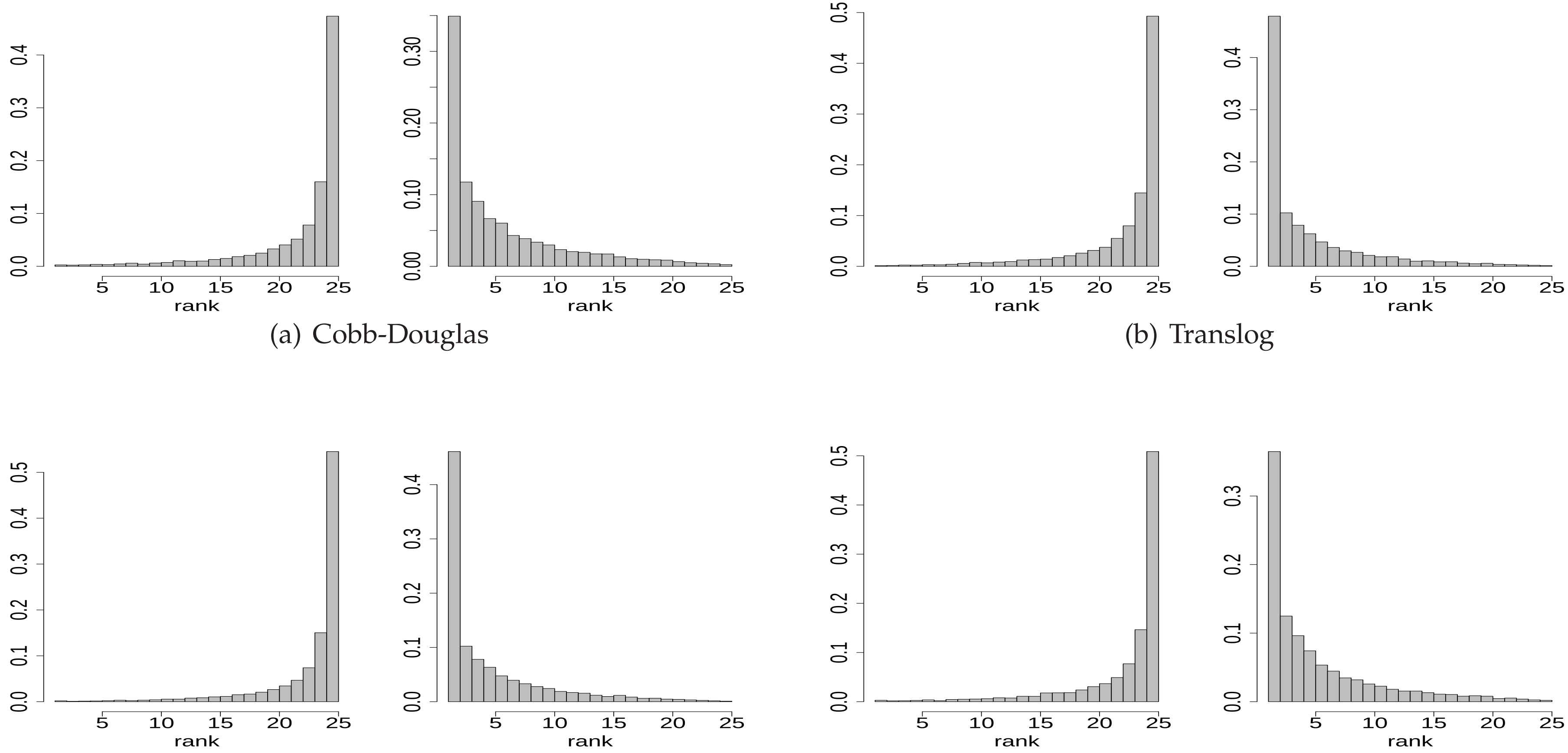
From Table 1 the most adequate model is the one that uses the GPF and a truncated normal distribution for the inefficiency terms. Note that model comparison is much easier in the transformed scale of the DIC weights. The weight for the second best model (CES + truncated normal inefficiencies) is less than half the weight for the most adequate model. Another subset of models (GPF + lognormal, CES + lognormal and Cobb-Douglas + truncated normal) received small weights.

Figure 1 shows the posterior distributions of ranks associated with $\exp(-u_i)$ of the best firm (left column) and the worst firm (right column) using the truncated normal considering all production functions. The truncated normal frontier model differentiates pretty well the more efficient from the less efficient firms.

Table 1: Model comparison based on 10 000 simulations.

Prod Function	inefficiency	$\overline{D}_i$	$D(\bar{\xi}_i)$	p_i	DIC	weight	rank
Cobb-Douglas	exp	-19.5609	-28.3636	8.8027	-10.7581	0.00044	23
Cobb-Douglas	tnorm	-30.9996	-41.6037	10.6041	-20.3956	0.05449	5
Cobb-Douglas	halfnorm	-19.0528	-27.2502	8.1974	-10.8554	0.00046	22
Cobb-Douglas	gamma	-19.8941	-28.6195	8.7254	-11.1687	0.00054	21
Cobb-Douglas	gen.gamma	-20.2863	-28.6803	8.3940	-11.8924	0.00078	19
Cobb-Douglas	weibull	-18.6541	-26.1266	7.4726	-11.1815	0.00054	20
Cobb-Douglas	lognorm	-27.6661	-37.6017	9.9355	-17.7306	0.01437	6
Translog	exp	-15.0361	-23.8715	8.8353	-6.2008	0.00005	25
Translog	tnorm	-23.9300	-33.2167	9.2867	-14.6432	0.00307	7
Translog	halfnorm	-12.8873	-20.6193	7.7320	-5.1554	0.00003	27
Translog	gamma	-15.1713	-23.7525	8.5812	-6.5900	0.00005	24
Translog	gen.gamma	-13.8555	-21.7858	7.9303	-5.9252	0.00004	26
Translog	weibull	-12.8305	-20.5585	7.7280	-5.1025	0.00003	28
Translog	lognorm	-22.0603	-31.4122	9.3519	-12.7084	0.00117	13

Prod Function	inefficiency	$\overline{D}_i$	$D(\bar{\xi}_i)$	p_i	DIC	weight	rank
CES	exp	-19.5195	-26.4384	6.9190	-12.6005	0.00111	15
CES	tnorm	-31.2346	-39.2628	8.0282	-23.2065	0.22216	2
CES	halfnorm	-19.5870	-25.8898	6.3028	-13.2842	0.00156	11
CES	gamma	-20.2690	-26.6339	6.3649	-13.9041	0.00212	8
CES	gen.gamma	-20.1451	-26.8770	6.7319	-13.4132	0.00166	10
CES	weibull	-18.8055	-24.9562	6.1507	-12.6548	0.00114	14
CES	lognorm	-28.1078	-35.6027	7.4949	-20.6130	0.06074	4
GPF	exp	-19.9535	-27.9918	8.0382	-11.9153	0.00078	18
GPF	tnorm	-33.4537	-41.8647	8.4110	-25.0428	0.55643	1
GPF	halfnorm	-20.0239	-27.4804	7.4566	-12.5673	0.00109	16
GPF	gamma	-20.9083	-28.9271	8.0188	-12.8895	0.00128	12
GPF	gen.gamma	-20.8332	-28.1687	7.3355	-13.4978	0.00173	9
GPF	weibull	-19.6832	-27.0591	7.3759	-12.3073	0.00095	17
GPF	lognorm	-28.8473	-36.7637	7.9164	-20.9309	0.07121	3



(a) Cobb-Douglas (b) Translog

(c) CES (d) GPF

Figure 1: Posterior distributions of the ranks of the best firm (left column) and the worst firm (right column) for a truncated normal inefficiency term considering the four production functions.

JAGS commands

JAGS commands used for the model with a GPF and a truncated normal distribution for u_i.

```
data {
  for (i in 1:N) {
    zeros[i] <- 0
    X[i,1] <- Xreg[i,1]-xbar[1]
    X[i,2] <- Xreg[i,2]-xbar[2]
  }
  xbar[1] <- mean(Xreg[,1])
  xbar[2] <- mean(Xreg[,2])
}
model {
  for(i in 1:N) {
    zeros[i] ~ dpois(p[i])
    p[i] <- -0.5*log(tau/(2*3.141593))
    +0.5*tau*pow(y[i]+gamma*exp(y[i])-mu[i],2)
    -log(1+gamma*exp(y[i]))+10000
    mu[i] <- alpha0 +inprod(beta[],X[i,]) - u[i]
  }
  beta[1] ~ dnorm(0.0,0.001)T(0,)
  beta[2] ~ dnorm(0.0,0.001)T(0,)

  (*)
```

```
  (*)
  alpha0 ~ dnorm(0.0,0.001)
  gamma ~ dgamma(0.1,0.01)
  tau ~ dgamma(0.01,0.01)
  sigma2 <- 1/tau
  for (i in 1:N) {
    u[i] ~ dnorm(zeta,invlambda)T(0,)
    eff[i] <- exp(-u[i])
  }
  invlambda ~ dgamma(1,1/37.5)
  zeta ~ dnorm(0.0, 1E-3)T(0,)
  lambda <- 1/invlambda
  alpha <- alpha0-inprod(beta[],xbar[])
}
```

References

Burnham, K. P. and D. R. Anderson (1998). *Model Selection and Inference: A Practical Information-Theoretic Approach*. Springer: New York.

Griffin, J. E. and M. F. J. Steel (2004). Flexible mixture modelling of stochastic frontiers. Technical Report 418, University of Warwick.

Plummer, M. (2003). JAGS: A program for analysis Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, and A. Zeileis (Eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC, 2003)*. Technische Universitaet Wien, Vienna, Austria. URL http://www.ci.tuwien.ac.at/Conferences/DSC.html.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B 64*, 1–34.

van den Broeck, J., G. Koop, J. Osiewalski, and M. F. J. Steel (1994). Stochastic frontier models: A Bayesian perspective. *Journal of Econometrics 61*, 273–303.