Wageningen Summer School in Econometrics

The Bayesian Approach in Theory and Practice

September 2008

Slides for Lecture on

The Linear Regression Model with Panel Data

Gary Koop, University of Strathclyde

1 Summary

- Readings: Chapter 7 of textbook.
- I will cover the pooled and individual effects models.
- A very popular version of the individual effects model: the stochastic frontier model will also be discussed.
- Computational tools: Gibbs sampler with data augmentation.

2 Notation

- y_{it} and ε_{it} denote t^{th} observations (for $t = 1, .., T$) of the dependent variable and error, respectively, for i^{th} individual for $i = 1, .., N$.

- y_i and ε_i denote vectors of T observations of dependent variable and error, respectively, for i^{th} individual.

- Sometimes it is important to distinguish between the intercept and slope coefficients. Hence, define X_i to be a $T \times k$ matrix containing the T observations on each of k explanatory variables (including intercept) for i^{th} individual. $\widetilde{X}_i$ is $T \times (k-1)$ matrix equal to X_i with intercept removed.

- If we stack observations for all N individuals together, we obtain the $TN-$vectors:

$$y = \begin{bmatrix} y_1 \\ . \\ . \\ y_N \end{bmatrix} \text{ and } \varepsilon = \begin{bmatrix} \varepsilon_1 \\ . \\ . \\ \varepsilon_N \end{bmatrix}.$$

Similarly, stacking observations on all explanatory variables together yields the $TN \times K$ matrix:

$$X = \begin{bmatrix} X_1 \\ . \\ . \\ X_N \end{bmatrix}.$$

3 The Pooled Model

Assume that same linear regression relationship holds for every individual and, hence,

$$y_i = X_i\beta + \varepsilon_i,$$

for $i = 1, .., N$ where β is the $k-$vector of regression coefficients, including the intercept.

This is just a linear regression model of sort discussed in previous lectures. No new issues arise.

4 Individual Effects Models

- Model is of the form:

$$y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it},$$

4.1 The Likelihood Function

- Likelihood function for this model is based on the regression equation:

$$y_i = \alpha_i \iota_T + \widetilde{X}_i \widetilde{\beta} + \varepsilon_i,$$

- Properties of multivariate Normal imply likelihood function of the form:

$p(y|\alpha, \widetilde{\beta}, h) =$
$\prod_{i=1}^{N} \frac{h^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \left\{ \exp \left[-\frac{h}{2} \left(y_i - \alpha_i - \widetilde{X}_i \widetilde{\beta} \right)' \left(y_i - \alpha_i - \widetilde{X}_i \widetilde{\beta} \right) \right] \right\}$,

where $\alpha = (\alpha_1, .., \alpha_N)'$.

4.2 The Prior

- Any sort of prior, including a noninformative one. Here we consider two types of priors which are computationally simple and commonly used.

4.2.1 A Non-hierarchical Prior

- Individual effects model can be written as:

$$y = X^*\beta^* + \varepsilon,$$

where X^* is a $TN \times (N + k - 1)$ matrix given by

$$X^* = \begin{bmatrix} \iota_T & 0_T & . & . & 0_T & \widetilde{X}_1 \\ 0_T & \iota_T & . & . & . & \widetilde{X}_2 \\ . & 0_T & . & . & . & . \\ . & . & . & . & 0_T & . \\ 0_T & . & . & . & \iota_T & \widetilde{X}_N \end{bmatrix}$$

and

$$\beta^* = \begin{bmatrix} \alpha_1 \\ . \\ . \\ \alpha_N \\ \widetilde{\beta} \end{bmatrix}.$$

- Thus, individual effects model can be written as a regression model (with individual dummy variables).

- Here we use the independent Normal-Gamma prior (but could also use natural conjugate prior):

$$\beta^* \sim N\left(\underline{\beta^*}, \underline{V}\right),$$

and

$$h \sim G\left(\underline{s}^{-2}, \underline{\nu}\right).$$

4.2.2 A Hierarchical Prior

- Hierarchical priors are increasingly popular in many fields in models with high-dimensional parameter spaces (such as the individual effects model).

- Consider a prior:

$$\alpha_i \sim N\left(\mu_\alpha, V_\alpha\right)$$

with α_i and α_j being independent of one another for $i \neq j$.

- Hierarchical structure of the prior arises if we treat μ_α and V_α as unknown parameters which require their own prior.

- We assume μ_α and V_α to be independent of one another with:

$$\mu_\alpha \sim N\left(\underline{\mu}_\alpha, \underline{\sigma}_\alpha^2\right)$$

and

$$V_\alpha^{-1} \sim G\left(\underline{V}_\alpha^{-1}, \underline{\nu}_\alpha\right).$$

- Hierarchical prior assumes all intercepts are drawn from same distribution.

- This extra structure (if consistent with patterns in the data), allows for more accurate estimation.

- For the remaining parameters, we assume a non-hierarchical prior of the independent Normal-Gamma variety.

$$\widetilde{\beta} \sim N\left(\underline{\beta}, \underline{V}_{\beta}\right),$$

and

$$h \sim G\left(\underline{s}^{-2}, \underline{\nu}\right).$$

- This model is analogous to the frequentist random effects model.

4.3 Bayesian Computation

4.3.1 Posterior Inference under the Non-hierarchical Prior

- Under the non-hierarchical prior, we have a linear regression model with independent Normal-Gamma prior. Hence, posterior inference can be can be carried out using methods in Chapter 4.

4.3.2 Posterior Inference under the Hierarchical Prior

- A Gibbs sampler can be used

- The relevant posterior distributions for $\widetilde{\beta}$ and h, conditional on α, are

$$\widetilde{\beta}|y, h, \alpha, \mu_\alpha, V_\alpha \sim N\left(\overline{\beta}, \overline{V}_\beta\right).$$

$$h|y, \widetilde{\beta}, \alpha, \mu_\alpha, V_\alpha \sim G(\overline{s}^{-2}, \overline{\nu})$$

$$\alpha_i|y, \widetilde{\beta}, h, \mu_\alpha, V_\alpha \sim N\left(\overline{\alpha}_i, \overline{V}_i\right),$$

$$\mu_\alpha|y, \widetilde{\beta}, h, \alpha, V_\alpha \sim N\left(\overline{\mu}_\alpha, \overline{\sigma}_\alpha^2\right),$$

$$V_\alpha^{-1}|y, \widetilde{\beta}, h, \alpha, \mu_\alpha \sim G\left(\overline{V}_\alpha^{-1}, \overline{\nu}_\alpha\right),$$

where formulae for arguments of these densities given in textbook, pages 152-154.

- Key things are that derivations above simple extensions of those for Normal linear regression model and that is that Gibbs sampler requires only random number generation from the Normal and Gamma distributions.

- Note: the random coefficients model is given by:

$$y_i = X_i\beta_i + \varepsilon_i,$$

where β_i varies over observation. Discussed in textbook, pages 155-157. (Simple extension of individual effects model so I will not discuss it here).

5 Efficiency Analysis and the Stochastic Frontier Model

- To motivate model, let output of firm i at time t, Y_{it}, be produced using a vector of inputs, X_{it}^{*}, .

- Firms have access to a common best-practice technology for turning inputs into output:

$$Y_{it} = f(X_{it}^{*}; \beta).$$

- *Production frontier* measures the maximum amount of output that can be obtained from a given level of inputs.

- Deviation of actual from maximum feasible output is a measure of inefficiency.

- Formally:

$$Y_{it} = f(X_{it}^*; \beta)\tau_i,$$

where $0 < \tau_i \leq 1$ is a measure of firm-specific efficiency and $\tau_i = 1$ indicates firm i is fully efficient.

- Example: $\tau_i = 0.75$ means that firm i is producing only 75% of the output it could have if it were operating according to best-practice technology.

- In this specification, we have assumed each firm has a particular efficiency level which is constant over time. This assumption can be relaxed.

- Adding a random error to the model, ζ_{it}, to capture measurement (or specification) error:

$$Y_{it} = f(X_{it}^*; \beta)\tau_i\zeta_{it}.$$

- Common for $f()$ to be log-linear (e.g. Cobb-Douglas or translog):

$$y_{it} = X_{it}\beta + \varepsilon_{it} - z_i,$$

where $y_{it} = ln(Y_{it})$, $\varepsilon_{it} = ln(\zeta_{it})$, $z_i = -ln(\tau_i)$ and X_{it} is the counterpart of X_{it}^* with the inputs transformed to logarithms or, if we stack into matrices:

$$y_i = X_i\beta + \varepsilon_i - z_i\iota_T$$

- z_i is referred to as inefficiency and, since $0 < \tau_i \leq 1$, it is a non-negative random variable. X_{it} is assumed to contain an intercept and β_1 is its coefficient.

- Note that this model is of the form of an individual effects model: $\beta_1 - z_i$ plays the same role that α_i did earliuer on.

5.1 Bayesian Inference in the Stochastic Frontier Model

- Bayesian inference in stochastic frontier model very similar to individual effects model, so we will only sketch out details.

- The important new issue here is the inefficiency term, z_i, so focus on that.

- Hierarchical prior for inefficiencies: Since $z_i > 0$, the Normal hierarchical prior we used with the individual effects model.

- Common choices for this prior include the truncated-Normal and members of the family of Gamma distributions.

- Here we will use the exponential distribution, which is the Gamma with two degrees of freedom

$$z_i \sim G(\mu_z, 2).$$

- μ_z requires a prior. It is computationally convenient to use:

$$\mu_z^{-1} \sim G\left(\underline{\mu}_z^{-1}, \underline{\nu}_z\right).$$

- Posterior inference can be carried out by setting up a Gibbs sampler. Thus, we must derive the full conditional posterior distributions. Derivations which are very similar to those done previously imply:

$$\beta | y, h, z, \mu_z \sim N\left(\overline{\beta}, \overline{V}\right).$$

$$h|y, \beta, z, \mu_z \sim G(\overline{s}^{-2}, \overline{\nu})$$

$$p(z_i|y_i, X_i, \beta, h, \mu_z) \propto \\ f_N(z_i|\overline{X}_i\beta - \overline{y}_i - (Th\mu_z)^{-1}, (Th)^{-1})1(z_i \geq 0) \; ,$$

$$\mu_z^{-1}|y, \beta, h, z \sim G(\overline{\mu}_z, \overline{\nu}_z)$$

where formulae for arguments of densities are given in the book.

- Gibbs sampler involves drawning from Normal, truncated Normal and Gamma distributions – all straightforward to do.

5.2 Empirical Illustration: Efficiency Analysis with Stochastic Frontier Models

- To illustrate Bayesian inference in the stochastic frontier model, artificial data was generated from:

$$y_{it} = 1.0 + 0.75 x_{2,it} + 0.25 x_{3,it} - z_i + \varepsilon_{it}$$

for $i = 1, .., 100$ and $t = 1, .., 5$. We assume $\varepsilon_{it} \sim N(0, 0.04)$, $z_i \sim G(-\ln[.85], 2)$, $x_{2,it} \sim U(0,1)$ and $x_{2,it} \sim U(0,1)$.

- The inefficiency distribution is selected so as to imply the median of the efficiency distribution is 0.85.

- Priors are relatively noninformative (see textbook).

- Posterior results for all models are based on Gibbs sampler taking $30,000$ replications, with $5,000$ burn in replications discarded and $25,000$ replications retained.

- MCMC diagnostics indicate convergence of all the Gibbs samplers and numerical standard errors indicate an approximation error which is small relative to posterior standard deviations of all parameters.

- Table 7.3 contains posterior means and standard deviations for the parameters of the stochastic frontier model.

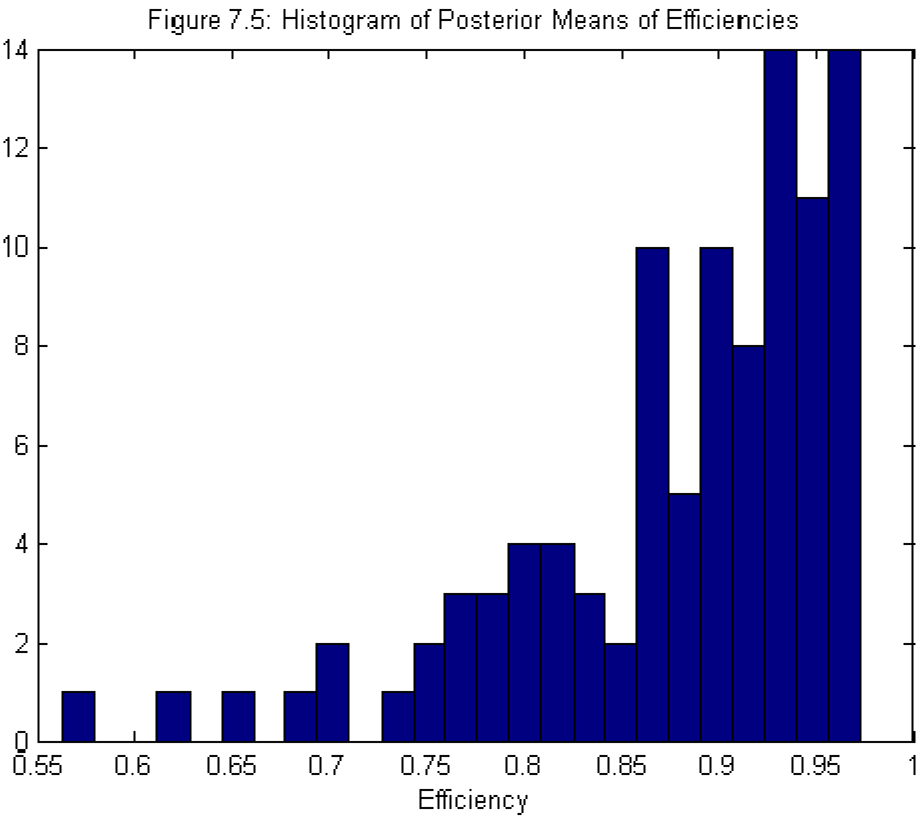
- With stochastic frontier models, interest often centers on the firm-specific efficiencies, τ_i for $i = 1, .., N$. Since $\tau_i = \exp(-z_i)$, and the Gibbs sampler yields draws of z_i, we can simply transform them and average to obtain $E(\tau_i|y)$

- For the sake of brevity, we do not present results for all $N = 100$ efficiencies. Rather we select the firms which have the minimum, median and maximum values for $E(\tau_i|y)$. These are labelled τ_{min}, τ_{med} and τ_{max} in Table 7.3.

- The histogram in Figure 7.5 plots the posterior means of the efficiencies of all 100 firms, might be presented to give a rough idea of how efficiencies are distributed across firms.

Table 7.3: Posterior Results for Artificial Data Set from Stochastic Frontier Model

	Mean	Standard Deviation
β_1	0.98	0.03
β_2	0.74	0.03
β_3	0.27	0.03
h	26.69	1.86
μ_z	0.15	0.02
τ_{min}	0.56	0.05
τ_{med}	0.89	0.06
τ_{max}	0.97	0.03



- An important issue in efficiency analysis is whether point estimates can be treated as a reliable guide to the ranking of firms. Important policy recommendations may hang on a finding that firm A is less efficient than firm B.

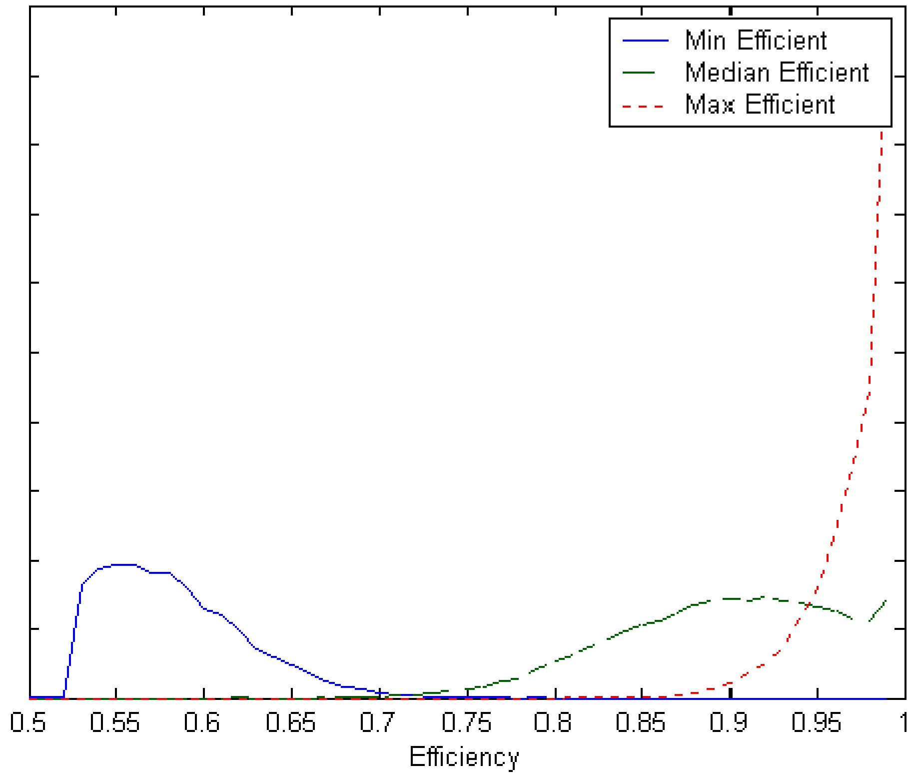
- Simply relying on point estimates which indicate that firm A is less efficient than firm B may lead to inappropriate policy advice. But Gibbs sampler output can be used in a straightforward manner to shed light on this issue. For instance, $p\left(\tau_A < \tau_B | y\right)$ is the probability firm A is less efficient than firm B.

- We find $p\left(\tau_{\text{max}} > \tau_{med} | y\right) = 0.89, p\left(\tau_{\text{max}} > \tau_{\text{min}} | y\right) = 1.00$ and $p\left(\tau_{med} > \tau_{\text{min}} | y\right) = 1.00$.

- Thus, we can conclude that firms which are ranked far apart in terms of their efficiency estimates do truly differ in efficiency. However, it is likely the case that, e.g., the researcher would be very uncertain about saying the 12^{th} ranked firm is more efficient than the 13^{th} ranked.

- Figure 7.6 plots the full posteriors for $\tau_{\text{min}}, \tau_{med}$ and τ_{max}.

Figure 7.6: Posteriors of Min/Med/Max Efficient Firms



6 Extensions/Applications

- Panel data topics very popular right now in the econometrics literature.

- The panel data models introduced in this chapter are useful for modeling heterogeneity of various sorts. This is a crucial issue in many fields.

- E.g. marketing has consumer heterogeneity, labor economics, individuals may vary in many ways that cannot be directly observed by the econometrician (e.g. they may differ in their returns to schooling, their value of leisure, their productivity, etc.).

- Dynamic panel data models are very hot these days (i.e. T is large enough that you have to start worrying about time series and unit root issues).