PREFACE

Stochastic Frontier and Data Envelopment Analysis Software

Only two of the large integrated econometrics programs currently in general use provide programs and routines for frontier and efficiency analysis, LIMDEP/NLOGIT and Stata. The freeware program, FRONTIER 4.1 by Tim Coelli (find it on the web) can also be used for a small range of stochastic frontier models. FRONTIER is now rather old, however, and is not being updated or maintained. This document, prepared by Richard Hofler of the University of Central Florida, compares the capabilities of LIMDEP and Stata for estimation of stochastic frontier models. The description is current as of 2006. Some additional capabilities have since been added to LIMDEP, notably the simulation based estimator for a sample selection corrected stochastic frontier model. Other features and models are added to LIMDEP on an ongoing basis – this will be one of the substantive updates in the next version of LIMDEP (10.0). To my knowledge, no further development of the frontier capabilities beyond those listed here has been done or is ongoing in Stata.

Data Envelopment Analysis

There are several packages that specialize in Data Envelopment Analysis (DEA) – a search of the web for this topic will amply demonstrate the variety of tools available for this mini-industry. Some of them are quite extensive. LIMDEP also contains a program for data envelopment analysis. To my knowledge, LIMDEP is the only program (small or large) that provides both stochastic frontier and DEA capabilities. This routine is, as of early 2008, also undergoing development and will be extended in version 10.0 of LIMDEP. This document does not describe features of LIMDEP related to DEA.

Tim Coelli has also developed another, separate program, DEAP for data envelopment analysis. Like FRONTIER, however, DEAP is rather old – 1996, and is not current with methodological developments of the last decade. However, it is freeware, and it does provide the basic capabilities needed by the entry level analyst.

William Greene, New York, February 1, 2009
Cross Section Data Models

Base model (Normal - half-normal)
\[ y = \beta'x + v - u \]
where \( u = |U|, \ U \sim N(0, \sigma_u^2) \) and \( v \sim N(0, \sigma_v^2) \)

Cross Section Formulations

<table>
<thead>
<tr>
<th>Behavioral Assumptions</th>
<th>LIMDEP 8.0</th>
<th>Stata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximizing (e.g., production)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Minimizing (e.g., cost)</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distributional Variations</th>
<th>LIMDEP 8.0</th>
<th>Stata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal - half-normal</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Normal – exponential parameter = ( \lambda )</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Normal – gamma, parameters = ( \lambda, \rho )</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Normal - truncated normal</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Assumptions about mean of \( u \) (inefficiency term)

<table>
<thead>
<tr>
<th>( E[u] = 0 ) (Normal - half-normal)</th>
<th>LIMDEP 8.0</th>
<th>Stata</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E[u] = \mu ) (N-truncated N with constant mean)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>( E[u] = \mu_i = \alpha'z_i ) (N-truncated N with heterogeneous mean)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>( u = \alpha'z + w ) (w truncated N such that ( u \geq 0 ) or ( u \leq 0 ))</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

Variance of \( u \)

<table>
<thead>
<tr>
<th>( \text{var}[u] = \sigma_u^2 ) (homoskedasticity)</th>
<th>LIMDEP 8.0</th>
<th>Stata</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{var}[u] = \sigma_{ui}^2 = \exp(\gamma'z_i) ) (heteroskedasticity)</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Scaling in Exponential (and Gamma) \( \lambda = \exp(\delta'w) \)

Variance of \( v \)

<table>
<thead>
<tr>
<th>( \text{var}[v] = \sigma_v^2 ) (homoskedasticity)</th>
<th>LIMDEP 8.0</th>
<th>Stata</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{var}[v] = \sigma_v^2 = \exp(\delta'w) ) (heteroskedasticity)</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Doubly Heteroskedastic

<table>
<thead>
<tr>
<th>( \text{var}[u] = \sigma_{ui}^2 = \exp(\gamma'z_i) ) and ( \text{var}[v] = \sigma_v^2 = \exp(\delta'w) )</th>
<th>LIMDEP 8.0</th>
<th>Stata</th>
</tr>
</thead>
</table>

Doubly Heteroscedastic and Nonzero Truncation \( E[u] = \mu_i = \alpha'z_i \)

Other

Requires all data be in natural logarithms

Confidence intervals (note 1)

<table>
<thead>
<tr>
<th></th>
<th>LIMDEP 8.0</th>
<th>Stata</th>
</tr>
</thead>
</table>
Panel Data Models
Random Effects Base model (Normal - half-normal)
\[ y_{it} = \beta' x_{it} + v_{it} - u_i \]
where \( u_i = | N(0, \sigma_u^2) | \)

<table>
<thead>
<tr>
<th>LIMDEP 8.0</th>
<th>Stata</th>
</tr>
</thead>
</table>

Panel Data Formulations
Random Effects Formulations (\( v_{it} - u_i \) : note \( u \) time invariant)

<table>
<thead>
<tr>
<th>Behavioral Assumptions</th>
<th>X</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>maximizing (e.g., production)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>minimizing (e.g., cost)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alvarez-Amsler Scaling Model</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distributional Variations</th>
<th>X</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal - half-normal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal - exponential</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal - gamma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal - truncated normal</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assumptions about mean of ( u ) (inefficiency term)</th>
<th>X</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( E[u] = 0 ) (Normal - half-normal)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( E[u] = \mu ) (N-truncated N with constant mean)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( E[u] = \mu_i = \alpha' z_i ) (N-truncated N with heterogeneous mean)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other</th>
<th>X</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Battese-Coelli (1992, 1995) model (note 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Requires all data be in natural logarithms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence intervals (note 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time-invariant inefficiency terms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time-varying inefficiency terms (note 3 &amp; note 4)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Panel Data Models

Fixed Effects Base model (Normal – half-normal) (note 5)
Unobserved heterogeneity enters intercept $\alpha_i$ (note $u_i$ time invariant)
\[ y_i = \alpha_i + \beta x_i + v_i - u_i \]
where $u_i = | N(0, \sigma_u^2) |$

Panel Data Formulations

Fixed Effects Formulations ($\alpha_i$ contains unobserved heterogeneity; $v_i - u_i$ : note $u_i$ time invariant)

Behavioral Assumptions
maximizing (e.g., production) $X$
minimizing (e.g., cost) $X$

Distributional Variations
Normal - half-normal $X$
Normal - exponential $X$
Normal - gamma $X$
Normal - truncated normal $X$

NOTE: LIMDEP fits several kinds of fixed effects models with $v_i +/- u_i$ AND a fixed effect in addition. Additional fixed effects may appear in the main equation, in the mean of the truncated normal or in the variance of the truncated normal (or any two of the three).

Assumptions about mean of $u$ (inefficiency term)
\[ E[u] = 0 \] (Normal - half-normal) $X$
\[ E[u] = \mu \] (N-truncated N with constant mean) $X$
\[ E[u] = \mu_i = \alpha' z_i \] (N-truncated N with heterogeneous mean) $X$
3. Unobserved heterogeneity enters mean of $u$ (inefficiency term) 
(note $u_{it}$ time varying)

$$y_{it} = \beta' x_{it} + v_{it} - u_{it}$$

where $u_{i} = |N(\mu_i, \sigma^2_u)|$

Assumptions about mean of $u$ (inefficiency term)

- $E[u] = 0$ (Normal - half-normal)
- $\mu_i = \alpha_i + \mu$ (N-truncated N with constant mean) $X$
- $\mu_i = \alpha_i + \alpha' z_i$ (N-truncated N with doubly heterogeneous mean) $X$

Panel Data Models: Fixed Effects model (cont.)

4. Unobserved heterogeneity enters inefficiency term $u$ variance $\sigma^2_{uit}$ 
(note $u_{it}$ time varying)

$$y_{it} = \beta' x_{it} + v_{it} - u_{it}$$

where $u_{i} = |N(0, \sigma^2_u)|$

Assumptions about mean of $u$ (inefficiency term)

- $E[u] = 0$ (Normal - half-normal) $X$
- $u_i = \alpha_i + \mu$ (N-truncated N with constant mean) $X$ (see above)
- $u_i = \alpha_i + \alpha' z_i$ (N-truncated N with doubly heterogeneous mean) $X$ (see above)

Variance of $u$

$$\text{var}[u] = \sigma^2_{uit} = \sigma^2_u \times \exp(\alpha_i + \delta' z_{it})$$ $X$
Random Parameters (Random Effects) Formulation

\[ y_{it} = \beta_i x_{it} + v_{it} - u_{it} \quad \text{(note: } \beta_i) \]

where

\[ v_{it} = N(0, \sigma_v^2) \]

\[ u_{it} = |N(\mu_u, \sigma_u^2)| \]

\[ \mu_{it} = \delta_i z_{it} \]

\[ \sigma_u^2 = \sigma_v^2 \times \exp(\gamma_i w_{it}) \]

Behavioral Assumptions

- maximizing (e.g., production)  \( X \)
- minimizing (e.g., cost)  \( X \)

Distributional Variations

- Normal - half-normal  \( X \)
- Normal - exponential
- Normal - gamma
- Normal - truncated normal  \( X \)

Assumptions about mean of u (inefficiency term)

- \( E[u] = 0 \) (Normal - half-normal)  \( X \)
- \( E[u] = \mu \) (N-truncated N with constant mean)  \( X \)
- \( E[u] = \mu_{it} = \delta_i z_{it} \) (N-truncated N with heterogeneous mean)  \( X \)

Random Parameters (Random Effects) Formulation (cont.)

Variance of u

- \( \text{var}[u] = \sigma_u^2 \) (homoskedasticity)  \( X \)
- \( \text{var}[u] = \sigma_u^2 = \sigma_v^2 \times \exp(\gamma_i z_i) \) (firmwise heteroskedasticity)  \( X \)
- \( \text{var}[u] = \sigma_u^2 = \sigma_v^2 \times \exp(\gamma_i z_i) \) (timewise heteroskedasticity)  \( X \)
- \( \text{var}[u] = \sigma_u^2 = \sigma_v^2 \times \exp(\gamma_i z_{it}) \) (firmwise/ timewise heteroskedasticity)  \( X \)
Latent Class Formulation

\[ y_{it} | j = \beta_j x_{it} + v_{it} - u_{it} \] (note: \( \beta_j \) - J classes)

where

\[ v_{it} | j = N(0, \sigma_v^2) \]

\[ u_{it} | j = N(0, \sigma_u^2) \]

**Behavioral Assumptions**

- Maximizing (e.g., production) \( \times \)
- Minimizing (e.g., cost) \( \times \)

**Distributional Variations**

- Normal - half-normal \( \times \)
- Normal - exponential
- Normal - gamma
- Normal - truncated normal

**Assumptions about mean of u (inefficiency term)**

- \( E[u] = 0 \) (Normal - half-normal) \( \times \)
- \( E[u] = \mu \) (N-truncated N with constant mean)
- \( E[u] = \mu_{it} = \delta_i z_{it} \) (N-truncated N with heterogeneous mean)

**Variance of u**

- \( \text{var}[u] = \sigma_u^2 \) (homoskedasticity) \( \times \)
- \( \text{var}[u] = \sigma_{u2}^2 = \sigma_u^2 \times \exp(\gamma' z) \) (firmwise heteroskedasticity)
- \( \text{var}[u] = \sigma_{u2}^2 = \sigma_u^2 \times \exp(\nu' z_{it}) \) (timewise heteroskedasticity)
- \( \text{var}[u] = \sigma_{u2}^2 = \sigma_u^2 \times \exp(\nu' z_{it}^2) \) (firmwise/timewise heteroskedasticity)

**Notes**

1. Horrace & Schmidt (1996) contains formulas for calculating the correct CIs in the stochastic frontier model. I have written LIMDEP commands to calculate those H & S CIs.
2. The Battese-Coelli (1995) model concurrently estimates the parameters of the stochastic frontier model and the coefficients of a model that relates selected determinants of inefficiency ("z" variables) to the inefficiency estimates.
3. Stata: It estimates the Battese-Coelli (1992) model in which all firms’ series of inefficiency estimates follow the same time path.
4. LIMDEP: I have written LIMDEP commands to allow each firm’s series of inefficiency estimates to (potentially) follow its own unique time path.