E63: Heteroscedasticity and Truncation in Stochastic Frontier Models

E63.1 Introduction

This chapter develops several extensions of the stochastic frontier model presented in Chapter E62. The four models considered here are as follows:

- Heteroscedasticity in v and/or u
- Truncated normal with nonzero, heterogeneous mean in the underlying U
- Heterogeneity in the parameter of the exponential or gamma distribution
- Amsler et al.'s 'scaling model'

E63.2 Heteroscedasticity and Heterogeneity

In the development of the frontier model, an important question concerns how to introduce observed heterogeneity into the specification. Suppose the vector of variables $\mathbf{z}_i$ contains the information. For example, in the airline data, we have data on load factor, stage length and number of points in the route map, that may also impact production, cost and efficiency. In the model proposed thus far, the only point at which one might introduce $\mathbf{z}_i$ appears to be in the goal function itself, which would become

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + \boldsymbol{\alpha}'\mathbf{z}_i + v_i - u_i.$$

This is a common approach. (See, e.g., Greene (2004a,b).) In this chapter, we present two other methods of introducing observed heterogeneity in the frontier model, in the variance parameters and in the mean of the underlying inefficiency.

E63.2.1 Heterogeneity in the Scale Parameters

A natural departure point is to allow observable variation in σ_v^2 and/or σ_u^2. For the first of these, the term heteroscedasticity is appropriate. (The papers by Hadri et al. (1999, 2003a,b) develop heteroscedasticity models for frontier specifications.) For the second of these, a result which seems routinely to be overlooked in the literature is that allowing σ_u^2 to vary over observations, call it $\sigma_{u,i}^2$, induces more than just heteroscedasticity. Unavoidably in all model specifications, when this parameter varies over individuals, then both the variance and the mean of u_i do also. For the half normal model, regardless of how $\sigma_{u,i}$ varies,

$$E[u_i] = \sigma_{u,i}\phi(0)/\Phi(0) = 0.79788\sigma_{u,i}.$$

A like result emerges in the truncated normal model. In the exponential model, the mean of u_i equals its standard deviation, while in the gamma model, it is a multiple, $P^{1/2}$, of it. Thus, in all cases, as regards u_i, the term heteroscedasticity, while not inappropriate, is nonetheless ambiguous. These models cannot be heteroscedastic without also having a heterogeneous mean. In what follows, therefore, we continue to use the familiar terminology, but we emphasize the nature of the model as well.

The models of scale heterogeneity may extend either variance parameter with the specification of the variance functions

$$\text{Var}[U|\mathbf{z}_i] = \sigma_{ui}^2 = \sigma_u^2 \exp(\boldsymbol{\gamma}'\mathbf{z}_i) \qquad \text{(heteroscedastic)}$$

$$\text{Var}[v|\mathbf{z}_i] = \sigma_v^2 = \sigma_v^2 \exp(\boldsymbol{\delta}'\mathbf{w}_i) \qquad \text{(heteroscedastic)}$$

$$\text{Var}[u|\mathbf{z}_i] = \sigma_u^2 \exp(\boldsymbol{\gamma}'\mathbf{z}) \text{ and } \text{Var}[v|\mathbf{z}_i] = \sigma_v^2 \exp(\boldsymbol{\delta}'\mathbf{w}_i) \qquad \text{(doubly heteroscedastic)}$$

There is no requirement that the same variables enter the two functions, and either or both may be heterogeneous. The model specification is

; Heteroscedasticity or ; Het

and either or both of

; Hfv = variables in the variance of v
; Hfu = variables in the variance of u

If either variance is not given, it is assumed to be constant. The variance function is the exponential format used throughout *LIMDEP* If either variance is unspecified, the implied model is $\sigma_{ji}^2 = \exp(\delta$ or $\gamma)$ which is the same as

; Hfv = one or **; Hfu =one**.

If both are unspecified, then the implied model

; Het ; Hfv = one ; Hfu = one

is the default, normal-half normal stochastic frontier model. It provides identical estimates. (Try it.) A constant (*one*) is automatically inserted into both lists if you do not include it. This form may be used with the normal-half normal and normal-truncated normal models.

E63.2.2 Exponential and Gamma Models with Heterogeneity

The one sided component of the normal-exponential and normal-gamma models is parameterized with a scale parameter, θ, which is thus far taken to be a constant. In these models,

$$E[u_i] = P/\theta = P\times\sigma_u$$

where $P = 1$ in the exponential model. The exponential heteroscedasticity model for u_i is extended to these two models by using

$$\theta_i = \theta \exp(-\boldsymbol{\delta}'\mathbf{z}_i).$$

With this parameterization, the estimates from this model will be comparable to those for the half normal and truncated normal models. (See the examples below.) To request this form, use

; Het ; Hfu = the list of variables.

The list should not contain a constant term, *one*. This may be used in all implementations of the exponential gamma model. Note, however, that in the panel data settings, the parameter is assumed to be time invariant. The values for $\mathbf{z}_i$ are taken from the data record for the last period for firm i. We will return to this subject below. The symmetric component, v, may also be heteroscedastic, as in the other models, with

; Hfv = list of variables.

E63.2.3 Efficiency Estimation with Heteroscedasticity

This extension does not change the computation of measures of efficiency or inefficiency. The central results are the JLMS estimators,

$$\hat{E}[u\,|\,\varepsilon] = \frac{\sigma\lambda}{1+\lambda^2}\left[\frac{\phi(w)}{1-\Phi(w)} - w\right],\ \varepsilon = v - u\,,\ w = S\varepsilon\lambda/\sigma$$

for the half normal models and

$$\hat{E}[u\,|\,\varepsilon] = \sigma_v\left[\frac{\phi(w)}{1-\Phi(w)} - w\right],\ w = (S\varepsilon/\sigma_v + \theta\sigma_v)$$

for the exponential models. These functions are evaluated for each observation at

$$\lambda_i = \sigma_{u,i} / \sigma_{v,i}$$

and

$$\sigma_i^2 = \sigma_{u,i}^2 + \sigma_{v,i}^2$$

for the half normal model and $\sigma_{v,i}$ and θ_i likewise in the exponential and gamma models.

E63.2.4 Application

The estimates below show a production frontier based on the six inputs. The second set of results presents the heteroscedastic model, with the variance of v a function of the log of the average stage length and the variance of u depending on the load factor and the log of the number of points served. We examine the efficiency results, then compute the average partial effects of the environmental variables on technical efficiency.

```
FRONTIER   ; Lhs = lq ; Rhs = one,ll,lp,lf,le,lm,lk ; Techeff = eu $
FRONTIER   ; Lhs = lq ; Rhs = one,ll,lp,lf,le,lm,lk ; Techeff = euhet
           ; Het ; Hfv = lstage ; Hfu = loadfctr,points $
PARTIALS   ; Effects: lstage / loadfctr / points ; Summary $
KERNEL     ; Rhs = eu,euhet
           ; Title = Kernel Estimators for Technical Efficiency $
PLOT       ; Lhs = eu ; Rhs = euhet ; Rh2 = eu ; Fill ; Grid
           ; Title = Estimates of Technical Efficiency
           ; Vaxis = exp(-E[u|e]) for Heteroscedastic Model $
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                    LQ
Log likelihood function        108.43918
Estimation based on N =    256, K =   9
Variances: Sigma-squared(v)=      .01902
           Sigma-squared(u)=      .01692
           Sigma(v)        =      .13791
           Sigma(u)        =      .13007
Sigma = Sqr[(s^2(u)+s^2(v)]=      .18957
Gamma = sigma(u)^2/sigma^2 =      .47074
Var[u]/{Var[u]+Var[v]}     =      .24425
Stochastic Production Frontier, e = v-u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0           108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =      .730
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -2.98823***      .72136    -4.14  .0000    -4.40206  -1.57439
      LL|    -.42909***      .06315    -6.79  .0000     -.55287   -.30530
      LP|     .44533***      .09498     4.69  .0000      .25917    .63149
      LF|     .37257***      .07038     5.29  .0000      .23463    .51052
      LE|    2.09473***      .68790     3.05  .0023      .74647   3.44299
      LM|     .69910***      .07580     9.22  .0000      .55054    .84766
      LK|   -2.09806***      .76556    -2.74  .0061    -3.59853   -.59759
        |Variance parameters for compound error
  Lambda|     .94309***      .16870     5.59  .0000      .61244   1.27373
   Sigma|     .18957***      .00064   297.81  .0000      .18832    .19082
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                       LQ
Log likelihood function       149.30854
Estimation based on N =    256, K =  12
Inf.Cr.AIC  =   -274.6 AIC/N =   -1.073
Variances: Sigma-squared(v)=     .01292
           Sigma-squared(u)=     .03575
           Sigma(v)        =     .11367
           Sigma(u)        =     .18907
Sigma = Sqr[(s^2(u)+s^2(v)]=     .22061
Gamma = sigma(u)^2/sigma^2 =     .73450
Var[u]/{Var[u]+Var[v]}     =     .50132
Variances averaged over observations
Stochastic Production Frontier, e = v-u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  2
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  3
LogL when sigma(u)=0          108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =   82.468
Kodde-Palm C*: 95%: 8.761,  99%: 12.483
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -3.29243***      .72664     -4.53  .0000    -4.71662  -1.86824
      LL|    -.47507***      .08890     -5.34  .0000     -.64932   -.30083
      LP|     .50435***      .10452      4.83  .0000      .29950    .70920
      LF|     .53204***      .07550      7.05  .0000      .38406    .68003
      LE|    2.36654***      .69245      3.42  .0006     1.00936   3.72372
      LM|     .53413***      .08670      6.16  .0000      .36419    .70406
      LK|   -2.43136***      .77258     -3.15  .0016    -3.94558   -.91713
        |Parameters in variance of v (symmetric)
Constant|   -3.97891***      .86601     -4.59  .0000    -5.67626  -2.28155
  LSTAGE|    -.06406         .13359      -.48  .6315     -.32590    .19777
        |Parameters in variance of u (one sided)
Constant|    9.96191**      4.51238      2.21  .0273     1.11781  18.80600
LOADFCTR|   -25.9711***     9.37571     -2.77  .0056    -44.3471   -7.5950
  POINTS|    -.00353         .01288      -.27  .7840     -.02877    .02171
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

The figure below displays the kernel density estimators for the two sets of estimated inefficiencies. The upper one is for the heteroscedastic model. The figure shows clearly the influence of the heterogeneity. The means of the two distributions are virtually the same, but the variance in the heteroscedastic model is considerably higher.

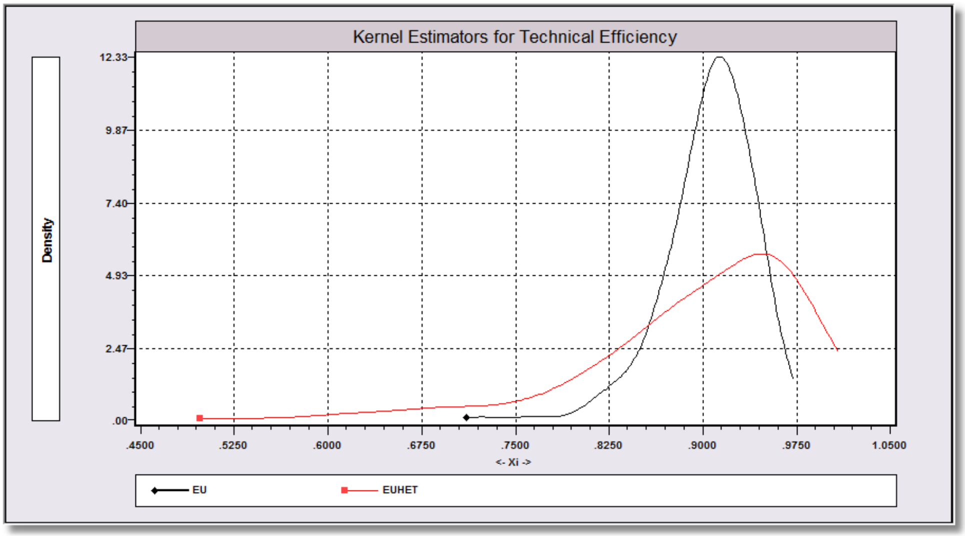

Figure E63.1 Kernel Estimators for Density of E[u|ε] with and without Heteroscedasticity

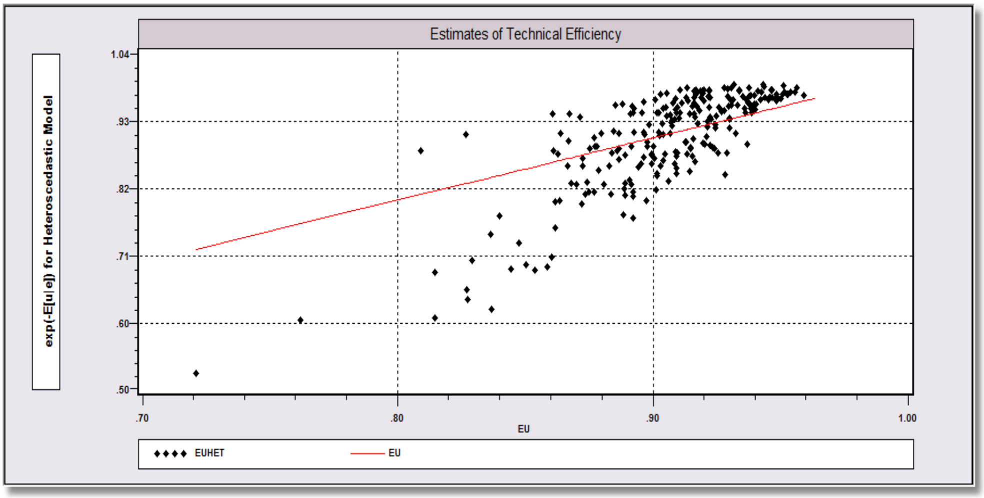

Figure E63.2 Plot of Estimated Inefficiencies, Heteroscedastic vs. Homoscedastic

```
----------------------------------------------------------------------
Partial Effects for JLMS Estimator in Normal/het SF Model
Partial Effects Averaged Over Observations
* ==> Partial Effect for a Binary Variable
----------------------------------------------------------------------
                     Partial     Standard
(Delta method)       Effect       Error      |t|  95% Confidence Interval
----------------------------------------------------------------------
      LSTAGE        -.00034      .00071      .48     -.00174      .00105
      LOADFCTR       .62934      .17576     3.58      .28485      .97382
      POINTS         .00009      .00031      .28     -.00052      .00069
----------------------------------------------------------------------
```

E63.2.5 Technical Details

For the models with heteroscedasticity, we revert to the original structural form of the model to form the log likelihoods. For the normal-half normal model, for example, we use

$$\log L_i = -\log(2/\pi) - \log\sigma_i - ½(\varepsilon_i/\sigma_i)^2 + \log\Phi[-S\varepsilon_i\lambda_i/\sigma_i]$$

where

$$\sigma_i = \sqrt{\sigma_{ui}^2 + \sigma_{ui}^2}$$

$$\lambda_i = \sigma_{ui} / \sigma_{vi}$$

$$\sigma_{ui}^2 = \exp(\boldsymbol{\gamma}'\mathbf{z}_i)$$

$$\sigma_{vi}^2 = \exp(\boldsymbol{\delta}'\mathbf{w}_i),$$

where S = +1 for a production frontier and -1 for a cost frontier. Likewise, for the truncation model,

$$\log L_i = -½\log 2\pi - \log\sigma_i - ½[(S\varepsilon_i + \mu)/\sigma_i]^2$$
$$+ \log\Phi[(\mu/\lambda_i - S\varepsilon_i\lambda_i)/\sigma_i] - \log\Phi(\mu/\sigma_{u.i}).$$

We build the structure of the model with two freely varying variance parameters, $\sigma_{u,i}$ and $\sigma_{v,i}$, rather than the reduced form parameters λ and σ. The use of λ_i as a free parameter would not be appropriate because the numerator and denominator of λ_i must be allowed to vary freely and independently. A like consideration rules out the composed parameter σ_i. The formulation of the log likelihood and its derivatives follows the results given earlier for the homogeneous cases. Where the derivatives with respect to $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ emerge, we use the chain rule to differentiate with respect to $\sigma_{u,i}$ and $\sigma_{v,i}$ first. Note that the independent parameter σ_u and σ_v have been absorbed into the exponential functions. Thus, σ_v is $\exp(\gamma_0)$. This ensures that the variances are always positive.

The normal-gamma and normal-exponential models are not reparameterized. The log likelihood for the exponential model with variance heterogeneity is

$$\log L_i = \log\theta_i + ½\theta_i^2\sigma_{i,v}^2 + \theta_i S\varepsilon_i + \log\Phi[-S\varepsilon_i/\sigma_{i,v} - \theta_i\sigma_{i,v}]$$

where

$$\theta_i = \theta\exp(-\boldsymbol{\gamma}'\mathbf{z}_i)$$

and

$$\sigma_{i,v} = \sigma_v \exp(\boldsymbol{\delta}'\mathbf{w}_i).$$

The sign change in θ_i is used to make the normal-exponential model comparable to the normal-half normal model, since $\text{Var}[u_i] = 1/\theta_i^2$.

E63.3 The Normal-Truncated Normal Model

The normal-truncated normal model relaxes an implicit restriction in the normal-half normal model, that the mean of the underlying inefficiency variable is zero. The extended model is obtained by allowing μ, the mean of *U*, to be nonzero;

$$y = \boldsymbol{\beta}'\mathbf{x} + v - u,\ u = |U|$$

$$U \sim N[\mu,\sigma_u^2]$$

$$v \sim N[0,\sigma_v^2]$$

(With a constant term in the model, no similar parameter can be introduced into the distribution of *v*.) The command for estimating this model is

FRONTIER ; Lhs = dependent variable
; Rhs = one, other independent variables
; Model = Truncated Normal $ (or ; Model = T)

The specification of the cost frontier and the estimator of technical inefficiency are requested in the same fashion,

; Cost
and **; Eff = variable name**

Other optional parts of the command are the same as that for the normal-half normal model.

We note, this model is extremely volatile, owing to the rather weak identification of the parameter μ. It is difficult to distinguish the mean from the variance parameter in this model. In the truncation model,

$$E[u_i] = \mu + \sigma_u\phi(\mu/\sigma_u)/\Phi(\mu/\sigma_u).$$

This implies that σ_u and μ can covary so as to produce little or no variation in the expectation of u_i. The likelihood is not a function of the square of u_i, so this mean is the only source of information about these two parameters. (By totally differentiating the expected value, one can solve for the implicit relationship, $d\mu/d\sigma_u$ that produces $dE[u_i] = 0$.) The example below suggests how this aspect of the model influences (or fails to) the estimates of inefficiency. For purposes of the JLMS estimator for the half normal model, when the mean of *U* is a nonzero μ, the argument to the function is replaced with

$$w = S\varepsilon\lambda/\sigma - \mu/(\sigma\lambda).$$

The remaining part of the computation is the same.

E63.3.1 Application

The results below show estimates of a stochastic cost frontier with the half normal then the truncated normal specifications. The additional parameterization appears to have had a large impact on the results; the estimates are noticeably different. The plot of the two sets of inefficiency estimates suggest that the effect of the new specification has been little more than to double the estimated values from the model – the dashed line in the figure shows the function $u_{TN} = 2u_{HN}$. The extremely large estimates of μ and the standard error do suggest that something is amiss with the model, however.

The commands are:

FRONTIER ; Lhs = lq ; Rhs = one,ll,lp,lf,le,lm,lk ; techeff = u $
FRONTIER ; Lhs = lq ; Rhs = one,ll,lp,lf,le,lm,lk ; techeff = ut ; Model = T $
PLOT ; Lhs = u ; Rhs = ut ; Rh2 = u ; Fill ; Grid
; Title = Truncated Normal Inefficiencies vs. Half Normal $
DSTAT ; Rhs = u,ut $

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                     LQ
Log likelihood function       108.43918
Variances: Sigma-squared(v)=     .01902
           Sigma-squared(u)=     .01692
           Sigma(v)        =     .13791
           Sigma(u)        =     .13007
Sigma = Sqr[(s^2(u)+s^2(v)]=     .18957
Gamma = sigma(u)^2/sigma^2 =     .47074
Var[u]/{Var[u]+Var[v]}     =     .24425
Stochastic Production Frontier, e = v-u
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0          108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =     .730
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
      LQ|  Coefficient      Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -2.98823***      .72136    -4.14  .0000    -4.40206  -1.57439
      LL|    -.42909***      .06315    -6.79  .0000     -.55287   -.30530
      LP|     .44533***      .09498     4.69  .0000      .25917    .63149
      LF|     .37257***      .07038     5.29  .0000      .23463    .51052
      LE|    2.09473***      .68790     3.05  .0023      .74647   3.44299
      LM|     .69910***      .07580     9.22  .0000      .55054    .84766
      LK|   -2.09806***      .76556    -2.74  .0061    -3.59853   -.59759
        |Variance parameters for compound error
  Lambda|     .94309***      .16870     5.59  .0000      .61244   1.27373
   Sigma|     .18957***      .00064   297.81  .0000      .18832    .19082
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                     LQ
Log likelihood function       109.49695
Estimation based on N =    256, K =  10
Variances: Sigma-squared(v)=     .01896
           Sigma-squared(u)=    2.48813
           Sigma(v)        =     .13771
           Sigma(u)        =    1.57738
Sigma = Sqr[(s^2(u)+s^2(v)]=    1.58338
Gamma = sigma(u)^2/sigma^2 =     .99244
Var[u]/{Var[u]+Var[v]}     =     .97946
Stochastic Production Frontier, e = v-u
Half Normal:u(i)=|U(i)|; frontier model
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0          108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =    2.845
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -3.11541***      .77143    -4.04  .0001    -4.62739  -1.60343
      LL|    -.44532***      .07797    -5.71  .0000     -.59814   -.29249
      LP|     .46908***      .11368     4.13  .0000      .24628    .69188
      LF|     .37437***      .07465     5.02  .0000      .22807    .52068
      LE|    2.20830***      .73883     2.99  .0028      .76023   3.65637
      LM|     .67741***      .09341     7.25  .0000      .49433    .86048
      LK|   -2.20620***      .82402    -2.68  .0074    -3.82126   -.59115
        |Offset [mean=mu(i)] parameters in one sided error
      Mu|   -31.5468       5061.203     -.01  .9950  -9951.3228  9888.2292
        |Variance parameters for compound error
  Lambda|    11.4545       907.8501      .01  .9899  -1767.8991  1790.8081
   Sigma|    1.58338       124.7546      .01  .9899  -242.93113  246.09790
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------

Descriptive Statistics
--------+---------------------------------------------------------------------
Variable|       Mean        Std.Dev.      Minimum       Maximum     Cases Missing
--------+---------------------------------------------------------------------
       U|     .902312      .035500      .703534      .963108       256        0
      UT|     .925474      .039335      .608274      .972355       256        0
--------+---------------------------------------------------------------------
```

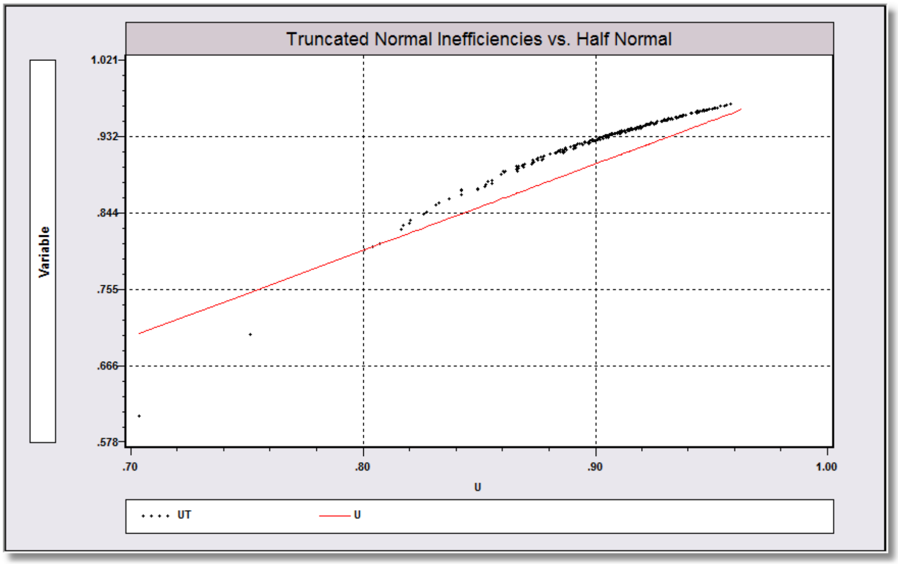

Figure E63.3 Inefficiency Estimates from Truncated Normal Model

E63.3.2 Battese and Coelli (1995) Formulation

There are (apparently) two formulations of the normal – truncated normal model in the literature. The formulated above,

$$y = \boldsymbol{\beta}'\mathbf{x} + v - u,\ u = |U|$$
$$U \sim N[\mu,\sigma_u^2] \leftarrow$$
$$v \sim N[0,\sigma_v^2]$$

is due to Stevenson (1980). Note that the inefficiency term is the absolute value of a normally distributed variable with a nonzero mean. Battese and Coelli proposed an apparently different formulation of the truncation model;

$$u = \mu + w$$

where w is a truncated normal, such that

$$w \geq -\mu.$$

This is actually the same model. You can obtain the estimates using this alternative formulation with

; Model = BC95

in place of **; Model = T**. The log likelihood for this formulation involves a one to one reparameterization of the Stevenson model, which has slightly different numerical properties. You can see this in the application below. The estimated inefficiency and efficiency values produced by the two models are the same to five or six digits, however.

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                   LQ
Log likelihood function       109.48819
Variances: Sigma-squared(v)=     .01918
           Sigma-squared(u)=    2.25705
           Sigma(v)        =     .13850
           Sigma(u)        =    1.50235
Sigma = Sqr[(s^2(u)+s^2(v)]=    1.50872
Gamma = sigma(u)^2/sigma^2 =     .99157
Var[u]/{Var[u]+Var[v]}     =     .97715
Stochastic Production Frontier, e = v-u
Battese/Coelli 1995 truncated normal model
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     1
Deg. freedom for inefficiency model:  2
LogL when sigma(u)=0          108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =    2.828
Kodde-Palm C*: 95%: 5.138,  99%:  8.273
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -3.09929***      .76919    -4.03  .0001    -4.60687  -1.59172
      LL|    -.44370***      .07771    -5.71  .0000     -.59600   -.29140
      LP|     .46535***      .11351     4.10  .0000      .24288    .68781
      LF|     .37430***      .07432     5.04  .0000      .22863    .51997
      LE|    2.18991***      .73664     2.97  .0030      .74613   3.63369
      LM|     .67921***      .09322     7.29  .0000      .49651    .86191
      LK|   -2.18647***      .82171    -2.66  .0078    -3.79700   -.57594
        |Offset [mean=z(i)*delta] parameters in one sided error
Constant|   -29.6062       4821.053     -.01  .9951  -9478.6972  9419.4848
        |Variance parameters for compound error
   Gamma|     .99157        1.34377      .74  .4606    -1.64216   3.62531
SigmaSqd|    2.27624       363.5754      .01  .9950  -710.31839  714.87086
--------+--------------------------------------------------------------------
 (Stevenson formulation)
Log likelihood function        94.86417
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -3.11541***      .77143    -4.04  .0001    -4.62739  -1.60343
      LL|    -.44532***      .07797    -5.71  .0000     -.59814   -.29249
      LP|     .46908***      .11368     4.13  .0000      .24628    .69188
      LF|     .37437***      .07465     5.02  .0000      .22807    .52068
      LE|    2.20830***      .73883     2.99  .0028      .76023   3.65637
      LM|     .67741***      .09341     7.25  .0000      .49433    .86048
      LK|   -2.20620***      .82402    -2.68  .0074    -3.82126   -.59115
        |Offset [mean=mu(i)] parameters in one sided error
      Mu|   -31.5468       5061.203     -.01  .9950  -9951.3228  9888.2292
        |Variance parameters for compound error
  Lambda|    11.4545       907.8501      .01  .9899  -1767.8991  1790.8081
   Sigma|    1.58338       124.7546      .01  .9899  -242.93113  246.09790
```

E63.3.3 Technical Details on the Truncated Normal Model

The individual term in the log likelihood for the normal-truncated normal model is

$$\log L_i = -½\log 2\pi - \log\sigma - ½[(S\varepsilon_i + \mu)/\sigma]^2 - \log\Phi(\mu/\sigma_u) + \log\Phi[(\mu/\lambda - S\varepsilon_i\lambda)/\sigma].$$

The definitions above imply that

$$\sigma_u = \sigma\lambda/\sqrt{1+\lambda^2}\ .$$

Using this and the reparameterization

$$\alpha = \mu/(\lambda\sigma)$$

produces the log likelihood for this model,

$$\text{Log } L_i = -½\log 2\pi - \log\sigma - ½(d\varepsilon_i/\sigma + \alpha\lambda)^2 - \log\Phi\left(\alpha\sqrt{1+\lambda^2}\right) + \log\Phi(\alpha - d\varepsilon_i\lambda/\sigma).$$

The function is then maximized with respect to β, σ, λ and α. After optimization, the structural parameter μ is recovered from the result $\mu = \alpha\sigma\lambda$. For the model with heterogeneity in the mean presented in Section E63.3.4,

$$\mu_i = \boldsymbol{\theta}'\mathbf{z}_i$$

we simply replace α with $\alpha_i = \boldsymbol{\alpha}'\mathbf{z}_i$, then recover the parameter vector $\boldsymbol{\theta}$ from the same transformation as before, $\boldsymbol{\theta} = \sigma\lambda\boldsymbol{\alpha}$.

For purposes of the JLMS estimator for the half normal model, when the mean of U is a nonzero μ, the argument to the function is replaced with

$$w = S\varepsilon\lambda/\sigma - \mu/(\sigma\lambda).$$

The remaining part of the computation is the same.

E63.3.4 Heterogeneity in the Mean in the Truncation Model

The models listed above are all ‘homogeneous.’ Both the means and the variances of the underlying disturbance distributions are constant. There are several models of heterogeneity available as well. Use

; Model = T ; Rh2 = list of variables that enter the mean

to specify the heterogeneity in mean model, $U_i \sim N[\boldsymbol{\alpha}'\mathbf{z}_i, \sigma_u^2]$. In formulating this model, though it is not required, you should include a constant in $\mathbf{z}_i$ (the Rh2 variables) so that the homogeneous model becomes a special case. Also, if you are fitting a panel data version of this, note that the assumption underlying the model is that the same u_i occurs in every period. Therefore, the $\boldsymbol{\alpha}'\mathbf{z}_i$ should be the same in every period. *LIMDEP* will assume this is the case, and only use the Rh2 variables provided for the first period.

E63.3.5 Truncation and Heteroscedasticity

The doubly heteroscedastic model is also available for the truncated normal stochastic frontier model. In

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + v_i - u_i$$

you may specify **; Model = Truncated Normal; Rh2 = list of variables**

and $\text{Var}[u_i] = \sigma_u^2 \exp(\boldsymbol{\delta}'\mathbf{z}_i)$ with

; Het ; Hfu = list of variables in $\mathbf{z}_i$

and/or $\text{Var}[v_i] = \sigma_v^2 \exp(\boldsymbol{\gamma}'\mathbf{w}_i)$ with

; Het ; Hfv = list of variables in $\mathbf{w}_i$

Note that since both variance functions have a free multiplicative constant, you should not include *one* in either variable list.

In the absence of the Rh2 list, the mean of the underlying truncated variable is taken to be a constant to be estimated. This formulation encompasses all of Stevenson (1980), Reifschneider and Stevenson (1991), Huang and Liu (1994), and Battese and Coelli (1995). (Notwithstanding the assertion in the Battese and Coelli paper, the latter is not a panel data treatment as observations are still assumed to be independent.)

To illustrate the truncated normal estimator, we have refit the stochastic frontier production function with a complete set of firm dummy variables (less the last one) and the load factor variable in the mean of the underlying distribution. In the second model below, we have made the variance of *v* a function of the log of the average stage length. The command set begins with a small repair to the data set. One of the firms has no observations for the load factor, stage length or points served variables – they are coded as zero in the data. These observations are bypassed, then the firm dummies for the fixed effects model are assembled.

```
SAMPLE      ; All $
REJECT      ; loadfctr = 0 $
CREATE      ; i = Seq(firm) $
CREATE      ; Expand(i,0) $
CREATE      ; lk = Log(k) $
NAMELIST    ; xp = one,lf,lm,le,ll,lp,lk $
FRONTIER    ; Lhs = lq ; Rhs = xp ; Model = T ; Rh2 = loadfctr,_i_ $
FRONTIER    ; Lhs = lq ; Rhs = xp ; Model = T ; Rh2 = loadfctr,_i_
            ; Het ; Hfv = lstage $
```

(These are ‘true fixed effects’ models.)

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                    LQ
Log likelihood function      196.20748
Estimation based on N =    256, K =  34
Inf.Cr.AIC  =   -324.4 AIC/N =   -1.267
Model estimated: Aug 22, 2011, 22:29:09
Variances: Sigma-squared(v)=     .00960
           Sigma-squared(u)=     .00389
           Sigma(v)        =     .09799
           Sigma(u)        =     .06241
Sigma = Sqr[(s^2(u)+s^2(v)]=     .11618
Gamma = sigma(u)^2/sigma^2 =     .28856
Var[u]/{Var[u]+Var[v]}     =     .12845
Stochastic Production Frontier, e = v-u
Half Normal:u(i)=|U(i)|; frontier model
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:    25
Deg. freedom for inefficiency model: 26
LogL when sigma(u)=0         108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =  176.266
Kodde-Palm C*: 95%:38.301,  99%: 45.026
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -2.92400***      .68225    -4.29  .0000    -4.26118  -1.58682
      LF|     .31938***      .09026     3.54  .0004      .14246    .49629
      LM|     .81647***      .08387     9.73  .0000      .65209    .98086
      LE|    1.99934***      .64368     3.11  .0019      .73776   3.26092
      LL|    -.42790***      .10954    -3.91  .0001     -.64260   -.21321
      LP|     .42291***      .10529     4.02  .0001      .21654    .62929
      LK|   -2.07145***      .72267    -2.87  .0042    -3.48786   -.65503
        |Offset [mean=mu(i)] parameters in one sided error
LOADFCTR|    -.83124       6.87337     -.12  .9037   -14.30280  12.64031
     I01|     .63250       4.90139      .13  .8973    -8.97405  10.23904
     I02|     .58118       4.27763      .14  .8919    -7.80282   8.96519
(Firms 3-21 omitted)
     I22|     .45249       4.00889      .11  .9101    -7.40480   8.30977
     I23|     .64687      99.45841      .01  .9948  -194.28803 195.58176
     I24|    -.19804       7.26011     -.03  .9782   -14.42760  14.03152
        |Variance parameters for compound error
  Lambda|     .63686**      .28984     2.20  .0280      .06879   1.20494
   Sigma|     .11618***     .01008    11.53  .0000      .09643    .13593
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                    LQ
Log likelihood function        215.58601
Estimation based on N =    256, K =  35
Variances: Sigma-squared(v)=     .00634
           Sigma-squared(u)=     .01037
           Sigma(u)        =     .10183
           Sigma(v)        =     .07961
Sigma = Sqr[(s^2(u)+s^2(v)]=     .12926
Variances averaged over observations
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:    25
Deg. freedom for inefficiency model: 26
LogL when sigma(u)=0           108.07431
Chi-sq=2*[LogL(SF)-LogL(LS)] =  215.023
Kodde-Palm C*: 95%:38.301,  99%: 45.026
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LQ|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|   -1.98442*       1.05055    -1.89  .0589    -4.04346    .07463
      LF|     .45669***      .11002     4.15  .0000      .24105    .67233
      LM|     .59013***      .10421     5.66  .0000      .38589    .79437
      LE|    1.11856        1.00928     1.11  .2677     -.85959   3.09671
      LL|    -.29237***      .10923    -2.68  .0074     -.50646   -.07827
      LP|     .31311**       .14333     2.18  .0289      .03220    .59402
      LK|   -1.14743        1.10875    -1.03  .3007    -3.32054   1.02568
        |Mean of underlying truncated distribution
LOADFCTR|   -2.20067***      .42161    -5.22  .0000    -3.02701  -1.37433
     I01|    1.44767***      .25736     5.63  .0000      .94326   1.95208
     I02|    1.39624***      .22401     6.23  .0000      .95718   1.83529
(Firms 3-22 omitted)
     I24|    1.29355***      .24998     5.17  .0000      .80360   1.78349
        |Scale parms. for random components of e(i)
ln_sgmaU|   -2.28443***      .02100  -108.79  .0000    -2.32559  -2.24328
ln_sgmaV|   -3.22203***     1.20573    -2.67  .0075    -5.58522   -.85884
        |Heteroscedasticity in variance of symmetric v(i)
  LSTAGE|     .11855         .19755      .60  .5485     -.26865    .50574
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

E63.4 Alvarez et al. – Equality Constrained Scaling Model

Alvarez, Amsler, Orea and Schmidt (2006) have suggested a form of the truncation model which encompasses a number of ideas in stochastic frontier modeling. Their formulation is a normal-truncated normal frontier model with

$$\mu_i = \mu\times\boldsymbol{\delta}'\mathbf{z}_i \text{ and } \sigma_{u,i} = \sigma_u\times\boldsymbol{\delta}'\mathbf{z}_i.$$

The mean and standard deviation of the underlying truncated normal variable u_i are scaled by the same linear function of the data. We are skeptical of the linear scaling of the variance, and propose our usual exponential form instead. The linear form may be natural for the mean, but it allows the variance to be negative, which is unacceptable. The model used here is

$$\mu_i = \mu\times\exp(\boldsymbol{\delta}'\mathbf{z}_i) \text{ and } \sigma_{u,i} = \sigma_u\times \exp(\boldsymbol{\gamma}'\mathbf{z}_i).$$

The Alvarez model results if $\boldsymbol{\delta} = \boldsymbol{\gamma}$. Otherwise, we allow these to be free and to produce another variant of the frontier model. Note that as stated, this model is now merely a change of the normal-truncated normal model with heteroscedasticity in which the variables enter the truncation mean function in the exponential function rather than linearly.

The equality constrained scaling model is requested with

```
FRONTIER    ; Lhs  = y ; Rhs = one, x...
            ; Model = Scaling
            ; Heteroscedasticity
            ; Rh2 = variables in mean of truncated distribution
            ; Hfu  = the same list of variables $
```

Note in this case, Rh2 and Hfu give the same list. To obtain the scaling model without forcing the equality of $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$, use

```
FRONTIER    ; Lhs  = y ; Rhs = one, x...
            ; Model = S
            ; Heteroscedasticity
            ; Rh2 = variables in mean of truncated distribution
            ; Hfu  = the same list of variables $
```

Note, **; Model = Scaling** in the equality constrained case and **; Model = S** when the equality constraint is relaxed. (In this formulation, the variable lists could differ.) To constrain $\boldsymbol{\delta} = \mathbf{0}$, which just produces the heteroscedasticity model, use

```
FRONTIER    ; Lhs  = y ; Rhs = one, x...
            ; Model = T
            ; Heteroscedasticity
            ; Hfu  = list of variables $
```

To constrain γ = **0**, you would use the available setup for the truncated normal form, but **; Model = S** rather than **; Model = T** to obtain the exponential scaling of the mean.

```
FRONTIER    ; Lhs = y ; Rhs = one, x...
            ; Model = S
            ; Rh2 = variables in mean of truncated distribution $
```

Finally, with both δ = **0** and γ = **0**, this is just the standard normal-truncated normal model.

Technical Details

The implementation of the scaling model in *LIMDEP* is just a version of the truncation model with heteroscedasticity. The modifications of that model are:

- The constant terms in the mean and variance are enforced by the program.
- The mean function is exponential.
- In the first form of the model, a constraint is imposed that the coefficients in the mean and variance functions are the same.

As Alvarez et al. note in their paper, this model is not supported by any particular theory of the frontier framework. They suggest it as a natural extension of the familiar model with truncation. Rather, they argue that the unnatural form of the model would be the one with different scaling factors in the mean and variance functions.

Application

To illustrate the scaling model, we use the airlines cost data. The cost function is fit with truncation mean and variance functions that depend on the load factor and (log of) the average stage length. The equality constraint is imposed in the first model and relaxed in the second.

```
FRONTIER    ; Lhs = lc ; Cost ; Rhs = x
            ; Model = Scaling ; Het
            ; Rh2 = loadfctr,lstage
            ; Hfu = loadfctr,lstage $
FRONTIER    ; Lhs = lc ; Cost ; Rhs = x
            ; Model = S ; Het
            ; Rh2 = loadfctr,lstage
            ; Hfu = loadfctr,lstage $
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                     LC
Log likelihood function       172.27160
Estimation based on N =    256, K =  13
Variances: Sigma-squared(v)=     .01528
           Sigma-squared(u)=     .00000
           Sigma(v)        =     .12361
           Sigma(u)        =     .00169
Sigma = Sqr[(s^2(u)+s^2(v)]=     .12363
Stochastic Frontier Scaling Model
Mean scale factor for E[u] =        .6996
Mean scale factor for V[u] =        .6996
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  2
Deg. freedom for truncation mean:     2
Deg. freedom for inefficiency model:  5
LogL when sigma(u)=0          157.91523
Chi-sq=2*[LogL(SF)-LogL(LS)] =   28.713
Kodde-Palm C*: 95%:10.371,  99%: 14.325
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    18.9477       27.00668      .70  .4829    -33.9844   71.8798
      LY|     .95234***      .02117    44.98  .0000      .91084    .99383
     LY2|     .07740***      .01534     5.04  .0000      .04733    .10747
    LPKP|    1.50434        1.86479      .81  .4198    -2.15058   5.15926
    LPLP|     .12682         .08328     1.52  .1278     -.03640    .29003
    LPMP|    -.16640        1.21907     -.14  .8914    -2.55574   2.22294
    LPEP|    -.52809         .60356     -.87  .3816    -1.71105    .65488
    LPFP|     .00151         .02141      .07  .9436     -.04045    .04348
        |Mean of Truncated Distribution, Mu then scale
    Mu_0|    2.50985       11.12070      .23  .8214   -19.28633  24.30603
LOADFCTR|    -.56559        3.85231     -.15  .8833    -8.11597   6.98479
  LSTAGE|    -.00823         .05624     -.15  .8837     -.11845    .10200
        |Standard Deviation of u: Sigma(u) then scale
Sigmau_0|     .00241        9.18604      .00  .9998   -18.00191  18.00673
LOADFCTR|    -.56559        3.85231     -.15  .8833    -8.11597   6.98479
  LSTAGE|    -.00823         .05624     -.15  .8837     -.11845    .10200
        |Standard deviation of v
Sigma(v)|     .12361         .08711     1.42  .1559     -.04713    .29435
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Limited Dependent Variable Model - FRONTIER
Dependent variable                    LC
Log likelihood function       173.52520
Estimation based on N =    256, K =  15
Variances: Sigma-squared(v)=     .01334
           Sigma-squared(u)=     .00121
           Sigma(v)        =     .11551
           Sigma(u)        =     .03476
Sigma = Sqr[(s^2(u)+s^2(v)]=     .19230
Stochastic Frontier Scaling Model
Mean scale factor for E[u] =        .3459
Mean scale factor for V[u] =        .2261
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  2
Deg. freedom for truncation mean:     2
Deg. freedom for inefficiency model:  5
LogL when sigma(u)=0          157.91523
Chi-sq=2*[LogL(SF)-LogL(LS)] =   31.220
Kodde-Palm C*: 95%:10.371,  99%: 14.325
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
      LC|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    11.6452       24.94703      .47  .6406    -37.2501   60.5405
      LY|     .94078***      .02140    43.97  .0000      .89884    .98272
     LY2|     .06680***      .01579     4.23  .0000      .03585    .09776
    LPKP|     .85146        1.94378      .44  .6614    -2.95828   4.66120
    LPLP|     .16345**       .07956     2.05  .0399      .00751    .31939
    LPMP|     .25417        1.26886      .20  .8412    -2.23275   2.74109
    LPEP|    -.34167         .62932     -.54  .5872    -1.57511    .89178
    LPFP|     .00164         .02164      .08  .9395     -.04078    .04406
        |Mean of Truncated Distribution, Mu then scale
    Mu_0|    1.92288***      .44030     4.37  .0000     1.05991   2.78584
LOADFCTR|   -1.74305        4.08382     -.43  .6695    -9.74720   6.26110
  LSTAGE|    -.01930         .04649     -.42  .6781     -.11042    .07182
        |Standard Deviation of u: Sigma(u) then scale
Sigmau_0|     .15374        1.11571      .14  .8904    -2.03301   2.34049
LOADFCTR|   -14.5014       10.21457    -1.42  .1557    -34.5216    5.5188
  LSTAGE|    1.02454        1.26499      .81  .4180    -1.45479   3.50388
        |Standard deviation of v
Sigma(v)|     .11551***      .00793    14.56  .0000      .09996    .13106
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```