

Stochastic Frontier Models and Economic Efficiency Estimation

William Greene

Stern School of Business, New York University

Assignment 3

Stochastic Frontier Production and Cost Models

Part I. Production and Cost

This part of assignment 3 will involve fitting stochastic frontier models with production and cost function frameworks. Use the electricity generation data, Christensen-Greene.lpj

1. In the C&G study, it was found that the preferred functional form for the cost model was the full translog model

$$(1) \quad \log(C/PF) = \alpha + \sum_{j=k,l} \beta_j \log(P_j/PF) + \sum_{i=k,l} \sum_{j=k,l} \gamma_{ij} \log(P_j/PF) \log(P_j/PF) \\ + \theta_q \log(Q) + \theta_{qq} \frac{1}{2} \log^2(Q) + \sum_{j=k,l} \lambda_j \log(P_j/PF) \log(Q) + \varepsilon.$$

(We have imposed the restriction of linear homogeneity in the input prices). The stochastic frontiers literature contains at least seven studies of the Christensen and Greene data (that I have found), all of which are based on the restricted (hybrid) model

$$(2) \quad \log(C/PF) = \alpha + \sum_{j=k,l} \beta_j \log(P_j/PF) + \theta_q \log(Q) + \theta_{qq} \frac{1}{2} \log^2(Q) + \varepsilon$$

Christensen and Greene tested the hypothesis of this restriction in the context of the multivariate regression model (cost function plus two of the factor shares). You can test it again using the likelihood ratio test with the stochastic frontier model. You can do this easily with the electricity data already set up. The restricted and unrestricted models are specified with these two commands. You can use the likelihood ratio test to test for the restrictions of the homothetic cost function. (How many degrees of freedom are there for the test?)

```
FRONTIER ; Cost ; Lhs = logc_pf ; Rhs = costcd,logqsq $
CALC ; lcd = logl $
FRONTIER ; Cost ; Lhs = logc_pf ; Rhs = costtlog $
CALC ; ltl = logl ; list ; chisq = 2*(ltl - lcd) $
```

What do you find? In fact, for purposes of the stochastic cost frontier model, the translog function is overspecified. What happens when you try to fit a stochastic frontier translog model? This may provide some insight into why researchers have always used model (2) above, instead of (1). [Hint: watch for the dreaded “wrong skewness” message.] Note, however, the MLEs are still reported, and you can do the likelihood ratio test of translog vs. hybrid, in spite of the adverse outcome for the translog model. Note, finally, Christensen and Greene rejected the restrictions, but your results will not. The reason is that C&G used FIML in a 3 equation model, whereas you are only using the cost function. They used more information.

2. Use the restricted model, and fit a cost frontier

```
FRONTIER ; Cost ; Lhs = logc_pf ; Rhs = costcd,logqsq ; Costeff = euicost $
DSTAT ; Rhs = euicost $
```

Save the resulting inefficiencies as variable *euicost* by adding `;CostEFF=euicost` to the command. Use `DSTAT` to examine the inefficiencies. What did you find?

Stochastic Frontier Models and Economic Efficiency Estimation
William Greene
Stern School of Business, New York University

3. The electricity data also contain the variables for the production side. Use the data on **logQ** (Lhs variable) and **one,logcap,loglabor,logfuel** (Rhs variables) to fit a stochastic frontier production model. Use **;techeff = euiprod** to save the estimates of inefficiency. Do they resemble their counterparts from the cost function? Use

```
FRONTEIR ; Lhs = logq ; Rhs = one,logcap,loglabor,logfuel ; techeff = euiprod $  
CALC ; List ; Cor (euiprod, eucost) $  
DSTAT ; Rhs = euiprod, eucost $
```

You can also plot them against each other,

```
PLOT ; Lhs = euiprod ; Rhs = eucost ; Rh2 = euiprod $
```

(The **;Rh2 = euiprod** setting above plots euiprod against itself, and connects the dots – this produces a 45 degree line which helps in the comparison.) The ranks of the firms are also interesting. You can use

```
CREATE ; rankp = Rnk (euiprod) $  
CREATE ; rankc = Rnk (eucost) $  
PLOT ; Lhs = rankp ; Rhs = rankc $
```

what do you find? A correlation coefficient for ranks is the Kendall rank correlation measure,

```
CALC ; List ; Rkc (rankp, rankc) $
```

4. Functional form matters. Though the translog function seems to be overspecified for the frontier model, it appears that the Cobb-Douglas model is underspecified. Try fitting the model with

```
FRONTIER ; Cost ; Lhs = logc_pf ; Rhs = costcd ; Maxit = 60 $
```

Notice that the estimation breaks down. The problem is that the economies of scale in these data run out at a fairly moderate output, while the Cobb-Douglas model assumes that they just continue for as large as output becomes. This appears to be a serious misspecification of the model. In practical terms, it is the difference between a linear function (in logs) vs. a quadratic function. By restricting the number of iterations, we can see what the failure looks like. Note in the results, the estimate of λ is huge, and the estimate of σ_v is getting close to zero. This means the estimator is essentially fitting a deterministic frontier model.

5. There is some evidence that a better model for the deterministic part of the frontier helps to produce better estimates of the inefficiencies. An interesting newly developed method suggests a nonparametric approach for the frontier function. For a cost function, this would appear

$$\log(C/P_f) = g(\text{output, prices}) + v + u$$

where g is a nonparametric approximation to the true cost (or production) frontier, and $v+u$ is our familiar half normal stochastic element of the SF model. Let's explore this with a nonparametric

Stochastic Frontier Models and Economic Efficiency Estimation
William Greene
Stern School of Business, New York University

frontier function and see how it compares to the more familiar approach. We will do the following:

? Estimate SF model with expanded Cobb-Douglas function.

**FRONTIER ; Cost ; Lhs = logc_pf
; Rhs = costcd,logqsq ; Eff = up \$**

? Estimate SF model with nonparametric cost function

**FRONTIER ; Cost ; Lhs = logc_pf
; Rhs = costcd,logqsq ; Lowess ; Eff = unip ; List \$**

? Plot estimated inefficiencies from nonparametric model vs. those from parametric.

**PLOT ; Lhs = up ; Rhs = unip ; Rh2 = up ; Grid
; Title=Inefficiencies: Nonparametric vs. Parametric Models\$**

Stochastic Frontier Models and Economic Efficiency Estimation

William Greene

Stern School of Business, New York University

Stochastic Frontier Models with Heterogeneity

Part II. Heterogeneity in a Stochastic Frontier Production Model

In this assignment, you will build a stochastic frontier model that incorporates heterogeneity. We have two data sets that you can use for this exercise, the WHO data on health care attainment and the airlines data. The assignment requests you to build a stochastic frontier model. Keep in mind that the objective is to study the efficiency, $\exp(-u_i)$, in the data. So, with each **FRONTIER** command, you can use

; Techeff = name for the saved exp(-u(i))

to keep the computed values. Then, you can explore the data with

KERNEL ; Rhs = the variable \$

and/or

DSTAT ; Rhs = variables \$

and, perhaps

PLOT ; Lhs = variable on horizontal axis

; Rhs = variable on vertical axis \$

You can add a 45⁰ line to the figure to help the comparison as follows:

PLOT ; Lhs = variable on horizontal axis

; Rhs = variable on vertical axis

; Rh2 = the same variable on the horizontal axis \$

1. We will build a cross section model using the 1997 data. We must eliminate the other four years of data from the sample and also the observations on internal political units such as the states of Mexico and the provinces of Canada. After loading the WHO project file, use this command to set the sample to be the 1997 observation for the 191 countries in the sample.

REJECT ; SMALL > 0 | Year # 1997 \$

You will also want to use logs of the variables in the model. The logs of the two attainment variables are LCOMP (log of the composite measure) and LDALE (log of life expectancy). The variables on the Rhs of the frontier model will be **x = one,lhexp,lhexp2,leduc** = a constant, log health expenditure, its square, and log of education. This is the base model. The other variables in the data set, that you can use to craft a frontier model are

GINI	= income distribution
GEFF	= government effectiveness
VOICE	= degree of democratization
TROPICS	= location, in tropics or not
LPOPDEN	= log of population density

Stochastic Frontier Models and Economic Efficiency Estimation
William Greene
Stern School of Business, New York University

LPUBTHE = log of public expenditure
 LGDPC = log of per capita GDP

Heterogeneity can enter the frontier model in (at least) four places:

- a. In the frontier function itself – add the variable to the **RHS**
- b. In the mean of the underlying distribution of u_i – use

;MODEL=Truncation ; RH2 = the variables. (Include ONE in this list)

- c. In the variance of u_i – use **;HFU = the list of variables.** (Do not include ONE)
- d. In the variance of v_i – use **;HFV = the list of variables.** (Do not include ONE)

Use the WHO data to design and estimate a model. Where do you find is the preferred location for the listed variables in the model? Notes:

- a. The truncation model does not work very well for these data.
- b. You can use different variables in different locations, for example, including LGDPC, TROPICS directly in the function and GINI in the mean of u_i .
- c. A variable can appear more than once in the model, such as in the function itself and in the variance of u_i .

The specifications of the frontier model differ greatly, and it is unclear how the environmental variables actually affect the efficiency results. Here is an experiment that might be helpful. In the first specification, we put the z 's in the mean of a truncated normal model. In the second, we put the z 's in the variance of a half normal model. With each specification, we compute the average partial effects of the environmental variables on the efficiency estimates. Note that (1) In both cases, the coefficients have the opposite signs of the derivatives and (2) the estimates are reasonably similar.

? How to interpret the coefficients on the environmental variables?

?

**FRONTIER ; quietly ; lhs = lcomp ; rhs = x ; model=t;rh2=one,z ; techeff=eutrunc\$
 PARTIALS ; effects: z ; summary \$
 FRONTIER ; quietly ; lhs = lcomp ; rhs = x ; het;hfu=z ; techeff=uhet_u\$
 PARTIALS ; effects: z ; summary \$**

Note, finally, in our setup of the model, the specification with heteroscedasticity in v breaks down completely.

Stochastic Frontier Models and Economic Efficiency Estimation
William Greene
Stern School of Business, New York University

For our specifications, we used

```
REJECT ; small > 0 | Year # 1997 $
NAMELIST ; x = one,lhexp,lhexp2,leduc $
NAMELIST ; z = gini,lgdpc,lpopden $
FRONTIER ; Lhs = lcomp ; Rhs = x,z ; techeff = euprod $
FRONTIER ; Lhs = lcomp ; Rhs = x ; Model=t;rh2=one,z ;techeff=eutrunc $
FRONTIER ; Lhs = lcomp ; Rhs = x ; Het;hfu=z ;techeff=euhet_u $
FRONTIER ; Lhs = lcomp ; Rhs = x ; Het;hfv=z ;techeff=euhet_v $
DSTAT ; Rhs = eu* $
KERNEL ; Rhs=euprod,eutrunc,euhet_u $
HISTOGRAM ;Rhs=euprod
;Title=Estimated Technical Efficiency - z in Production Function $
? How to interpret the coefficients on the environmental variables?
?
FRONTIER ; quietly ; lhs = lcomp ; rhs = x ;model=t;rh2=one,z ; techeff=eutrunc$
PARTIALS ; effects: z ; summary $
FRONTIER ; quietly ; lhs = lcomp ; rhs = x ;het;hfu=z ; techeff=euhet_u$
PARTIALS ; effects: z ; summary $
```


Stochastic Frontier Models and Economic Efficiency Estimation
William Greene
Stern School of Business, New York University

Part IV. Multiple Output Technology

Use the project file `banksdata.lpj` for this exercise. Be sure to exit NLOGIT and restart before you load these data. Then, also reload `Assignment3.lim`.

1. Multiple Outputs

The banking data correspond to a multiple output technology. There are five outputs and five inputs (described in the file). Begin this exercise by fitting a Cobb-Douglas style COST model in which the LHS is **C** and the RHS is **one,w1,w2,w3,w4,q1,q2,q3,q4,q5**. The linear homogeneity restrictions are built into the model. As usual, use `;CostEff = eui` to save the estimated efficiencies, and examine your results. What do you find? Now, since this is a panel, lets see if there are time effects in the data. Add T and T2 to the RHS of the model. Are the coefficients significant? Include `;Costeff = euit` in the model command. To examine the time effects, regress eui (from the model without the time trends) on one,t,t2. What do you find? How does including T and T2 in the equation differ from regression of the estimated UIs on T and T2 in a second step?

```
FRONTIER ; Cost ; Lhs = c ; Rhs=one,inputs,outputs ; Costeff = eui $  
FRONTIER ; Cost ; Lhs = c ; Rhs=one,inputs,outputs,t,t2 ; Costeff = euit $  
REGRESS ; Lhs = Eui ; Rhs = one,t,t2 $  
DSTAT ; Rhs = eui,euit $  
PLOT ; Lhs = eui ; Rhs = euit ; Rh2 = eui $
```

2. Latent Class Models

A number of experiments have been carried out analyzing the banking industry. (Kumbhakar, et al.) Here, we will fit a few latent class models to examine parameter heterogeneity. The first is a simple two class model. The LC model exploits the panel nature of this data set. The kernel estimator shows the difference in the efficiency estimates across the classes.

```
NAMELIST ; outputs=q1,q2,q3,q4,q5 $  
NAMELIST ; inputs=w1,w2,w3,w4 $  
FRONTIER ; cost ; lhs=c ; rhs = one,outputs,inputs $  
FRONTIER ; cost ; lhs=c ; rhs = one,outputs,inputs  
; lcm ; pds=5 ; pts = 2 ; Group=class ; techeff = eui $  
DSTAT ; rhs=eui ; str=class $  
KERNEL ; rhs=eui;group=class $
```

Stochastic Frontier Models and Economic Efficiency Estimation
William Greene
Stern School of Business, New York University

We now explore the Kumbhakar and Parmeter suggestion, that some banks might not be inefficient, at least not all the time. The command forces λ to equal zero in one of the classes. The technology parameters are the same in both classes. The variance parameters change across the classes.

```
FRONTIER ; cost ; lhs=c ; rhs = one,outputs,inputs $  
FRONTIER ; cost ; lhs=c ; rhs = one,outputs,inputs  
; lcm ; pds=5 ; pts = 2  
; Rst = 10_b,sigma1,lambda1,10_b,sigma2,0,pi1,pi2 $  
CALC ; loglsame=logl $
```

The final specification of the LC frontier allows the technology parameters to change across the classes, while again forcing $\lambda = 0$ in one of the classes. A likelihood ratio test rejects the assumption that the parameters are the same.

```
FRONTIER ; cost ; lhs=c ; rhs = one,outputs,inputs $  
FRONTIER ; cost ; lhs=c ; rhs = one,outputs,inputs  
; lcm ; pds=5 ; pts = 2  
; Rst = 10_b,sigma1,lambda1,10_c,sigma2,0,pi1,pi2 $  
CALC ; logldiff=logl $  
CALC ; list ; lrstat = 2*(logldiff-loglsame)$
```