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

Lab Session 2 Assignment
Stochastic Frontier Models

This assignment is based on the airlines data contained in airlines.lpj and described in airlines.lim and airlines.doc. Though they are a panel of 25 firms observed over varying numbers of years, for present purposes, we will treat them as a cross section.

1. Although the maximum likelihood estimator of the normal – half normal stochastic frontier model is now standard, researchers still use the corrected least squares and modified least squares estimators. One interpretation is that the frontier is deterministic – there is no ‘noise’ component, only inefficiency. This is consistent with the deterministic frontier model

\[ y(i) = \alpha + \beta'x(i) - u(i) \]

a. The corrected least squares (COLS) estimator is computed by fitting the frontier model by ordinary least squares, then computing \( u(i) = [\text{maximum } e(i)] - e(i) \) where \( e(i) \) is the OLS residuals. These corrected residuals provide the interesting information about inefficiency in the data. Using the airlines data, let \( X = \text{ONE,LF,LE,LL,LP} \) (i.e., omit materials), fit the production function model (\( LQ \) is the log of the output variable). Compute the residuals. (You can just add \textbf{;RES=E} to your \textbf{REGRESS} command. Now, compute the COLS residuals with

\[
\text{CALC ; Maxe = Max(e) } \\
\text{CREATE ; ui = Maxe – e }
\]

Use \textbf{DSTAT} to obtain the mean inefficiency suggested by this method. (Note, data are in logs, so the mean is roughly the percentage. For example, if the sample mean is 0.30, this suggests 30% inefficiency on average.) Are the results plausible? Look at the minimum and maximum as well. This is all automated. You can use

\[
\text{NAMELIST ; x = one,lf,le,ll,lp } \\
\text{FRONTIER ; Lhs = lq ; Rhs = x ; Model = cols }
\]

Give these commands. Note the results provided for analyzing the frontier function

b. The modified least squares (MOLS) estimator is also a deterministic frontier estimator (at least at this point). In this computation, we ‘estimate’ \( u(i) \) with

\[
\text{Est}.u(i) = \text{Est}[E[e(i)]] - e(i).
\]

In order to compute the expectation, we need to assume a distribution. (Note by construction, the OLS residuals have mean zero, so the sample mean won’t help.) Suppose we assume the distribution of \( u(i) \) is exponential with parameter \( \theta \). Then, the sample standard deviation of the OLS residuals is an estimator of \( \theta \), and our estimates of \( u(i) \) can be \( \theta - e(i) \). You can obtain the results using
CALC ; theta = sdv(e) $
CREATE ; ui = theta – e $

Now use DSTAT to examine your results. What is the estimate of average inefficiency. Note that the minimum value is negative, implying that at least one firm/year is ‘superefficient.’ This is an unfortunate shortcoming of this method of estimating the model. What is at work even more than the estimation method is the assumption of a deterministic frontier model. The MOLS method can also be applied to the stochastic frontier model (see Section 2.4.8 in the Greene survey). Since maximum likelihood estimation is so easy, and is more efficient, we will leave these corrected and modified OLS estimators at this point.

2. Still using the airlines data, fit a stochastic frontier model. Retain the Cobb-Douglas assumption $X = ONE,LF,LE,LL,LP$. Examine your results. Are the coefficient estimates plausible? What are the estimates of $\sigma_v$ and $\sigma_u$. Which appears to be the greater source of variation in log output, inefficiency, $u$, or noise, $v$? What is the log likelihood function for the frontier model? Now, fit the same model by ordinary least squares – this assumes $\sigma_u = 0$. Use a likelihood ratio test to assess the presence of inefficiency in the data.

```
FRONTIER ; Lhs = LQ ; RHS = ONE,LF,LE,LL,LP $
REGRESS ; Lhs = LQ ; RHS = ONE,LF,LE,LL,LP $
```

3. Using the model in problem 2 but with an exponential model, obtain the JLMS inefficiency estimates. This is done with

```
FRONTIER ; Lhs = LQ ; RHS = ONE,LF,LE,LL,LP ; Model=E 
    ; TechEFF = eui $
```

Now, use a kernel estimator to examine the estimated inefficiency distribution. What is the sample estimate of the expected value of $u$? You can use DSTAT to obtain this.

```
KERNEL ; Rhs = eui,eucols $
```

How do you explain the large difference between the deterministic measure of efficiency and the stochastic frontier estimate of efficiency?

4. The airlines data contain three variables that might (emphasize might) help to explain the inefficiency in the data, LOAD FACTOR, STAGE – this is the average length of flights, and POINTS which is the number of nodes in the route. There are two (more, as we will see tomorrow) ways to examine this proposition. First, you obtained estimates of $u(i)$ when you computed the model in 3. You could now regress these on a constant and the three factors noted. Does there appear to be a relationship between $u(i)$ and these variables? A second, indirect way to deal with this possibility is to include the three variables in the model as additional RHS variables. Fit the model with LOADFCTR, POINTS and LSTAGE (log of stage length) included, and save the inefficiencies as UIA. (That is, as a new variable.) Now, compare UI to UIA. You can use a correlation coefficient

```
CALC ; List ; Cor (UI, UIA) $
```

or you could regress UI on a constant and UIA. Or, you could plot UI against UIA using
PLOT ; LHS = UI ; RHS = UIA $

What do you find.  Are the two sets of estimates the same?

5.  (A vexing problem for stochastic frontier modelers)  Now, fit the stochastic frontier model in part 4, but include LM the model, as well as ONE,LL,LF,LE,LP.  What happened?  Can you explain the outcome?  What should you do next?

6.  Comparing SF and DEA.  Data envelopment is an alternative method of analyzing efficiency. Though we are not spending much class time on this topic, you can get a quick look at how it works, and compare it to the stochastic frontier approach we have been studying. The mechanics of DEA are described in the extract from the LIMDEP manual that is included in your course materials.  DEA produces two measures of efficiency, input oriented and output oriented.  LIMDEP computes both of these for you.  To obtain these for the airlines data, use to obtain the measure from the stochastic frontier model,

FRONTIER ; Lhs = LQ ; Rhs = X ; eff= uisf $
CREATE ; euisf=exp(-uisf)$

then

FRONTIER ; Lhs = output ; Rhs = fuel.eqpt,labor,prop ; alg=dea $
?  Pick up the input oriented efficiency from the computation
CREATE ; euidea=deaeff_i$

to obtain the DEA results. Notice that the analysis is based on the levels of the variables. We are not fitting a production function, we are analyzing inputs and outputs. The command produces a summary and computes the two variables that we want to analyze. We will first compare the input oriented measure to what was produced by the stochastic fronteir. The command file for this assignment discusses this further. Three comparisons are provided by

DSTAT ; Rhs = euisf.euidea $
PLOT ; Lhs=euidea ; Rhs=uisf$
CALC ; List ; Cor(euidea,uisf)$

As an additional exercise, repeat the computations using the output oriented measure, DEAEFF_O.