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

Assignment 4
Panel Data Stochastic Frontier Models

Four of the data sets for the course – all save the Christensen and Greene data – are panels. The WHO data have been used in many efficiency studies, both as a panel and, essentially as a cross section. (There is very little within groups variation in the WHO data – they actually are rather close to a cross section.) This exercise set will briefly examine a number of specification issues, one with each panel data set.

Part I. Banking Data

1. Standard panel modeling. The banking data are a five year panel on costs. We will fit pooled data models and panel data models based on the standard techniques, and examine the results. The frontier models will all be based on the core command

```
SETPANEL   ; Group = bank ; pds = Ti $
FRONTIER       ; LHS = C ; Cost
               ; RHS = ONE,W1,W2,W3,W4,Q1,Q2,Q3,Q4,Q5 $
```

a. **Cross section model:** As above. Add **; EFF = UICS** to the command.

```
FRONTIER       ; LHS = C ; Cost
               ; RHS = ONE,W1,W2,W3,W4,Q1,Q2,Q3,Q4,Q5
               ; EFF = UICS              ; Techeff=SFTE $
```

b. **Pitt and Lee's random effects model**. Add **; Panel ; EFF = UIRE**. How do the results change? What do you find?

```
FRONTIER       ; LHS = C ; Cost
               ; RHS = ONE,W1,W2,W3,W4,Q1,Q2,Q3,Q4,Q5
               ; Panel ; EFF = UIRE      ; Techeff=PLTE $
```

c. **Schmidt and Sickles, fixed effects model**: This requires a couple steps.

```
REGRESS        ; LHS = C ; RHS = ONE,W1,W2,W3,W4,Q1,Q2,Q3,Q4,Q5
               ; PANEL ; FIXED ; PARAMETERS $
CREATE         ; AI = ALPHAFE (Bank) $
CALC           ; MINAI = Min(AI) $
CREA           ; UIFE = AI - MINAI       ; CSTE=1-UIFE$
```

(The regression command creates a vector containin the estimated fixed effects. The CREATE command replicates the value once for each year, for each firm.) Use DSTAT to describe these three sets of estimates. What do you find. An important question is whether the RE and FE estimates are similar. Are they? Plot one against the other and report your finding. You can use to make the scatter diagram a little more attractive. Now, (one at a time) plot the two sets of panel data estimates against the cross section values. The main difference is that the panel data variables are the same in each period. The cross section version is allowed to change from one period to the next. The next several exercises examine this issue.

```
PLOT        ; LHS = UIFE ; RHS = UIRE ; RH2 = UIFE
            ; LIMITS=0,.3;endpoints=0,1$
KERNEL      ; Rhs = SFTE,PLTE,CSTE
            ; Grid
            ; Title=Pooled, Fixed and Random Effects Estimators$
```

The following are a variety of specifications for the banking data.

```
?======================================================
? Analysis of different panel models for banking data.
? Random parameters and latent class
?======================================================
    SETPANEL    ;group=bank ; pds=ti $
    FRONTIER    ;cost;lhs=c;rhs=one,linearw,linearq$
    FRONTIER    ;cost;lhs=c;rhs=one,linearw,linearq;panel
                ;rpm;fcn=one(n),w1(n),w2(n),w3(n),w4(n);pts=10;halton
                ;maxit=10;par;eff=u_rpm$
    MATRIX      ;rpmbeta=beta_i $
    FRONTIER    ;cost;lhs=c;rhs=one,linearw,linearq$
    FRONTIER    ;cost;lhs=c;rhs=one,linearw,linearq;panel
                ;lcm;pts=3;par
                ;eff=u_lcm $
    MATRIX      ;lcmbeta=beta_i$
```

Notice the value of λ in class 1 is -.00000053018. This indicates that firms in class 1, whichever they are, do not appear to be efficient. We will investigate this formally.

```
    FRONTIER    ;cost;lhs=c;rhs=one,linearw,linearq$
    FRONTIER    ;cost;lhs=c;rhs=one,linearw,linearq;panel
                ;lcm;pts=3;par ;eff=u_lcm ; maxit=25
                ;rst=10_b,sa,0,10_b,sb,lb,10_b,sd,ld,pi1,pi2,pi3$

?======================================================
? Create scale variable and namelist for RHS of frontier
?======================================================
    CREATE      ;q=log(exp(q1)+exp(q2)+exp(q3)+exp(q4)+exp(q5))$
    CREATE      ;qq=.5*q*q$
    NAMELIST    ; X = one,q,qq,w1,w2,w3,w4$

?======================================================
? Base case
?======================================================
    FRONTIER    ;cost;lhs=c;rhs=x;eff=u_sf$
    FRONTIER    ;cost;lhs=c;rhs=x;eff=uire;panel$
    FRONTIER    ;cost;model=t;lhs=c;rhs=x;eff=u_trunc;panel$
    PLOT        ;lhs=u_trunc;rhs=uire;rh2=u_trunc$
    KERNEL      ;rhs=u_sf,u_trunc;endpoints=0,1$
    KERNEL      ;For[t = 1] ; rhs=uire,u_trunc$
    PLOT        ;lhs=u_sf;rhs=u_rpm;endpoints=0,1
            ;Title=Inefficiencies: Pooled vs. Random Parameters Model$
    PLOT        ;lhs=u_rpm;rhs=u_lcm;limits=0,.8
            ;Title=Inefficiencies: Latent Class vs. Random Parameters Model$
    PLOT        ;lhs=u_sf;rhs=U_lcm ; rh2=u_sf
            ;Title=Inefficiencies: Pooled vs. Latent Class Model;limits=0,.8$
```

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

```
?===============================================================
? True fixed effects
?===============================================================
        FRONTIER    ;cost;lhs=c;rhs=x$
        FRONTIER    ;cost;lhs=c;rhs=x;panel;fem;eff=u_truefe; maxit=5$
        PLOT        ;lhs=u_sf;rhs=u_truefe ; rh2=u_sf $

?===============================================================
? True random effects
? True random effects is identical to base case.
?===============================================================
        FRONTIER    ;cost;lhs=c;rhs=x$
        FRONTIER    ;cost;lhs=c;rhs=x;panel;rpm;fcn=one(n)
                    ;halton;pts=20 ; eff=u_truere $

?==========================================================
? Time varying and time invariant inefficiency terms in the
? same model.  Most general true random effects in use.
?==========================================================
        FRONTIER    ; cost ; Lhs = c ; Rhs = x $
        FRONTIER    ; cost ; Lhs = c ; Rhs = x ; Panel
                    ; CSN ; Pts=100 ; Halton ; Maxit=25 $; List $

?===============================================================
? Pitt and Lee random effects
?===============================================================
        FRONTIER    ;cost;lhs=c;rhs=x;panel;eff=u_re$
        PLOT        ;lhs=u_re;rhs=u_sf;rh2=u_re$
        MATRIX      ;meanusf=gxbr(u_sf,bank)$
        MATRIX      ;meanure=gxbr(u_re,bank)$
        MPLOT       ;lhs=meanusf;rhs=meanure$
        REGRESS     ;lhs=u_sf;rhs=one;panel$
        DSTAT       ;rhs=u_truere,u_re,u_sf$
        PLOT        ;lhs=u_truere;rhs=u_re ; Rh2=u_truere$

?===============================================================
? Schmidt and Sickles fixed effects
?===============================================================
        REGRESS     ;lhs=c;rhs=x;panel;fixed$
        CREATE      ;u_fe=alphafe(bank)$
        CALC        ;minufe=min(u_fe)$
        CREATE      ;u_fe=u_fe-minufe$
        PLOT        ;lhs=u_re;rhs=u_fe;rh2=u_re$
        PLOT        ;lhs=u_truefe;rhs=u_fe;rh2=u_truefe$
        MATRIX      ;xbtfe=gxbr(u_truefe,bank)$
        CREATE      ;meantfe=xbtfe(bank)$
        PLOT        ;lhs=meantfe;rhs=u_fe;rh2=meantfe;limits=0,.8$$
```

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

```
?============================================================
? Latent class
?============================================================
        FRONTIER    ;cost;lhs=c;rhs=x$
        FRONTIER    ;cost;lhs=c;rhs=x;lcm;panel;pts=3;eff=u_lcm$
        PLOT        ;lhs=u_sf;rhs=u_lcm ; rh2=u_sf $

?============================================================
? Random Parameters Model - answer to the Bayesians
?============================================================
        FRONTIER    ;cost;lhs=c;rhs=x$
        FRONTIER    ;cost;lhs=c;rhs=x;rpm
                    ;fcn=one(n),w1(n),w2(n),w3(n),w4(n),q(n),qq(n)
                    ;halton;pts=20;pds=5;maxit=15;cor;par $ ;eff=u_rpm$
        PLOT        ;lhs=u_lcm;rhs=u_rpm;endpoints=0,.8;Rh2=u_lcm$
        DSTAT       ;rhs=u*$
        MATRIX      ;list;xcor(u_sf,u_trunc,u_truere,u_truefe,u_re,u_fe)$
```

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

Part II. Spanish Dairy Farm Data

Using the dairy farming data, compare the various random and fixed effects estimators. Note that the distinction in these panel data models is between models that assume inefficiency is time invariant (Pitt and Lee random effects, Schmidt and Sickles fixed effects) and models that allow the inefficiency to vary through time, and assume that the time invariant effect is heterogeneity that is unrelated to inefficiency (Greene's true fixed and random effects models.) Commands you can use to fit these models are

```
         SETPANEL          ; Group = Farm ; Pds = Ti $
         NAMELIST          ; CD = ONE,X1,X2,X3,X4 $
? 1.  Pooled, cross section
         FRONTIER          ; LHS = Yit ; RHS = CD ;eff=uics$
? Fixed effects, time invariant inefficiency (Schmidt and Sickles)
         REGRESS           ; LHS = Yit ; RHS = CD
                           ;  PANEL ; FIXED ; PARAMETERS $
         CREATE            ; AI = ALPHAFE (farm) $
         CALC              ; MaxAI = Max(AI) $
         CREA              ; UIFE = maxai-AI$
? Fixed effects, time varying (True fixed effects – does not work well)
         FRONTIER          ; LHS = Yit ; RHS = CD $
         FRONTIER          ; LHS = Yit ; RHS = CD ; PANEL ; FEM ; EFF = UITFE $
? Random effects, time invariant – Pitt and Lee
         FRONTIER          ; LHS = Yit ; RHS = CD ; PANEL ; EFF = UIRE $
? Random effects, time varying True Random Effects
         FRONTIER          ; LHS = Yit ; RHS = CD $
         FRONTIER          ; LHS = Yit ; RHS = CD ; PANEL ; RPM ; FCN = One(N)
                           ; PTS = 50 ; HALTON ; EFF = UITRE $
? Random effects with Mundlak's treatment
         CREATE            ; X1B = Group Mean (X1,Str=Farm) $
         CREATE            ; X2B = Group Mean (X2, Str=Farm) $
         CREATE            ; X3B = Group Mean (X3, Str=Farm) $
         CREATE            ; X4B = Group Mean (X4, Str=Farm) $
         NAMELIST          ; Mundlak = X1B,X2B,X3B,X4B $
         FRONTIER          ; LHS = Yit ; RHS = CD,Mundlak ; PANEL ; EFF = UIREM $
         FRONTIER          ; LHS = Yit ; RHS = CD,Mundlak $
         FRONTIER          ; LHS = Yit ; RHS = CD,Mundlak ; PANEL ; RPM
                           ; FCN = One(N) ; PTS = 50 ; HALTON ; EFF = UITREM $
         FRONTIER          ; Lhs = yit ; Rhs = CD,Mundlak $
         FRONTIER          ; Lhs = yit ; Rhs = CD,Mundlak
                           ; Panel ; Csn ; Pts=100 ; Halton ; Maxit=25 ; List $
```

Descriptive statistics for the estimates of $E[u_i|\varepsilon_i]$ will be useful.

```
         DSTAT        ; RHS = UI* $
```

You should also plot the estimates – scatter plots of UIRE vs. UIFE and UITRE vs UIFRE will demonstrate the similarity of the fixed and random effects approaches. On the other hand, if you plot UIFE vs. UITFE or UIRE vs. UITRE, you can see how the time varying and time invariant models produce very different results.

```
?
? Display of four different sets of estimates.
         PLOT  ; Lhs = UiFE ; Rhs = UITFE
```

```
      ; Title=Fixed Effects vs. True Fixed Effects$
PLOT  ; Lhs = UiRE ; Rhs = UITRE
      ; Title=Random Effects vs. True Random Effects$
PLOT  ; Lhs = UIRE ; Rhs = UIFE ; Rh2 = uire
      ; Vaxis=True Fixed Effects
      ; Title=Random Effects vs. Fixed Effects$
PLOT  ; Lhs = UITRE ; Rhs = UITFE ; Rh2 = UITRE
      ; Title=True Random Effects vs. True Fixed Effects$
```

Battese and Coelli Models

The preceding exercise using the dairy farming data, examines two extreme assumptions about technical inefficiency. In the 'fixed effects' and 'random effects' models, it is assumed that inefficiency is fixed through time, regardless of how many periods are observed. In the 'true fixed effects' and 'true random effects' models, it is assumed that inefficiency varies completely freely from one period to the next. Much of the literature on panel data models has examined intermediate cases, in which the inefficiency term was of a form more or less like

$$u_{it} = g(t) \times |U_i|$$

in which inefficiency varies through time, but in a somewhat restricted fashion. A popular form is the Battese and Coelli model,

$$u_{it} = \text{Exp}[\eta(t-T)] \times |U_i|$$

where t is the period, T is the last period, and note that the stochastic part is U_i which is time invariant. Thus, in this form, there is a patterned variation through time, a simple exponential function determined by the parameter η. Variants on this model will involve different g(.) functions involving data (heterogeneity), perhaps different functions of time, and so on. For example, we will consider the model

$$u_{it} = \exp[\ \eta_1(t-T) + \eta_2(t-T)^2] \times |U_i|$$

which makes the model somewhat more general. An interesting case is the model

$$u_{it} = \exp[\Sigma_t \eta_t d_t] \times |U_i|$$

in which the d_ts are time dummy variables and η_t are dummy variable coefficients. In this case, there is a separate coefficient for each period (one must be normalized at 0). But, this is not the same as time varying inefficiency, since U_i is still time invariant. It is interesting to compare these different cases. We will use the dairy farming data again. The commands you need to compute the model coefficients and the estimated inefficiencies are listed below. Estimate the models, then compare the different estimates of u_i. Use DSTAT, PLOT, and compute correlations. You can get a correlation matrix by adding ;OUTPUT=2 to your DSTAT command. A plot of UICS (freely varying inefficiency) against UIBCTIME (time dummy variables) should be interesting. Some basic commands are as follows:

```
? Battese and Coelli, base case
```

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

```
?
FRONTIER   ; LHS = YIT ; RHS = CD ; PANEL
           ; MODEL = BC  ; EFF=UIBC$
?
? Battese and Coelli, separate year effects
?
FRONTIER   ; LHS = YIT ; RHS = CD ; PANEL ; MODEL = BC
           ; HFU = YEAR93,YEAR94,YEAR95,YEAR96,YEAR97
           ; EFF = UIBCTIME$
?
? Battese and Coelli, quadratic term in time
?
CREATE     ; T = YEAR – 93  ; T2 = T^2 $
FRONTIER   ; LHS = YIT ; RHS = CD ; PANEL
           ; MODEL = BC  ; HFU = T,T2 ; EFF = UIBC2 $
?
? Comparison.  Reveals that inefficiency estimates from the second and third
? specifications are identical.
?
KERNEL     ; rhs = uibc*$
```

Stochastic Frontier Models and Economic Efficiency Estimation

William Greene
Stern School of Business, New York University

Stochastic Frontier Model for Swiss Railways

The Swiss railway data set contains a panel of data on production costs for 51 Swiss railway companies. The data were used in "Efficiency measurement in network industries: Application to the swiss railway companies" by Mehdi Farsi, Massimo Filippini and William Greene. The variables of interest for your study (there are a number of variables we will not use) are

ID	= firm ID
NI	= number of periods observed. Repeated
T	= time period, begins at 0
LNCT	= log of total cost/electricity price
LNPK	= log of capital price/electricity price
LNPL	= log of labor price/electricity price
MLNPK	= railway mean of LNPK – repeated for each year
MLNPL	= railway mean of LNPL – repeated for each year
LNQ2	= log of passenger output
LNQ3	= log of freight (goods) output
MLNQ1,2,3	= railway means of logs of outputs. Q1 is total output. for Mundlak formulation, use LNQ2,LNQ3,MLNQ1 and MLNQ3 (not MLNQ2)
NARROW_T	= dummy variable for narrow track
RACK	= dummy variable for a network type. See SwissRailways.lim
VIRAGE	= dummy variable for curvature. See SwissRailways.lim
TUNNEL	= dummy variable for network with long tunnels
LNSTOP	= log of number of stations
MLNSTOP	= railway mean of log of number of stops – repeated for each year
LNNET	= log of network length
MLNNET	= railway mean of log of network length – repeated for each year

Using these data, build a stochastic cost frontier model for swiss railways. Estimate technical efficiency, $E[\exp(-u_i)]$ either by railroad if you are using a panel model with time invariant inefficiency, or by railway-year if you are using a model with time varying inefficiency. Report your results for your model estimation and estimates of efficiency.