NYU STERN
NEW YORK UNIVERSITY LEONARD N. STERN SCHOOL OF BUSINESS

Department of Economics

Econometric Analysis of Panel Data

Assignment 4
Parameter Heterogeneity in Linear Models: RPM and HLM

The estimation parts of this assignment will be based on the Baltagi and Griffin gasoline market and the Cornwell and Rupert labor market data sets that are posted on the course website.

We will begin with the gasoline market.The basic linear regression model in use will be

$$y_{it} = \beta_1 + \beta_2 x_{it,1} + \beta_3 x_{it,2} + \beta_4 x_{it,3} + w_{it}$$

where

and i = 1,…,18 OECD countries

t = 1,…,19 years (1960 to 1978).

y_{it} = *lgaspcar* = log of per capita gasoline use

$x_{it,1}$ = *lincomep* = log of per capita income

$x_{it,2}$ = *lrpmg* = log of gasoline price index

$x_{it,3}$ = *lcarpcap* = log of cars per capita

w_{it} = a disturbance that may have have both permanent (time invariant) components and time varying components, and may, under some circumstances, be correlated with $\mathbf{x}_{it}$.

Denote $\mathbf{x}_{it}$ = $(1, x_{it,1}, x_{it,2}, x_{it,3})$ and

$\mathbf{X}_i$ = the 19×4 matrix containing all the data on $\mathbf{x}_{it}$ for country i.

Part I. Parameter Variation in the Gasoline Market

A. Homogeneous parameters: To begin, we *assume* that all parameters, including the constant term, are homogeneous across countries and through time and that $w_{it} = \varepsilon_{it}$, a classical zero mean, homoscedastic disturbances.

1. Under these assumptions, what are the properties of the pooled OLS estimator?
2. Estimate the parameters of the model using OLS and report your results.
3. As a first cut at assessing whether the assumptions are correct, compute the robust, cluster (country) corrected standard errors for the least squares estimator. Do they appear to be the same, or close to the same, as the uncorrected OLS standard errors? What do you conclude about the disturbances in the equation?

B. Heterogeneous Constant Terms: Now, consider fixed and random effects formulations of the model. We write the model as

$$y_{it} = \beta_{1i} + \beta_2 x_{it,1} + \beta_3 x_{it,2} + \beta_4 x_{it,3} + \varepsilon_{it}$$

where

$$\beta_{1i} = \beta_1 + u_i \text{ and } \mathrm{E}[u_i] = 0,\ i = 1,\ldots,17.$$

Thus, this is a model with a random constant term. By substituting the second equation into the first, you can see that it is the "effects" model we have discussed in class.

1. (**Fixed Effects**) Using the OECD gasoline data, estimate the parameters of the model under the assumption that $E[u_i|\mathbf{X}_i] = g(\mathbf{X}_i)$ for some nonzero function $g(.)$. Explain the estimator and the motivation for using it. Display your results with the OLS estimates so that you (and your reader) can see the difference between the two. Note that $E[u_i|\mathbf{X}_i] = g(\mathbf{X}_i)$ is still consistent with $E[u_i] = 0$. When averaged over $\mathbf{X}_i$, the overall mean is zero, but the mean is not zero for a specific $\mathbf{X}_i$. This implies that u_i and $\mathbf{X}_i$ are correlated.
2. (**Random Effects**) Estimate the parameters of the model under the more restrictive assumption that $\mathrm{E}[u_i|\mathbf{X}_i] = 0$.
3. Use the Wu/Mundlak variable addition test to test for the assumption of the (null) random effects model against the (alternative) fixed effects model. Report your results and your conclusions. (Recall, the Wu test is based on adding the group (country) means of the regressors to the model and testing down to the REM without the group means.)

C. General parameter heterogeneity: Let $\mathbf{x}_{it}$ denote $(1, lincomep, lrpmg, lcarpcap)_{it}$. We now consider the possibility that there are differences across countries. Write the model

$$(1) \qquad y_{it} = \boldsymbol{\beta}_i'\mathbf{x}_{it} + \varepsilon_{it}.$$

Absent any further assumptions about the variation in the parameters across countries, how would you proceed to examine the relationship between per capita gasoline consumption, y_{it} and the other variables, $\mathbf{x}_{it}$?

1. Suppose we now assume that all the parameters, not just the constant, are random;

(2) $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{u}_i$

where $\mathbf{u}_i$ has an overall mean of zero, however $E[\mathbf{u}_i|\mathbf{X}_i] = g(\mathbf{X}_i)$, where $\mathbf{X}_i$ is the 19 years of data on $\mathbf{x}_{it}$ for country i? Note that the assumption of the overall mean of zero states only that $\boldsymbol{\beta}_i$ varies around a mean.

We are particularly interested in the price elasticity of the demand for gasoline, the coefficient on *lrpmg*. To explore the cross country variation, compute the linear regression model for each of the 19 countries (separately). Display in a graph or a well labeled table the results of your estimation, to describe the variation in the estimated coefficients on *lrpmg*. Note that the assumption about $\mathbf{u}_i$ is equivalent to the "fixed effects" case, but here we are considering the entire parameter vector, not just the constant term.

2. If we add to A. the assumption $E[\mathbf{u}_i|\mathbf{X}_i] = 0$, the model turns into a 'random effects' model, though note, once again, we are considering the entire parameter vector. Under this new assumption, what are the properties of the pooled ordinary least squares estimator? What does $\mathbf{b}$ estimate in this case? For a useful step in the analysis, insert (2) into (1), expand, and analyze the implied model.

Part II. Theory and an Example for Simulation Based Estimation:

This theoretical exercise will begin to suggest how simulation based estimation works. Consider a simple regression model

$$y_{it} = \beta_i x_{it} + \varepsilon_{it}$$

There is only one variable and no constant in the model. Assume that $\varepsilon_{it} \sim N[0,\sigma^2]$. We suppose as well that β_i is random; $\beta_i = \beta + w_i$ where $w_i \sim N[0,\theta^2]$. A useful way to write this is

$$\beta_i = \beta + \theta u_i \text{ where } u_i \sim N[0,1].$$

Putting θ specifically in the equation simplifies the derivation a bit. The contribution of individual i to the likelihood function is the product of the normal densities,

$$L_i|u_i = \prod_{t=1}^{T} \frac{1}{\sigma}\phi\left(\frac{y_{it} - \beta_i x_{it}}{\sigma}\right) \text{ (Note, } u_i \text{ is in } \beta_i.)$$

This is not useable for maximum likelihood estimation because $\beta_i = \beta + \theta u_i$ which means that the log likelihood to be maximized involves the unobserved u_i;

$$L_i|u_i = \prod_{t=1}^{T} \frac{1}{\sigma}\phi\left(\frac{y_{it} - \beta x_{it} - \theta u_i x_{it}}{\sigma}\right)$$

In principle, we would now maximize $\log L|\mathbf{u} = \Sigma_i \log L_i|u_i$ with respect to (β,θ,σ). The problem is that the unobserved u_i is still in the equation and must be integrated out to proceed.. The contribution of individual i to the *unconditional* log likelihood function is

$$\log L_i = \log \int_{-\infty}^{\infty} \left[\prod_{t=1}^{T} \frac{1}{\sigma} \phi\left(\frac{y_{it} - \beta x_{it} - \theta u_i x_{it}}{\sigma} \right) \right] \phi(u_i) du_i$$

where $\phi(u_i)$ is the standard normal density. The integral of the product above does not exist in closed form, so we will approximate it by simulation. (It could be approximated with quadrature.) Adding up the individual contributions, the *simulated* log likelihood is

$$\log L_S = \sum_{i=1}^{n} \log \left\{ \frac{1}{R} \sum_{r=1}^{R} \left[\prod_{t=1}^{T} \frac{1}{\sigma} \phi\left(\frac{y_{it} - \beta x_{it} - \theta u_{ir} x_{it}}{\sigma} \right) \right] \right\}$$

where u_{ir} is a set of R random draws on the standard normal population for each individual i. (The same random draws are reused every time the function or its derivatives are computed. There are a total of nR random draws used in the simulation.) An additional simplification is obtained by using $\gamma = 1/\sigma$. (We make use of the invariance principle for maximum likelihood estimation.) Then,

$$\log L_S = \sum_{i=1}^{n} \log \left\{ \frac{1}{R} \sum_{r=1}^{R} \left[\prod_{t=1}^{T} \gamma \phi\left(\gamma \left(y_{it} - \beta x_{it} - \theta u_{ir} x_{it} \right) \right) \right] \right\}.$$

The maximum simulated likelihood estimator is the (β,θ,γ) that maximizes this function.

1. Derive the necessary (first order) conditions for maximizing this function. Hint: your derivation is simplified greatly by using the result $d\phi(t)/dt = -t\phi(t)$. You can then just use the chain rule.
2. How would you obtain asymptotic standard errors for your estimator?
3. The following small exercise will show this computation at work. This application estimates the parameters of a model that precisely satisfies the assumptions of the model above. Execute these commands and report all of your results

? 1,000 observations in total will be n=100, T=10. The x(i,t) is normally distributed
? with mean zero and standard deviation 1. Variable i is the 1,1,1,1...,2,2,2,2... etc.

```
Sample ; 1 - 1000 $
Create ; xit = Rnn(0,1) ; i = Trn(10,0) $
```

? We generate b(i) = 0.5 + u(i) where u(i) is normal with mean 0, standard deviation
? .5. Then, y(i,t) = b(i)*x(i,t) + e(i,t) where e(i,t) is normally distributed (0,1).

```
Matrix ; bi = init(100,1,.5) + .5*rndm(100)$
Create ; yit = bi(i)*xit + rnn(0,1) $
```

? This command estimates the random parameters model exactly as shown in
? part 2. above.

```
Regress ; lhs = yit ; rhs = xit ;rpm ; fcn=xit(n)
        ; pds=10 ; pts=100 ; halton $
```

Part III. Random Parameters Models

This exercise will demonstrate the computation of a fairly elaborate, hierarchical linear model. The computations are based on the Cornwell and Rupert data. Note that the simulations below are based on Halton sequences, not pseudorandom random numbers. As such, the results you obtain below are replicable – in principle, you and I (and your colleagues) should all get the same results. Also, if you fit these equations more than once, you will get the same answers.

1. A simple RPM with one random coefficient. The first model is the regression model discussed in class, now with a random coefficient on education. After fitting the random parameters model, this program computes the posterior estimates of $E[\beta_{i,Ed}|y_i,X_i]$ and plots the distribution with a kernel density estimator and a histogram. Estimate the model and report all results. (Note, you can copy/paste the figure into a Word document.)

```
Sample ; All $
Regress ; Lhs = Lwage ;Rhs = One,Exp,Occ,Ind,South,SMSA,MS,FEM,Union,Ed
;Pds=7 ;RPM ; Halton ; Pts=100 ;Fcn = Ed(N) ;Parameters $
Sample ; 1 - 595 $
Create ; Ed_Coeff = 0 $
Create ; Ed_Coeff = beta_i $
Kernel ; Rhs = Ed_Coeff$
Histogram ; Rhs = Ed_Coeff $
```

2. The second model is a typical hierarchical model. The model is

$$y_{it} = \beta_{1,i} + \beta_{2,i}Exp_{it} + \beta_3 OCC_{it} + ... + \beta_8 Union_{it} + \varepsilon_{it}$$
$$\beta_{1,i} = \beta_1 + \delta_{1,1}Fem_i + \delta_{1,2}Ed_i + u_{1i},$$
$$\beta_{2,i} = \beta_2 + \delta_{2,1}Fem_i + \delta_{2,2}Ed_i + u_{2i},$$

This is a common sort of model in which the regression of interst is based on the time varying attributes and the variation in the parameters is explained by the randomness, $u_{k,i}$ and by the demographics that do not vary across time, here Gender, Education and Race. Fit the model and report all results, identifying what parameter is what in your report.

```
Sample ; All $
Regress ; Lhs = Lwage
;Rhs = One,Exp,Occ,Ind,South,SMSA,MS,Union
;Pds=7
;RPM=Fem,Ed
;Halton ; Pts=100
;Fcn = one(n),exp(n)
;Parameters $
```

3. Construct a different random parameters specification, modify the command above accordingly and fit your model. Report your results and interpret the estimates you obtain. (Note in formulating an RPM, the variables at the 'upper' level do not vary through time. Thus, it will be inappropriate to include time variables in the ;RPM list.)