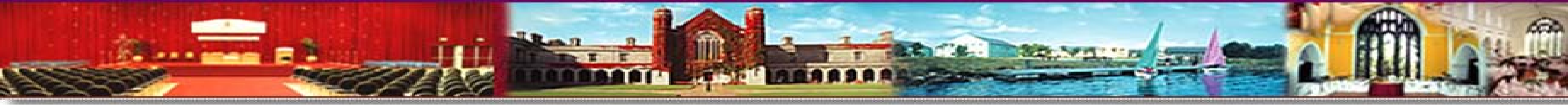
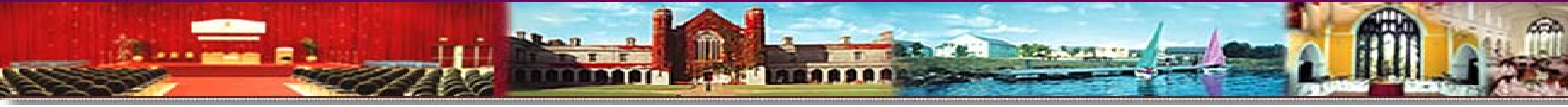
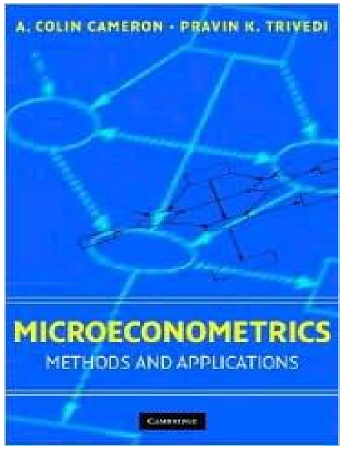
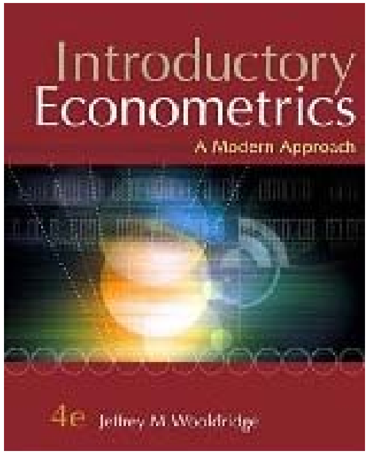


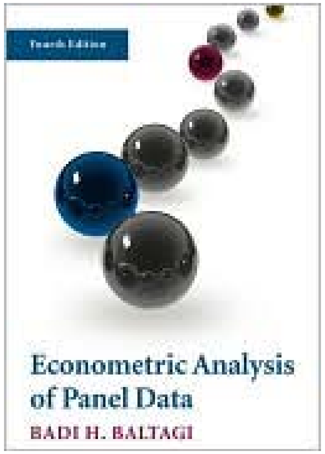
Topics in Microeconometrics

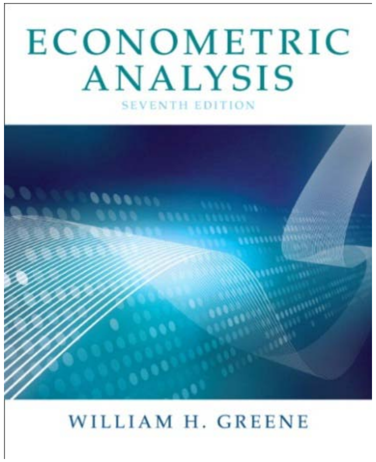
William Greene
Department of Economics
Stern School of Business



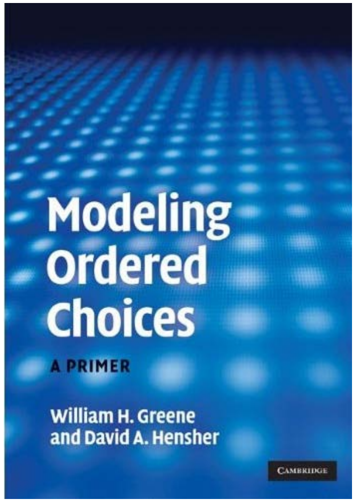


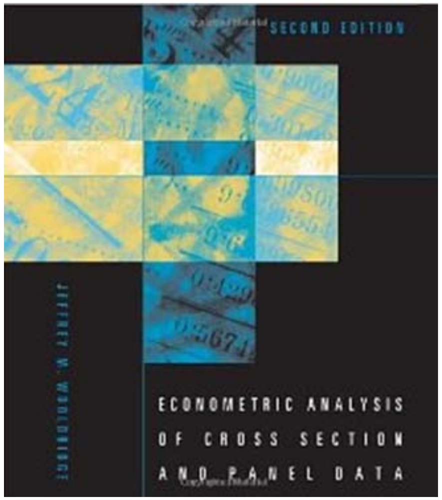


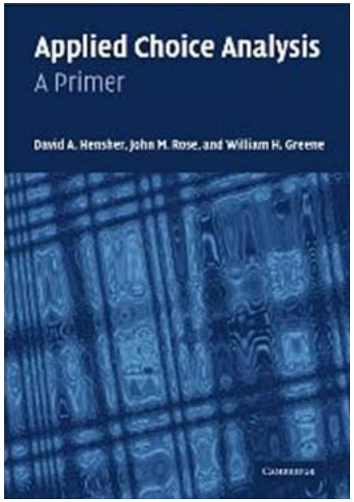


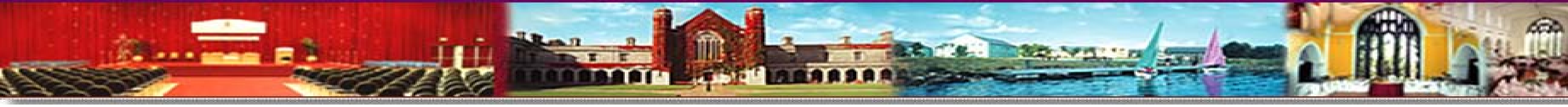


8. Random Parameters and Hierarchical Linear Models









About CIC   Penn World Table   About ICP   Research Papers   Contact Us

- About CIC
- Penn World Table
- About ICP
- Research Papers
- Contact Us

Announcements

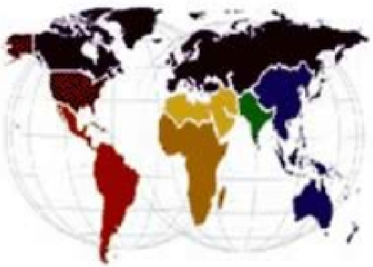
PWT 7.0

Last updated on June 3, 2011!

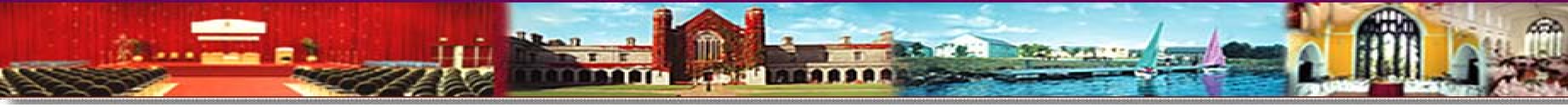
And follow us on twitter for new announcement.

Follow @PennWorldTable

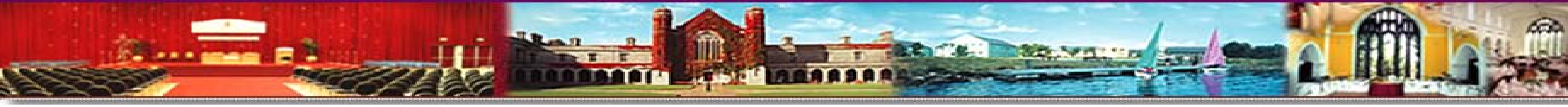
Center for International Comparisons at the University of Pennsylvania



Center for International Comparisons of Production, Income and Prices
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
(215) 898-7624



- **PWT 7.0** (189 countries and territories, 1950-2009, 2005 as reference year, **last updated: June 3, 2011**)
  - Data Download (online database which provides html, csv, sas format downloads for selected countries, years)
  - Complete Data Download (zipped file, excel spreadsheets, all the countries, all the years)
  - about the variables in PWT7.0
  - Description of PWT 7.0
  - Brief Note on the Tables for PWT 7.0
  - National account data for PWT 7.0
    Variables in the national account data

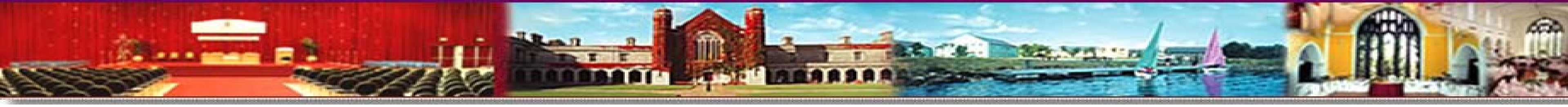


Heterogeneous Dynamic Model

$$\log Y_{i,t} = \mu_i + \lambda_i \log Y_{i,t-1} + \theta_i x_{it} + \varepsilon_{i,t}$$

Long run effect of interest is $\beta_i = \dfrac{\theta_i}{1-\lambda_i}$

Average (over countries) effect: $\bar{\beta}$ or $\dfrac{\bar{\theta}}{1-\bar{\lambda}}$



“Fixed Effects” Approach

$$\log Y_{i,t} = \mu_i + \lambda_i \log Y_{i,t-1} + \theta_i x_{it} + \varepsilon_{i,t}$$

$$\beta_i = \frac{\theta_i}{1-\lambda_i},\ \ \overline{\beta} = \frac{\overline{\theta}}{1-\overline{\lambda}}$$

(1) Separate regressions; $\hat{\mu}_i, \hat{\lambda}_i, \hat{\theta}_i$

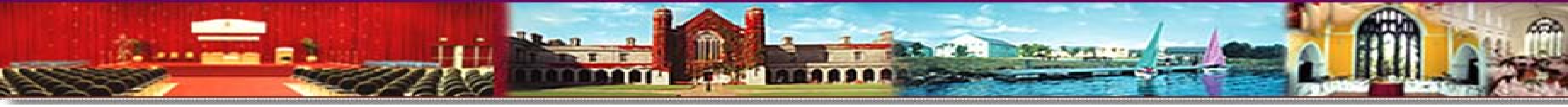
(2) Average estimates = $\overline{\hat{\beta}} = \frac{1}{N}\Sigma_{i=1}^{N} \frac{\hat{\theta}_i}{1-\hat{\lambda}_i} = \frac{1}{N}\Sigma_{i=1}^{N}\hat{\beta}_i$ or

Function of averages: $\hat{\overline{\beta}} = \frac{(1/N)\Sigma_{i=1}^{N}\hat{\theta}_i}{1-(1/N)\Sigma_{i=1}^{N}\hat{\lambda}_i}$

In each case, each term i has variance $O(1/T_i)$

Each average has variance $O(1/N) \times \sum_{i=1}^{N}(1/N)O(1/T_i)$

Expect consistency of estimates of long run effects.



A Mixed/Fixed Approach

$$\log y_{i,t} = \mu_i + \Sigma_{i=1}^{N} \lambda_i d_{i,t} \log y_{i,t-1} + \theta_i x_{i,t} + \varepsilon_{i,t}$$

$d_{i,t}$ = country specific dummy variable.

Treat μ_i and θ_i as random, λ_i is a 'fixed effect.'

This model can be fit consistently by OLS and efficiently by GLS.

A Mixed Fixed Model Estimator

$$\log y_{i,t} = \mu_i + \Sigma_{i=1}^{N} \lambda_i d_{i,t} \log y_{i,t-1} + \theta_i x_{i,t} + \varepsilon_{i,t}$$

$$\theta_i = \theta + w_i$$

$$\log y_{i,t} = \mu_i + \Sigma_{i=1}^{N} \lambda_i d_{i,t} \log y_{i,t-1} + \theta_i x_{i,t} + (w_i x_{i,t} + \varepsilon_{i,t})$$

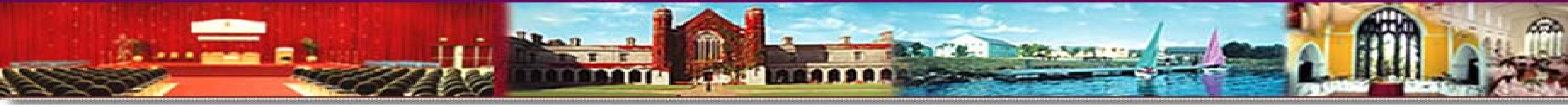
Heteroscedastic : $\text{Var}[w_i x_{i,t} + \varepsilon_{i,t}] = \sigma_w^2 x_{i,t}^2 + \sigma_\varepsilon^2$

Use two step least squares.

(1) Linear regression of $\log y_{i,t}$ on dummy variables, dummy variables times $\log y_{i,t-1}$ and $x_{i,t}$.

(2) Regress squares of OLS residuals on $x_{i,t}^2$ and 1 to estimate σ_w^2 and σ_ε^2.

(3) Return to (1) but now use weighted least squares.

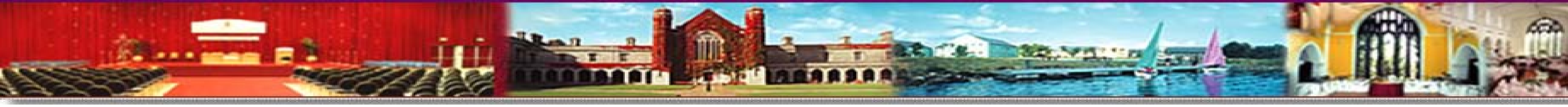


Baltagi and Griffin's Gasoline Data

World Gasoline Demand Data, 18 OECD Countries, 19 years
Variables in the file are

COUNTRY = name of country
YEAR = year, 1960-1978
LGASPCAR = log of consumption per car
LINCOMEP = log of per capita income
LRPMG = log of real price of gasoline
LCARPCAP = log of per capita number of cars

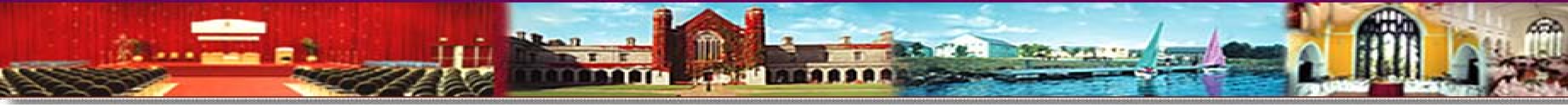
See Baltagi (2001, p. 24) for analysis of these data. The article on which the analysis is based is Baltagi, B. and Griffin, J., "Gasoline Demand in the OECD: An Application of Pooling and Testing Procedures," European Economic Review, 22, 1983, pp. 117-137. The data were downloaded from the website for Baltagi's text.



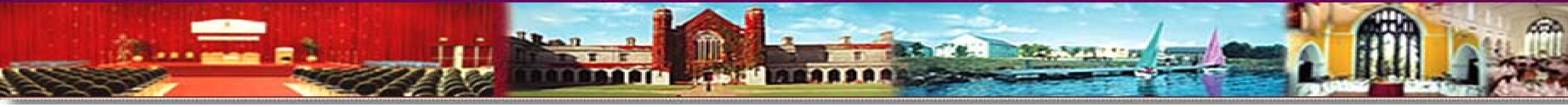
Baltagi and Griffin’s Gasoline Market

COUNTRY = name of country		
YEAR = year, 1960-1978		
LGASPCAR = log of consumption per car		y
LINCOMEP = log of per capita income		z
LRPMG = log of real price of gasoline		x1
LCARPCAP = log of per capita number of cars		x2

$$y_{it} = \beta_{1i} + \beta_{2i}\, z_{it} + \beta_{3i}\, x1_{it} + \beta_{4i}\, x2_{it} + \varepsilon_{it}.$$



FIXED EFFECTS



Parameter Heterogeneity

Unobserved Effects $\equiv$ **Random Constants**

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + \varepsilon_{it}$$

$$y_{it} = \alpha_i + \mathbf{x}_{it}\boldsymbol{\beta} + \varepsilon_{it}$$

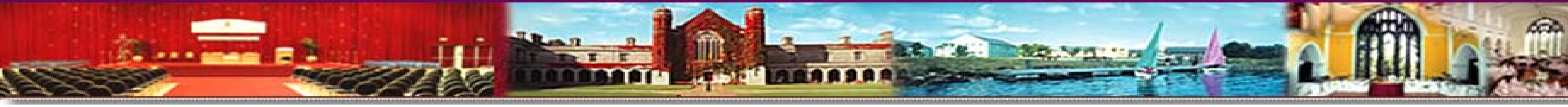
$$\alpha_i = \alpha + u_i,$$

$$E[u_i \mid \mathbf{X}_i] = 0 \text{ --> Random effects}$$

$$E[u_i \mid \mathbf{X}_i] \neq 0 \text{ --> Fixed effects}$$

$$E_X E[u_i \mid \mathbf{X}_i] = 0.$$

$Var[u_i \mid \mathbf{X}_i]$ not yet defined - so far, constant.



Parameter Heterogeneity

Generalize to Random Slopes

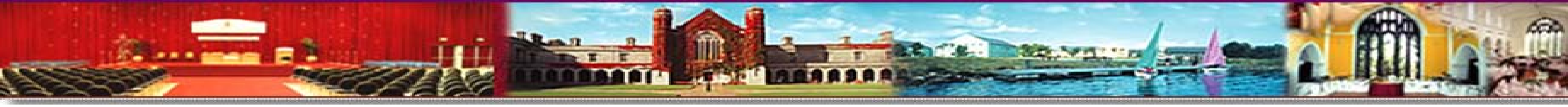
$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta}_{i} + \varepsilon_{it}$

$\boldsymbol{\beta}_{i} = \boldsymbol{\beta} + \mathbf{u}_{i}$

$E[\mathbf{u}_i \mid \mathbf{X}_i]$ zero or nonzero - to be defined

$E_X[E[\mathbf{u}_i \mid \mathbf{X}_i]] = \mathbf{0}$

$Var[\mathbf{u}_i \mid \mathbf{X}_i]$ to be defined, constant or variable



Fixed Effects (Hildreth, Houck, Hsiao, Swamy)

$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta}_i + \varepsilon_{it}$, each observation

$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i$, T_i observations

$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{u}_i$

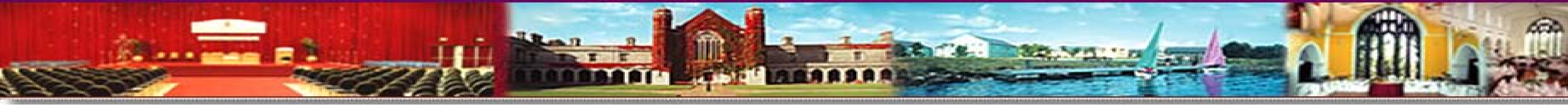
Assume (temporarily) $T_i > K$.

$E[\mathbf{u}_i \mid \mathbf{X}_i] = g(\mathbf{X}_i)$ (conditional mean)

$P[\mathbf{u}_i \mid \mathbf{X}_i] = (\mathbf{X}_i - E[\mathbf{X}_i])\boldsymbol{\theta}$ (projection)

$E_X[E[\mathbf{u}_i \mid \mathbf{X}_i]] = E_X[P[\mathbf{u}_i \mid \mathbf{X}_i]] = \mathbf{0}$

$Var[\mathbf{u}_i \mid \mathbf{X}_i] = \boldsymbol{\Gamma}$ constant but nonzero



OLS and GLS Are Inconsistent

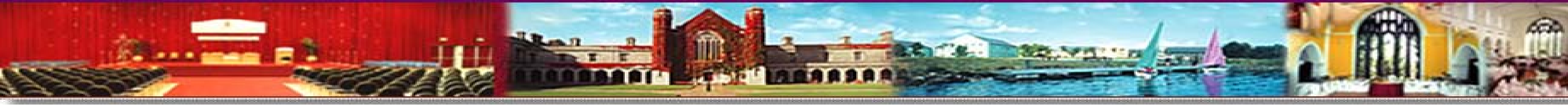
$$\mathbf{y_i} = \mathbf{X_i}\boldsymbol{\beta} + \boldsymbol{\varepsilon_i},\ T_i \text{ observations}$$

$$\boldsymbol{\beta_i} = \boldsymbol{\beta} + \mathbf{u_i}$$

$$\mathbf{y_i} = \mathbf{X_i}\boldsymbol{\beta} + \mathbf{X_i}\mathbf{u_i} + \boldsymbol{\varepsilon_i},\ T_i \text{ observations}$$

$$\mathbf{y_i} = \mathbf{X_i}\boldsymbol{\beta} + \mathbf{w_i}$$

$$E[\mathbf{w_i} \mid \mathbf{X_i}] = \mathbf{X_i}E[\mathbf{u_i} \mid \mathbf{X_i}] + E[\boldsymbol{\varepsilon_i} \mid \mathbf{X_i}] \neq$$

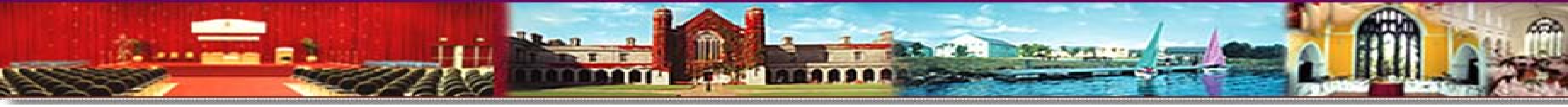


Estimating the Fixed Effects Model

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{...} \\ \mathbf{y}_N \end{pmatrix} = \begin{bmatrix} \mathbf{X}\boldsymbol{\beta}_1 & \mathbf{0}\ \boldsymbol{\varepsilon} & \mathbf{...} & \mathbf{0} \\ \boldsymbol{\beta} & \mathbf{X}_2\ \boldsymbol{\varepsilon} & \mathbf{...} & \mathbf{0} \\ \mathbf{...} & \mathbf{...} & \mathbf{...} & \mathbf{...} \\ \boldsymbol{\beta} & \mathbf{0}\ \boldsymbol{\varepsilon} & \mathbf{...} & \mathbf{X}_N \end{bmatrix} \begin{pmatrix} _1 \\ _2 \\ \mathbf{...} \\ _N \end{pmatrix} + \begin{pmatrix} _1 \\ _2 \\ \mathbf{...} \\ _N \end{pmatrix}$$

Estimator: Equation by equation OLS or (F)GLS

Estimate $\boldsymbol{\beta}$? $\frac{1}{N}\Sigma_{i=1}^{N}\hat{\boldsymbol{\beta}}_i$ is consistent in N for $E[\boldsymbol{\beta}_i]$.



Random Effects and Random Parameters

THE Random Parameters Model

$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta}_{\mathbf{i}} + \varepsilon_{it}$, each observation

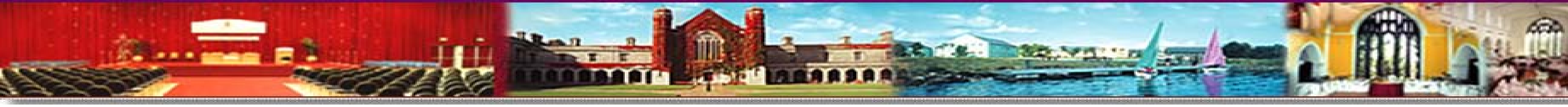
$\mathbf{y}_{\mathbf{i}} = \mathbf{X}_i\boldsymbol{\beta}_{\mathbf{i}} + \boldsymbol{\varepsilon}_{\mathbf{i}}$, T_i observations

$\boldsymbol{\beta}_{\mathbf{i}} = \boldsymbol{\beta} + \mathbf{u}_i$

Assume (temporarily) $T_i > K$.

$E[\mathbf{u}_i \mid \mathbf{X}_i] = \mathbf{0}$

$Var[\mathbf{u}_i \mid \mathbf{X}_i] =$ constant but nonzero



Estimating the Random Parameters Model

$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$, T_i observations

$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{u}_i$

$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i$, T_i observations

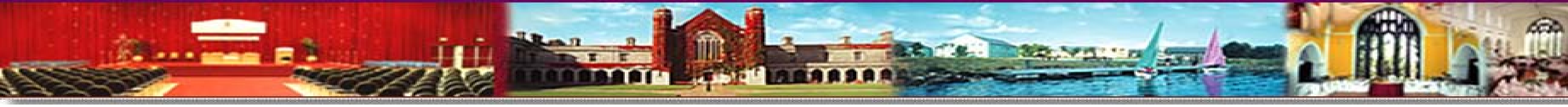
$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{w}_i$

$E[\mathbf{w}_i \mid \mathbf{X}_i] = \mathbf{X}_i E[\mathbf{u}_i \mid \mathbf{X}_i] + E[\boldsymbol{\varepsilon}_i \mid \mathbf{X}_i] = \mathbf{0}$

$\text{Var}[\mathbf{w}_i \mid \mathbf{X}_i] = \mathbf{X}_i\boldsymbol{\Gamma}\mathbf{X}_i' + \sigma^2_{\varepsilon,i}\mathbf{I} = \boldsymbol{\Sigma}_i$ Should $\sigma^2_{\varepsilon,i}$ vary by i?

Objects of estimation : $\boldsymbol{\beta}, \sigma^2_{\varepsilon,i}, \boldsymbol{\Gamma}$

Second level estimation : $\boldsymbol{\beta}_i$



Estimating the Random Parameters Model by OLS

$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$, T_i observations

$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{u}_i$

$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i$, T_i observations

$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{w}_i$

$$\mathbf{b} = [\Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i]^{-1}[\Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{y}_i]$$

$$= \boldsymbol{\beta} + [\Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i]^{-1}[\Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{w}_i]$$

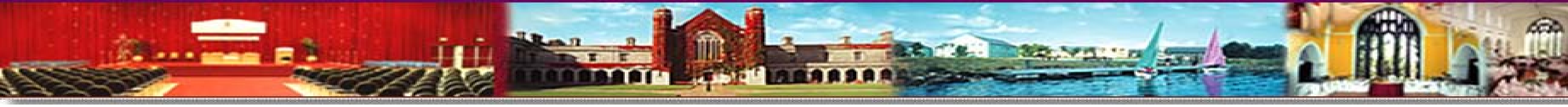
$$\mathbf{Var}[\mathbf{b}\,|\,\mathbf{X}] = [\Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i]^{-1}[\Sigma_{i=1}^{N}\mathbf{X}_i'(\mathbf{X}_i\boldsymbol{\Gamma}\mathbf{X}_i' + \sigma_\varepsilon^2\mathbf{I})\mathbf{X}_i][\Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i]^{-1}$$

$$= \sigma_\varepsilon^2[\Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i]^{-1} + [\Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i]^{-1}[\Sigma_{i=1}^{N}(\mathbf{X}_i'\mathbf{X}_i)\boldsymbol{\Gamma}(\mathbf{X}_i'\mathbf{X}_i)][\Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i]^{-1}$$

= the usual + the variation due to the random parameters

Robust estimator

$$\text{Est.Var}[\mathbf{b}] = [\Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i]^{-1}[\Sigma_{i=1}^{N}\mathbf{X}_i'\hat{\mathbf{w}}_i\hat{\mathbf{w}}_i'\mathbf{X}_i][\Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i]^{-1}$$



Estimating the Random Parameters Model by GLS

$\mathbf{y_i} = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i$, T_i observations

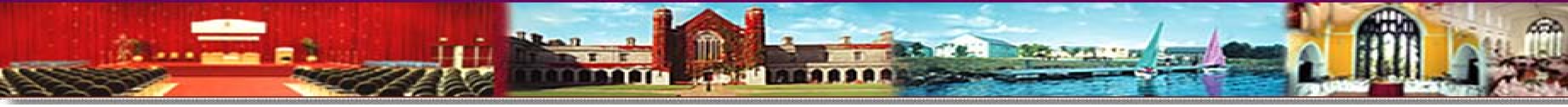
$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{u}_i$

$\mathbf{y_i} = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i$, T_i observations

$\mathbf{y_i} = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{w}_i$, $\text{Var}[\mathbf{w}_i \mid \mathbf{X}_i] = \boldsymbol{\Omega}_i = (\mathbf{X}_i\boldsymbol{\Gamma}\mathbf{X}_i' + \sigma^2_{\varepsilon,i}\mathbf{I})$

$\hat{\boldsymbol{\beta}} = [\Sigma_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\Omega}_i^{-1}\mathbf{X}_i]^{-1}[\Sigma_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\Omega}_i^{-1}\mathbf{y}_i]$

For FGLS, we need $\hat{\boldsymbol{\Gamma}}$ and $\hat{\sigma}^2_{\varepsilon,i}$.



Estimating the RPM

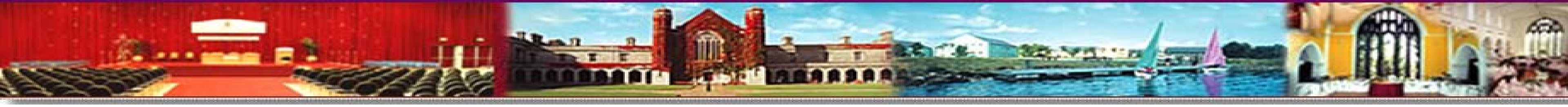
$$\mathbf{b}_i = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{y}_i = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'[\mathbf{X}_i(\boldsymbol{\beta} + \mathbf{u}_i) + \boldsymbol{\varepsilon}_i]$$

$$= \boldsymbol{\beta} + \mathbf{u}_i + (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\boldsymbol{\varepsilon}_i$$

$$\text{Var}[\mathbf{b}_i|\mathbf{X}_i] = \boldsymbol{\Gamma} + \sigma^2_{\varepsilon,i}(\mathbf{X}_i'\mathbf{X}_i)^{-1}$$

$$\hat{\sigma}^2_{\varepsilon,i} = \frac{\sum_{t=1}^{T_i}(y_{it} - \mathbf{x}_{it}'\mathbf{b}_i)^2}{T_i - K} \text{ is unbiased}$$

(but not consistent because T_i is fixed).



An Estimator for Γ

$$E[\mathbf{b_i}|\mathbf{X_i}] = \boldsymbol{\beta}$$

$$Var[\mathbf{b_i}|\mathbf{X_i}] = \boldsymbol{\Gamma} + \sigma^2_{\varepsilon,i}(\mathbf{X_i}'\mathbf{X_i})^{-1}$$

$$Var[\mathbf{b_i}] = Var_X E[\mathbf{b_i}|\mathbf{X_i}] \;+\; E_X Var[\mathbf{b_i}|\mathbf{X_i}]$$

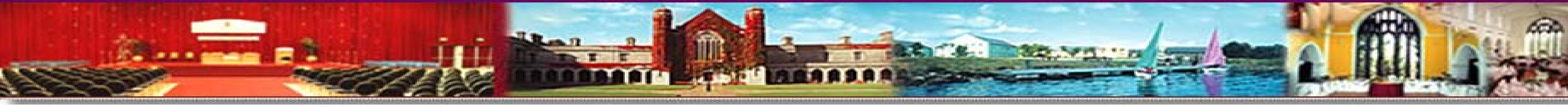
$$= \quad 0 \quad + \; E_X[\boldsymbol{\Gamma} + \sigma^2_{\varepsilon,i}(\mathbf{X_i}'\mathbf{X_i})^{-1}]$$

$$= \quad \boldsymbol{\Gamma} + E_X[\sigma^2_{\varepsilon,i}(\mathbf{X_i}'\mathbf{X_i})^{-1}]$$

Estimate $Var[\mathbf{b_i}]$ with $\frac{1}{N}\Sigma_{i=1}^{N}(\mathbf{b_i} - \overline{\mathbf{b}})(\mathbf{b_i} - \overline{\mathbf{b}})'$

Estimate $E_X[\sigma^2_{\varepsilon,i}(\mathbf{X_i}'\mathbf{X_i})^{-1}]$ with $\frac{1}{N}\Sigma_{i=1}^{N}\hat{\sigma}^2_{\varepsilon,i}(\mathbf{X_i}'\mathbf{X_i})^{-1}$

$$\hat{\boldsymbol{\Gamma}} = \frac{1}{N}\Sigma_{i=1}^{N}(\mathbf{b_i} - \overline{\mathbf{b}})(\mathbf{b_i} - \overline{\mathbf{b}})' - \frac{1}{N}\Sigma_{i=1}^{N}\hat{\sigma}^2_{\varepsilon,i}(\mathbf{X_i}'\mathbf{X_i})^{-1}$$



A Positive Definite Estimator for $\boldsymbol{\Gamma}$

$$\hat{\boldsymbol{\Gamma}} = \frac{1}{N}\Sigma_{i=1}^{N}(\mathbf{b}_i - \bar{\mathbf{b}})(\mathbf{b}_i - \bar{\mathbf{b}})' - \frac{1}{N}\Sigma_{i=1}^{N}\hat{\sigma}^2_{\varepsilon,i}(\mathbf{X}_i'\mathbf{X}_i)^{-1}$$

May not be positive definite. What to do?

(1) The second term converges (in theory) to 0 in T_i. Drop it.

(2) Various Bayesian "shrinkage" estimators,

(3) An ML estimator

Estimating $\boldsymbol{\beta}_i$

$$\hat{\boldsymbol{\beta}}_{GLS} = \Sigma_{i=1}^{N}\mathbf{W}_i\mathbf{b}_{i,OLS}$$

$$\mathbf{W}_i = \{\Sigma_{i=1}^{N}[\boldsymbol{\Gamma} + \sigma_{\varepsilon,i}^{2}(\mathbf{X}_i'\mathbf{X}_i)^{-1}]^{-1}\}^{-1}[\boldsymbol{\Gamma} + \sigma_{\varepsilon,i}^{2}(\mathbf{X}_i'\mathbf{X}_i)^{-1}]^{-1}$$

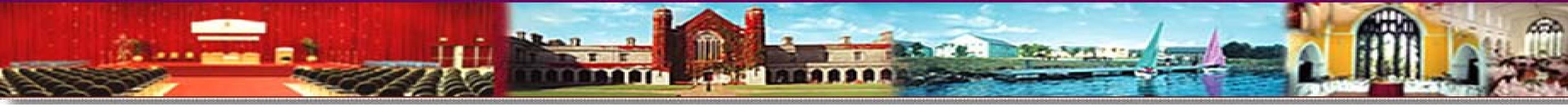
Best linear unbiased predictor based on GLS is

$$\hat{\boldsymbol{\beta}}_i = \mathbf{A}_i\hat{\boldsymbol{\beta}}_{GLS} + (\mathbf{I}-\mathbf{A}_i)\mathbf{b}_{i,OLS} = \mathbf{b}_{i,OLS} + \mathbf{A}_i(\hat{\boldsymbol{\beta}}_{GLS} - \mathbf{b}_{i,OLS})$$

$$\mathbf{A}_i = \{\boldsymbol{\Gamma}^{-1} + [\sigma_{\varepsilon,i}^{2}(\mathbf{X}_i'\mathbf{X}_i)^{-1}]^{-1}\}^{-1}\boldsymbol{\Gamma}^{-1}$$

$$\text{Var}[\hat{\boldsymbol{\beta}}_i \mid \text{all data}] = \mathbf{A}_i\text{Var}[\hat{\boldsymbol{\beta}}_{GLS}]\mathbf{A}_i' +$$

$$[\mathbf{A}_i \quad (\mathbf{I}-\mathbf{A}_i)]\begin{bmatrix} \text{Var}[\hat{\boldsymbol{\beta}}_{GLS}] & \text{Var}[\mathbf{b}_{i,OLS}]\mathbf{W}_i \\ \mathbf{W}'\text{Var}[\mathbf{b}_{i,OLS}]_i & \text{Var}[\mathbf{b}_{i,OLS}] \end{bmatrix}\begin{bmatrix} \mathbf{A}_i \\ (\mathbf{I}-\mathbf{A}_i) \end{bmatrix}$$



OLS and FGLS Estimates

Overall OLS results for pooled sample.

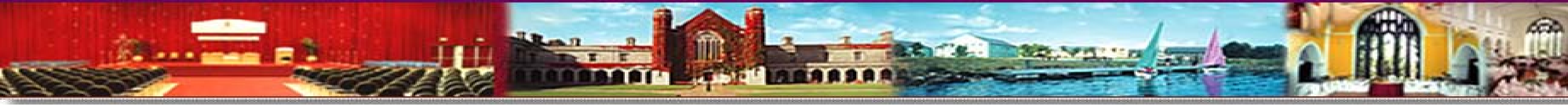
Residuals	Sum of squares	=	14.90436
	Standard error of e	=	.2099898
Fit	R-squared	=	.8549355

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]
Constant	2.39132562	.11693429	20.450	.0000
LINCOMEP	.88996166	.03580581	24.855	.0000
LRPMG	-.89179791	.03031474	-29.418	.0000
LCARPCAP	-.76337275	.01860830	-41.023	.0000

Random Coefficients Model

Residual standard deviation	=	.3498
R squared	=	.5976
Chi-squared for homogeneity test	=	22202.43
Degrees of freedom	=	68
Probability value for chi-squared	=	.000000

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]
CONSTANT	2.40548802	.55014979	4.372	.0000
LINCOMEP	.39314902	.11729448	3.352	.0008
LRPMG	-.24988767	.04372201	-5.715	.0000
LCARPCAP	-.44820927	.05416460	-8.275	.0000

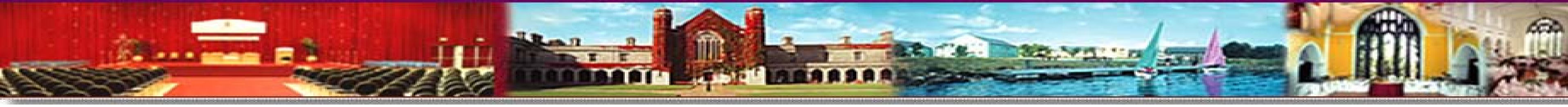


Best Linear Unbiased Country Specific Estimates

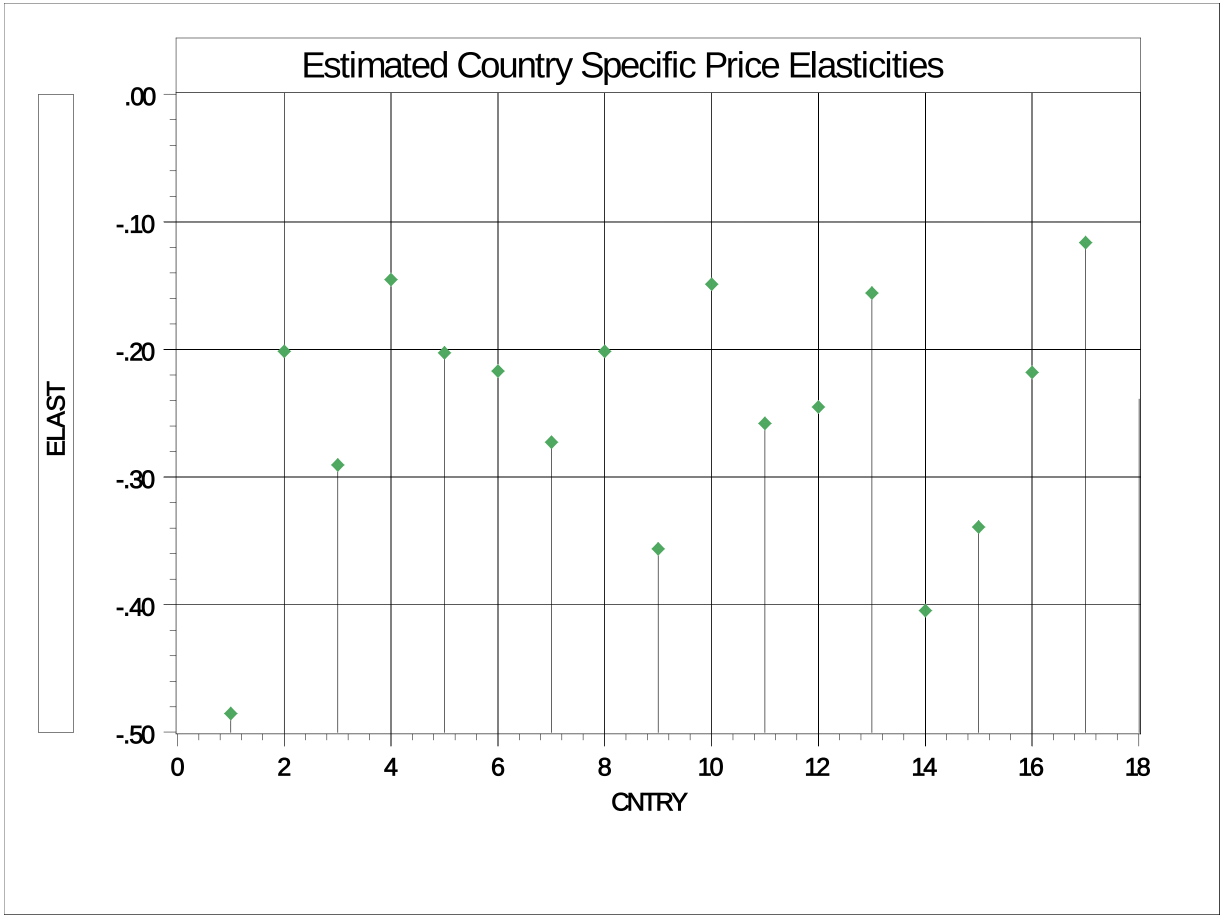
Matrix - BT

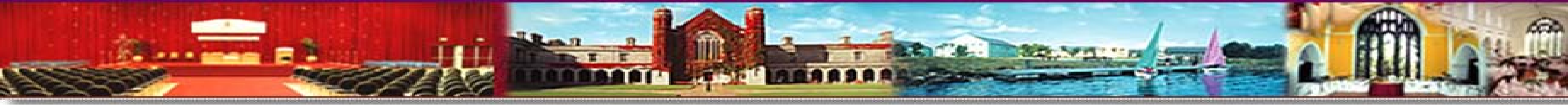
[18, 4] Cell: 3.7712

	1	2	3	4
1	3.7712	0.55187	-0.485435	-0.387265
2	2.55238	0.714802	-0.201465	-0.635849
3	3.28567	0.418864	-0.290662	-0.446393
4	0.291564	0.122538	-0.145417	-0.530267
5	3.1753	1.00926	-0.202796	-0.770054
6	4.13698	0.431209	-0.216905	-0.254492
7	3.58449	0.595853	-0.272624	-0.484243
8	4.16349	0.32981	-0.201541	-0.23423
9	1.30752	0.126914	-0.356194	-0.35955
10	-1.09617	-0.00292814	-0.148817	-0.576652
11	0.914642	0.383942	-0.258079	-0.60601
12	2.68978	0.684036	-0.245217	-0.60324
13	0.23914	-0.146933	-0.155894	-0.313708
14	0.823016	-0.0308741	-0.404606	-0.223692
15	4.60208	0.933709	-0.339252	-0.56949
16	0.601735	0.289237	-0.217959	-0.577492
17	4.03822	0.441826	-0.116262	-0.298403
18	4.21776	0.223542	-0.238852	-0.196738



Estimated Price Elasticities



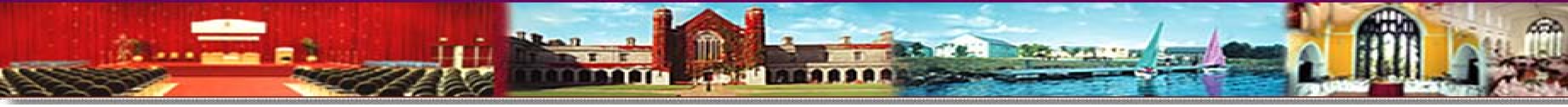


Estimated Γ

Matrix - GAMMA

[4, 4] Cell: 5.06761

	1	2	3	4
1	5.06761	0.792817	0.00189069	-0.0628475
2	0.792817	0.204007	-0.000836729	-0.0649743
3	0.00189069	-0.000836729	0.0219207	-0.00612883
4	-0.0628475	-0.0649743	-0.00612883	0.0435183



Two Step Estimation (Saxonhouse)

A Fixed Effects Model

$$y_{it} = \alpha_i + \mathbf{x}_{it}\boldsymbol{\beta} + \varepsilon_{it}$$

Secondary Model

$$\alpha_i = \mathbf{z}_i\boldsymbol{\delta}$$

Two approaches

(1) Reduced form is a linear model with time constant z_i

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\delta} + \varepsilon_{it}$$

(2) Two step

(a) FEM at step 1

(b) $a_i = \alpha_i + (a_i - \alpha_i) = \mathbf{z}_i\boldsymbol{\delta} + v_i$

$$\text{Var}[v_i] = \sigma_\varepsilon^2\left[\frac{1}{T_i} + \overline{\mathbf{x}}_i'(\mathbf{X}_i'\mathbf{M}_D^i\mathbf{X}_i)^{-1}\overline{\mathbf{x}}_i\right]$$

Use weighted least squares regression of a_i on $\mathbf{z}_i$

A Hierarchical Model

Fixed Effects Model

$$y_{it} = \alpha_i + \mathbf{x}_{it}\boldsymbol{\beta} + \varepsilon_{it}$$

Secondary Model

$\alpha_i = \mathbf{z}_i\boldsymbol{\delta} + u_i$ <========= (No u_i in Saxonhouse)

Two approaches

(1) Reduced form is an REM with time constant z_i

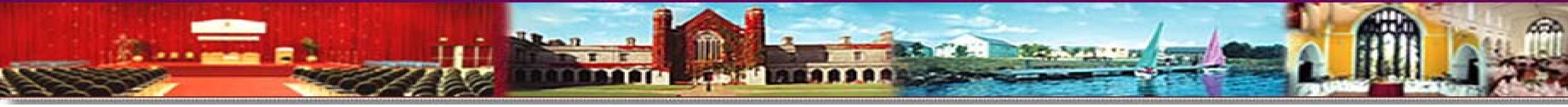
$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\delta} + u_i + \varepsilon_{it}$$

(2) Two step

(a) FEM at step 1

(b) $a_i = \alpha_i + (a_i - \alpha_i) = \mathbf{z}_i\boldsymbol{\delta} + u_i + v_i$

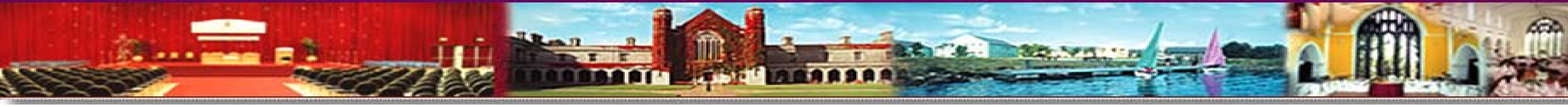
$$\text{Var}[u_i + v_i] = \sigma_u^2 + \sigma_\varepsilon^2\left[\frac{1}{T_i} + \bar{\mathbf{x}}_i'(\mathbf{X}_i'\mathbf{M}_D^i\mathbf{X}_i)^{-1}\bar{\mathbf{x}}_i\right]$$



Analysis of Fannie Mae

- Fannie Mae
- The Funding Advantage
- The Pass Through

Passmore, W., Sherlund, S., Burgess, G., “The Effect of Housing Government-Sponsored Enterprises on Mortgage Rates,” 2005, Real Estate Economics



Two Step Analysis of Fannie-Mae

Fannie Mae's GSE Funding Advantage and Pass Through

$$RM_{i,s,t} = \beta^0_{s,t} + (\beta^1_{s,t}LTV) + \beta^2_{s,t}Small_{i,s,t} + \beta^3_{s,t}Fees_{i,s,t}$$
$$+ \beta^4_{s,t}New_{i,s,t} + \beta^5_{s,t}MtgCo_{i,s,t} + \boxed{\alpha_{s,t}J_{i,s,t}} + \varepsilon_{i,s,t}$$

$i, s, t =$ individual, state, month

1,036,252 observations in 370 state,months.

$RM =$ mortgage

LTV= 3 dummy variables for loan to value

Small = dummy variable for small loan

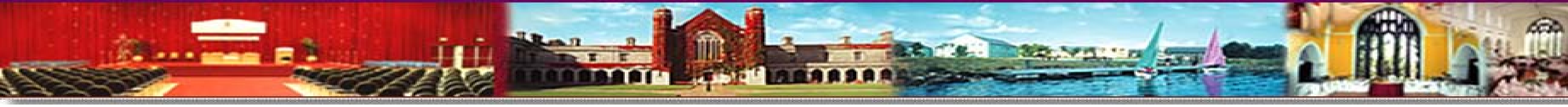
Fees = dummy variable for whether fees paid up front

New = dummy variable for new home

MtgCo = dummy variable for mortgage company

J = dummy variable for whether this is a JUMBO loan

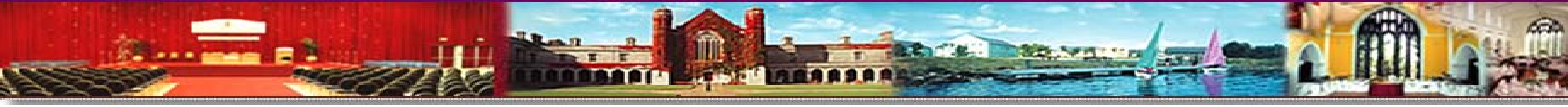
THIS IS THE COEFFICIENT OF INTEREST.



Average of 370 First Step Regressions

Symbol	Variable	Mean	S.D.	Coeff	S.E.
RM	Rate %	7.23	0.79		
J	**Jumbo**	**0.06**	**0.23**	**0.16**	**0.05**
LTV1	75%-80%	0.36	0.48	0.04	0.04
LTV2	81%-90%	0.15	0.35	0.17	0.05
LTV3	>90%	0.22	0.41	0.15	0.04
New	New Home	0.17	0.38	0.05	0.04
Small	< $100,000	0.27	0.44	0.14	0.04
Fees	Fees paid	0.62	0.52	0.06	0.03
MtgCo	Mtg. Co.	0.67	0.47	0.12	0.05

$R^2 = 0.77$



Second Step Uses 370 Estimates of α_{st}

$\alpha_{s,t} = \beta_0 +$

β_1 GSE Funding Advantage$_{s,t}$ - estimated separately

β_2 Risk free cost of credit$_{s,t}$

β_3 Corporate debt spreads$_{s,t}$ - estimated 4 different ways

β_4 Prepayment spread$_{s,t}$

β_5 Maturity mismatch risk$_{s,t}$

β_6 Aggregate Demand$_{s,t}$

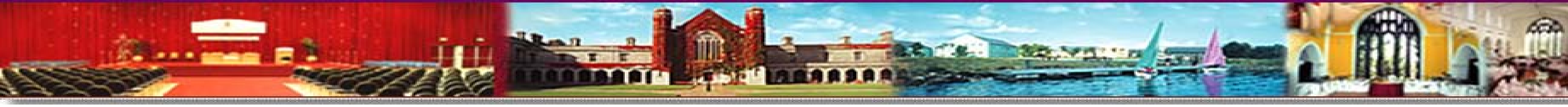
β_7 Long term interest rate$_{s,t}$

β_8 Market Capacity$_{s,t}$

β_9 Time trend$_{s,t}$

$\beta_{10\text{-}13}$ 4 dummy variables for CA, NJ, MD, VA$_{s,t}$

$\beta_{14\text{-}16}$ 3 dummy variables for calendar quarters$_{s,t}$



Estimates of $\boldsymbol{\beta}_1$

Second step based on 370 observations. Corrected for "heteroscedasticity, autocorrelation, and monthly clustering."

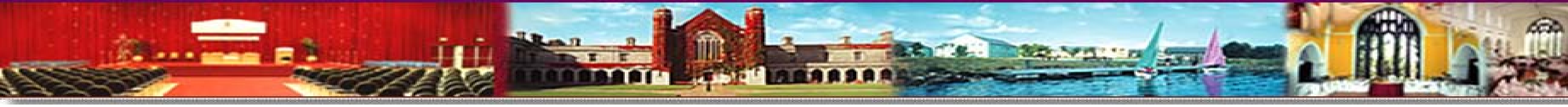
Four estimates based on different estimates of corporate credit spread:

0.07 (0.11)   0.31 (0.11)   017 (0.10)   0.10 (0.11)

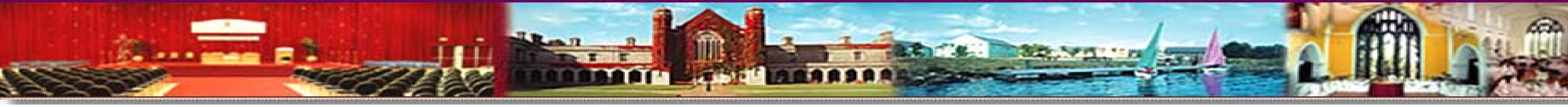
Reconcile the 4 estimates with a minimum distance estimator

$$\text{Minimize } [(\hat{\beta}_1^1-\beta_1),(\hat{\beta}_1^2-\beta_1),(\hat{\beta}_1^3-\beta_1),(\hat{\beta}_1^4-\beta_1)]'\hat{\boldsymbol{\Omega}}^{-1}\begin{bmatrix}(\hat{\beta}_1^1-\beta_1)\\(\hat{\beta}_1^2-\beta_1)\\(\hat{\beta}_1^3-\beta_1)\\(\hat{\beta}_1^4-\beta_1)\end{bmatrix}$$

Estimated mortgage rate reduction: About 7 basis points. .07%.



RANDOM EFFECTS - CONTINUOUS



Continuous Parameter Variation (The Random Parameters Model)

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta}_i + \varepsilon_{it}, \text{ each observation}$$

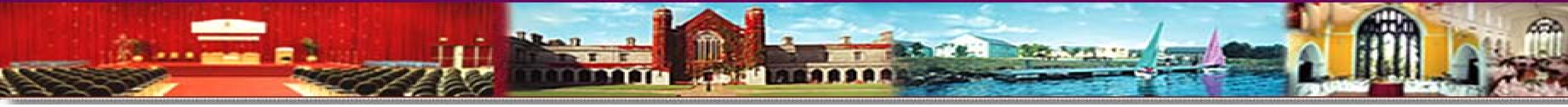
$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \; T_i \text{ observations}$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{u}_i$$

$$E[\mathbf{u}_i \mid \mathbf{X}_i] = \mathbf{0}$$

$$\text{Var}[\mathbf{u}_i \mid \mathbf{X}_i] = \boldsymbol{\Gamma} \text{ constant but nonzero}$$

$$f(\mathbf{u}_i \mid \mathbf{X}_i) = g(\mathbf{u}_i, \boldsymbol{\Gamma}), \text{ a density that does not involve } \mathbf{X}_i$$



OLS and GLS Are Consistent

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \; T_i \text{ observations}$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{u}_i$$

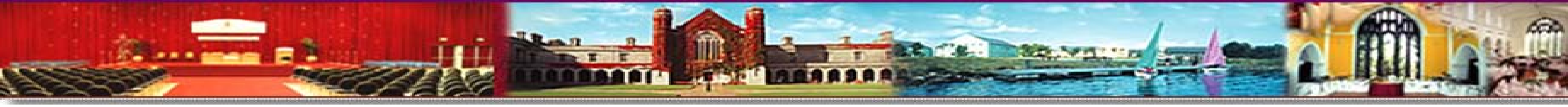
$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i, \; T_i \text{ observations}$$

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{w}_i$$

$$E[\mathbf{w}_i \mid \mathbf{X}_i] = \mathbf{X}_i E[\mathbf{u}_i \mid \mathbf{X}_i] + E[\boldsymbol{\varepsilon}_i \mid \mathbf{X}_i] = \mathbf{0}$$

$$\text{Var}[\mathbf{w}_i \mid \mathbf{X}_i] = \sigma_\varepsilon^2 \mathbf{I} + \mathbf{X}_i \boldsymbol{\Gamma} \mathbf{X}_i'$$

(Discussed earlier - two step GLS)



ML Estimation of the RPM

Sample data generation

$$y_{i,t} = \mathbf{x}_{i,t}'\boldsymbol{\beta}_i + \varepsilon_{i,t}$$

Individual heterogeneity

$$\boldsymbol{\beta}_i = \overline{\boldsymbol{\beta}} + \mathbf{u}_i$$

Conditional log likelihood

$$\log f(y_{i1},\ldots,y_{iT_i} \mid \mathbf{X}_i, \boldsymbol{\beta}_i, \sigma_\varepsilon) = \log \prod_{t=1}^{T_i} f(y_{it} \mid \mathbf{x}_{it}, \boldsymbol{\beta}_i, \sigma_\varepsilon)$$

Unconditional log likelihood

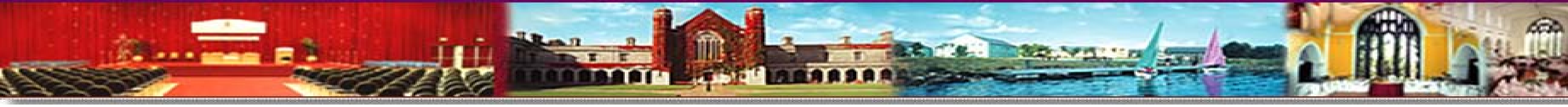
$$\log L(\overline{\boldsymbol{\beta}}, \boldsymbol{\Gamma}, \sigma_\varepsilon) = \int_{\boldsymbol{\beta}_i} \log \prod_{t=1}^{T_i} f(y_{it} \mid \mathbf{x}_{it}, \boldsymbol{\beta}_i, \sigma_\varepsilon) g(\boldsymbol{\beta}_i \mid \overline{\boldsymbol{\beta}}, \boldsymbol{\Gamma}) d\boldsymbol{\beta}_i$$

(1) Using simulated ML or quadrature, maximize to estimate $\overline{\boldsymbol{\beta}}, \boldsymbol{\Gamma}, \sigma_\varepsilon$.

(2) Using data and estimated structural parameters, compute $E[\boldsymbol{\beta}_i \mid \text{data}_i, \overline{\boldsymbol{\beta}}, \boldsymbol{\Gamma}, \sigma_\varepsilon]$

RP Gasoline Market

```
-----------------------------------------------------------------------------
Random Coefficients  LinearRg Model
Dependent variable              LGASPCAR
Log likelihood function        434.79652
Restricted log likelihood         .00000
Chi squared [  10 d.f.]        869.59305
Significance level                .00000
Estimation based on N =     342, K =  15
Inf.Cr.AIC  =   -839.6 AIC/N =    -2.455
Model estimated: Mar 19, 2012, 19:18:37
Sample is 19 pds and      18 individuals
LINEAR regression model
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
LGASPCAR|  Coefficient      Error       z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
        |Means for random parameters
Constant|     1.67771***     .08386    20.01  .0000     1.51334   1.84207
LINCOMEP|      .70469***     .02601    27.09  .0000      .65371    .75568
   LRPMG|     -.43202***     .01919   -22.52  .0000     -.46962   -.39441
LCARPCAP|     -.73081***     .01460   -50.05  .0000     -.75942   -.70219
        |Diagonal elements of Cholesky matrix
Constant|      .76412***     .07031    10.87  .0000      .62632    .90192
LINCOMEP|      .05570**      .02186     2.55  .0108      .01286    .09855
   LRPMG|      .00333        .01015      .33  .7427     -.01656    .02323
LCARPCAP|      .03151***     .00141    22.28  .0000      .02874    .03428
        |Below diagonal elements of Cholesky matrix
lLIN_ONE|      .13405***     .02569     5.22  .0000      .08369    .18440
lLRP_ONE|      .00447        .01799      .25  .8038     -.03080    .03974
lLRP_LIN|     -.19488***     .02580    -7.55  .0000     -.24545   -.14431
lLCA_ONE|     -.16537***     .01253   -13.20  .0000     -.18993   -.14082
lLCA_LIN|     -.08739***     .01399    -6.25  .0000     -.11480   -.05997
lLCA_LRP|     -.06151***     .00152   -40.41  .0000     -.06450   -.05853
        |Variance parameter given is sigma
Std.Dev.|      .05377***     .00140    38.43  .0000      .05103    .05651
--------+--------------------------------------------------------------------
```



Parameter Covariance matrix

Implied covariance matrix of random parameters

Var_Beta	1	2	3	4
1	.583880	.102430	.00341640	-.126363
2	.102430	.0210719	-.0102561	-.0270354
3	.00341640	-.0102561	.0380098	.0160854
4	-.126363	-.0270354	.0160854	.0397606

Implied standard deviations of random parameters

S.D_Beta	1
1	.764120
2	.145162
3	.194961
4	.199401

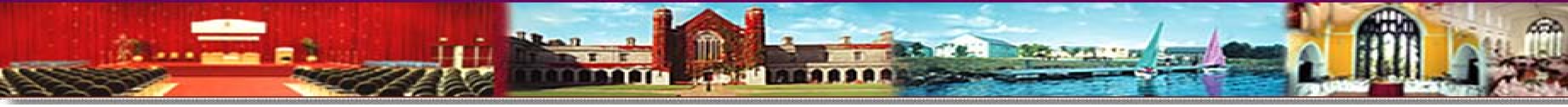
Implied correlation matrix of random parameters

Cor_Beta	1	2	3	4
1	1.00000	.923446	.0229329	-.829341
2	.923446	1.00000	-.362395	-.934017
3	.0229329	-.362395	1.00000	.413769
4	-.829341	-.934017	.413769	1.00000

RP vs. Gen1

	1	2	3	4
1	3.27213	0.503373	-0.147012	-0.42937
2	1.96274	0.640215	-0.482069	-0.642943
3	1.75442	0.676578	-0.487375	-0.788322
4	1.15407	0.784814	-0.478848	-0.859231
5	2.03636	0.657008	-0.376327	-0.656425
6	2.85148	0.465676	-0.539219	-0.410915
7	3.46929	0.356281	-0.543699	-0.345737
8	1.8807	0.65083	-0.500032	-0.705248
9	2.74472	0.47831	-0.562992	-0.446671
10	1.08883	0.853082	-0.277589	-0.888086
11	1.96602	0.705496	-0.250226	-0.703198
12	2.20724	0.633617	-0.351074	-0.622019
13	2.15123	0.592559	-0.531639	-0.569621
14	2.46213	0.575075	-0.401558	-0.600073
15	2.29889	0.524727	-0.679244	-0.517898
16	1.35328	0.775068	-0.38542	-0.798245
17	3.35646	0.425103	-0.368257	-0.351088
18	2.31756	0.597369	-0.415146	-0.676744

	1	2	3	4
1	3.7712	0.55187	-0.485435	-0.387265
2	2.55238	0.714802	-0.201465	-0.635849
3	3.28567	0.418864	-0.290662	-0.446393
4	0.291564	0.122538	-0.145417	-0.530267
5	3.1753	1.00926	-0.202796	-0.770054
6	4.13698	0.431209	-0.216905	-0.254492
7	3.58449	0.595853	-0.272624	-0.484243
8	4.16349	0.32981	-0.201541	-0.23423
9	1.30752	0.126914	-0.356194	-0.35955
10	-1.09617	-0.00292814	-0.148817	-0.576652
11	0.914642	0.383942	-0.258079	-0.60601
12	2.68978	0.684036	-0.245217	-0.60324
13	0.23914	-0.146933	-0.155894	-0.313708
14	0.823016	-0.0308741	-0.404606	-0.223692
15	4.60208	0.933709	-0.339252	-0.56949
16	0.601735	0.289237	-0.217959	-0.577492
17	4.03822	0.441826	-0.116262	-0.298403
18	4.21776	0.223542	-0.238852	-0.196738



Modeling Parameter Heterogeneity

Conditional Linear Regression

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta}_i + \varepsilon_{i,t}$$

Individual heterogeneity in the means of the parameters

$$\boldsymbol{\beta}_i = \bar{\boldsymbol{\beta}} + \boldsymbol{\Delta}\mathbf{z}_i + \mathbf{u}_i$$

$$E[\mathbf{u}_i \mid \mathbf{X}_i, \mathbf{z}_i]$$

Heterogeneity in the variances of the parameters

$$\text{Var}[u_{i,k} \mid \text{data}_i] = \theta_k \exp(\mathbf{h}_i'\boldsymbol{\delta}_k)$$

Estimation by maximum simulated likelihood

Hierarchical Linear Model

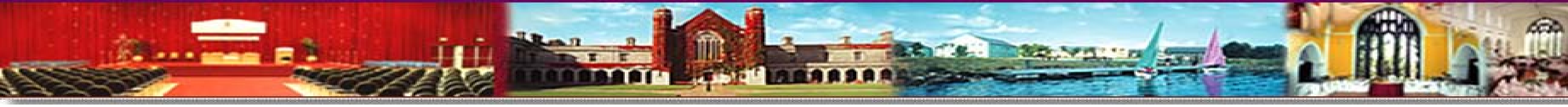
COUNTRY = name of country		
YEAR = year, 1960-1978		
LGASPCAR = log of consumption per car		y
LINCOMEP = log of per capita income		z
LRPMG = log of real price of gasoline		x1
LCARPCAP = log of per capita number of cars		x2

$$y_{it} = \beta_{1i} + \beta_{2i}\, x1_{it} + \beta_{3i}\, x2_{it} + \varepsilon_{it}.$$

$$\beta_{1i} = \beta_1 + \Delta_1\, z_i + u_{1i}$$

$$\beta_{2i} = \beta_2 + \Delta_2\, z_i + u_{2i}$$

$$\beta_{3i} = \beta_3 + \Delta_3\, z_i + u_{3i}$$

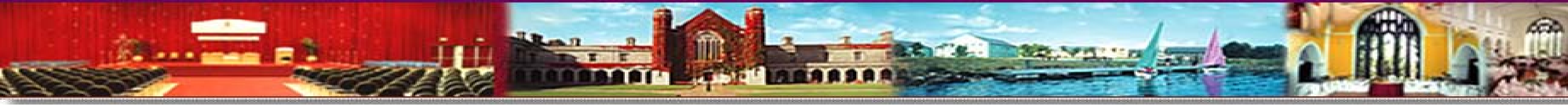


Estimated HLM

```
Random Coefficients  LinearRg Model
Dependent variable             LGASPCAR
Log likelihood function       413.16349
Restricted log likelihood        .00000
Chi squared [   9 d.f.]       826.32698
Significance level               .00000
Estimation based on N =    342, K =  13
Inf.Cr.AIC  =   -800.3 AIC/N =   -2.340
Model estimated: Mar 19, 2012, 19:24:17
Sample is 19 pds and     18 individuals
LINEAR regression model
```

LGASPCAR	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Means for random parameters					
Constant	5.72753***	.49404	11.59	.0000	4.75923	6.69582
LRPMG	-2.16228***	.13928	-15.52	.0000	-2.43527	-1.88928
LCARPCAP	.04564	.04960	.92	.3574	-.05157	.14284
	Diagonal elements of Cholesky matrix					
Constant	1.66588***	.07174	23.22	.0000	1.52526	1.80650
LRPMG	.08854***	.02222	3.98	.0001	.04499	.13209
LCARPCAP	.07959***	.00221	35.96	.0000	.07526	.08393
	Below diagonal elements of Cholesky matrix					
lLRP_ONE	.10055***	.01657	6.07	.0000	.06808	.13302
lLCA_ONE	-.17701***	.00745	-23.76	.0000	-.19161	-.16241
lLCA_LRP	-.05170***	.00222	-23.28	.0000	-.05605	-.04735
	Heterogeneity in the means of random parameters					
cONE_LIN	.69371***	.07948	8.73	.0000	.53794	.84949
cLRP_LIN	-.26680***	.02185	-12.21	.0000	-.30963	-.22397
cLCA_LIN	.04775***	.00790	6.05	.0000	.03227	.06323
	Variance parameter given is sigma					
Std.Dev.	.05739***	.00230	24.97	.0000	.05288	.06189

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.



RP vs. HLM

	1	2	3	4
1	3.27213	0.503373	-0.147012	-0.42937
2	1.96274	0.640215	-0.482069	-0.642943
3	1.75442	0.676578	-0.487375	-0.788322
4	1.15407	0.784814	-0.478848	-0.859231
5	2.03636	0.657008	-0.376327	-0.656425
6	2.85148	0.465676	-0.539219	-0.410915
7	3.46929	0.356281	-0.543699	-0.345737
8	1.8807	0.65083	-0.500032	-0.705248
9	2.74472	0.47831	-0.562992	-0.446671
10	1.08883	0.853082	-0.277589	-0.888086
11	1.96602	0.705496	-0.250226	-0.703198
12	2.20724	0.633617	-0.351074	-0.622019
13	2.15123	0.592559	-0.531639	-0.569621
14	2.46213	0.575075	-0.401558	-0.600073
15	2.29889	0.524727	-0.679244	-0.517898
16	1.35328	0.775068	-0.38542	-0.798245
17	3.35646	0.425103	-0.368257	-0.351088
18	2.31756	0.597369	-0.415146	-0.676744

	1	2	3
1	2.48532	-0.40819	-0.150259
2	1.01239	-0.478012	-0.318541
3	1.67214	-0.648518	-0.307997
4	0.277382	-0.620352	-0.427846
5	2.62543	-0.383924	-0.130059
6	3.38261	-0.429054	-0.0365439
7	2.03139	-0.226011	-0.263547
8	1.62824	-0.313808	-0.278519
9	0.838691	-0.336901	-0.320116
10	-1.36663	-0.485345	-0.596943
11	-1.36434	-0.835113	-0.583871
12	1.39568	-0.542576	-0.295864
13	0.490709	-0.629258	-0.407443
14	2.94572	0.0875361	-0.157653
15	0.770315	-0.499533	-0.352927
16	-0.217383	-0.169486	-0.473712
17	2.38066	-0.459525	-0.162769
18	4.03452	-0.422281	-0.0352291

Hierarchical Bayesian Estimation

Sample data generation: $y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta}_{\mathbf{i}} + \varepsilon_{i,t},\ \varepsilon_{i,t} \sim N[0,\sigma_\varepsilon^2]$

Individual heterogeneity: $\boldsymbol{\beta}_{\mathbf{i}} = \boldsymbol{\beta} + \mathbf{u}_{\mathbf{i}},\ u_i \sim N[\mathbf{0},\boldsymbol{\Gamma}]$

What information exists about 'the model?'

Prior densities for structural parameters :

$p(\log\sigma_\varepsilon)$= uniform density with (large) parameter A^0

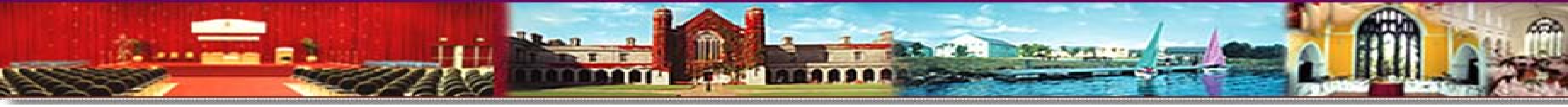
$p(\boldsymbol{\beta}) = N[\boldsymbol{\beta}^{\mathbf{0}},\boldsymbol{\Sigma}^0]$, e.g., $\mathbf{0}$ and (large) $v^0\mathbf{I}$

$p(\boldsymbol{\Gamma})$ = Inverse Wishart[...]

Priors for parameters of interest :

$p(\boldsymbol{\beta}_{\mathbf{i}}) = N[\boldsymbol{\beta},\boldsymbol{\Gamma}]$

$p(\sigma_\varepsilon)$ = as above.



Estimation of Hierarchical Bayes Models

(1) Analyze 'posteriors' for hyperparameters $\boldsymbol{\beta}, \boldsymbol{\Gamma}, \sigma$

(2) Analyze posterior for group level parameters, $\boldsymbol{\beta}_i$

Estimators are Means and Variances of posterior distributions.

Algorithm: Generally, Gibbs sampling from posteriors with resort to laws of large numbers

A Hierarchical Linear Model

German Health Care Data

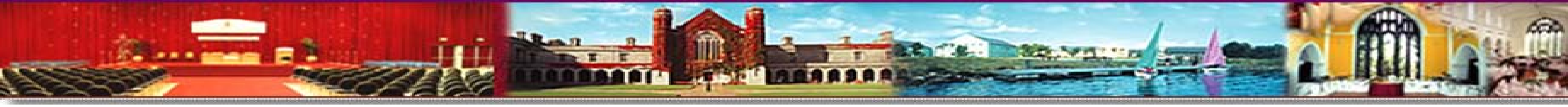
$$\text{Hsat} = \beta_1 + \beta_2\text{AGE}_{it} + \gamma_i\,\text{EDUC}_{it} + \beta_4\,\text{MARRIED}_{it} + \varepsilon_{it}$$

$$\gamma_i = \alpha_1 + \alpha_2\text{FEMALE}_i + u_i$$

```
Sample ; all $
Setpanel ; Group = id ; Pds = ti $
Regress ; For [ti = 7] ; Lhs = newhsat  ; Rhs = one,age,educ,married
                       ; RPM = female    ; Fcn = educ(n)
                       ; pts = 25 ; halton  ; panel ; Parameters$
Sample ; 1 – 887 $
Create  ; betaeduc = beta_i $
Dstat    ; rhs = betaeduc $
Histogram ; Rhs = betaeduc $
```

OLS Results

```
OLS Starting values for random parameters model...
Ordinary     least squares regression ............
LHS=NEWHSAT  Mean                    =        6.69641
             Standard deviation      =        2.26003
             Number of observs.      =           6209
Model size   Parameters              =              4
             Degrees of freedom      =           6205
Residuals    Sum of squares          =    29671.89461
             Standard error of e     =        2.18676
Fit          R-squared               =         .06424
             Adjusted R-squared      =         .06378
Model test   F[  3,  6205] (prob) =   142.0(.0000)
--------+-------------------------------------------------------------
        |                 Standard             Prob.         Mean
 NEWHSAT| Coefficient       Error        z     z>|Z|         of X
--------+-------------------------------------------------------------
Constant|    7.02769***      .22099    31.80   .0000
     AGE|    -.04882***      .00307   -15.90   .0000      44.3352
 MARRIED|     .29664***      .07701     3.85   .0001       .84539
    EDUC|     .14464***      .01331    10.87   .0000      10.9409
--------+-------------------------------------------------------------
```



Maximum Simulated Likelihood

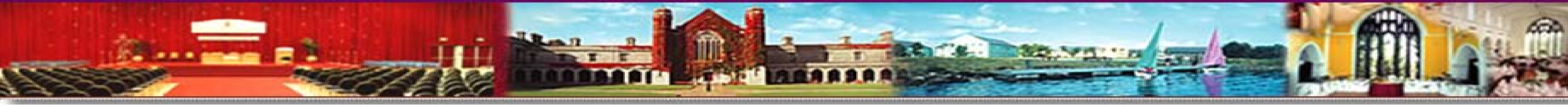
```
Normal exit:  27 iterations. Status=0. F=    12584.28
-----------------------------------------------------------------
Random Coefficients  LinearRg Model
Dependent variable               NEWHSAT
Log likelihood function     -12583.74717
Estimation based on N =   6209, K =   7
Unbalanced panel has     887 individuals
LINEAR regression model
Simulation based on     25 Halton draws
--------+--------------------------------------------------------
        |                 Standard            Prob.        Mean
 NEWHSAT| Coefficient       Error       z     z>|Z|        of X
--------+--------------------------------------------------------
        |Nonrandom parameters
Constant|    7.34576***      .15415    47.65  .0000
     AGE|    -.05878***      .00206   -28.56  .0000    44.3352
 MARRIED|     .23427***      .05034     4.65  .0000     .84539
        |Means for random parameters
    EDUC|     .16580***      .00951    17.43  .0000    10.9409
        |Scale parameters for dists. of random parameters
    EDUC|    1.86831***      .00179  1044.68  .0000
        |Heterogeneity in the means of random parameters
cEDU_FEM|    -.03493***      .00379    -9.21  .0000
        |Variance parameter given is sigma
Std.Dev.|    1.58877***      .00954   166.45  .0000
--------+--------------------------------------------------------
```

Simulating Conditional Means for Individual Parameters

$$\hat{E}(\boldsymbol{\beta}_i \mid \mathbf{y}_i, \mathbf{X}_i) = \frac{\frac{1}{R}\sum_{r=1}^{R} (\hat{\boldsymbol{\beta}} + \hat{\mathbf{L}}\mathbf{w}_{i,r}) \prod_{t=1}^{T_i} \frac{1}{\hat{\sigma}} \phi\left(\frac{y_{it} - (\hat{\boldsymbol{\beta}} + \hat{\mathbf{L}}\mathbf{w}_{i,r})'\mathbf{x}_{it}}{\hat{\sigma}} \right)}{\frac{1}{R}\sum_{r=1}^{R} \prod_{t=1}^{T_i} \frac{1}{\hat{\sigma}} \phi\left(\frac{y_{it} - (\hat{\boldsymbol{\beta}} + \hat{\mathbf{L}}\mathbf{w}_{i,r})'\mathbf{x}_{it}}{\hat{\sigma}} \right)}$$

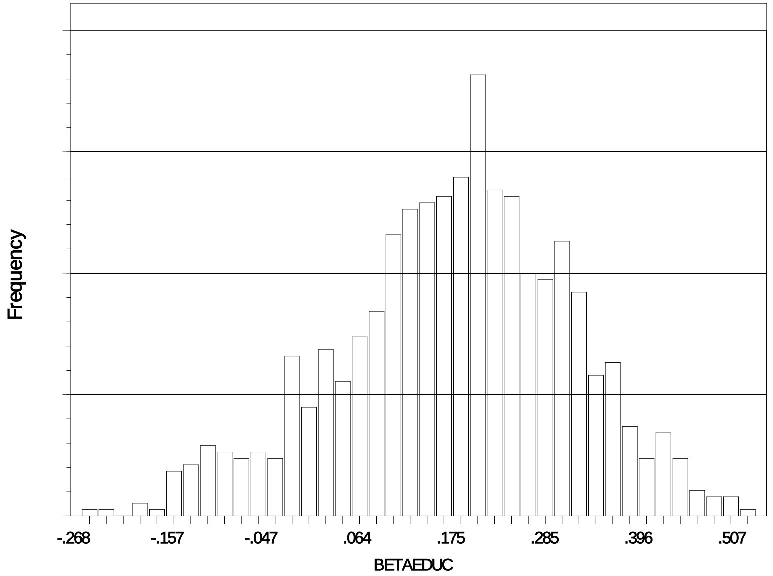
$$= \frac{1}{R}\sum_{r=1}^{R} \widehat{Weight}_{ir} \hat{\boldsymbol{\beta}}_{ir}$$

Posterior estimates of E[parameters(i) | Data(i)]



“Individual Coefficients”

```
--> Sample ; 1 - 887 $
--> create ; betaeduc = beta_i $
--> dstat   ; rhs = betaeduc $
Descriptive Statistics
All results based on nonmissing observations.
==============================================================================
Variable      Mean        Std.Dev.      Minimum      Maximum        Cases Missing
==============================================================================
All observations in current sample
--------+---------------------------------------------------------------------
BETAEDUC|  .161184      .132334     -.268006      .506677          887        0
```



A Hierarchical Linear Model

- A hedonic model of house values
- Beron, K., Murdoch, J., Thayer, M., “Hierarchical Linear Models with Application to Air Pollution in the South Coast Air Basin,” American Journal of Agricultural Economics, 81, 5, 1999.

Three Level HLM

y_{ijk} = log of home sale price i, neighborhood j, community k.

$$y_{ijk} = \sum_{m=1}^{M} \beta_{jk}^{m} x_{ijk}^{m} + \varepsilon_{ijk} \text{ (linear regression model)}$$

x_{ijk}^{m} = sq.ft, #baths, lot size, central heat, AC, pool, good view, age, distance to beach

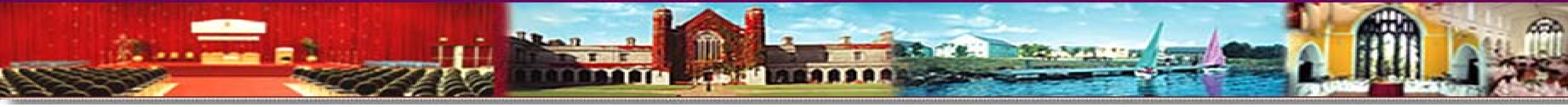
Random coefficients

$$\beta_{jk}^{m} = \sum_{q=1}^{Q_m} \pi_j^q N_{jk}^q + w_{jk}$$

N_{jk}^{q} = %population poor, race mix, avg age, avg. travel to work, FBI crime index, school avg. CA achievement test score

$$\pi_j^q = \sum_{s=1}^{S_{qm}} \delta^s E_j^{qm} + v_j$$

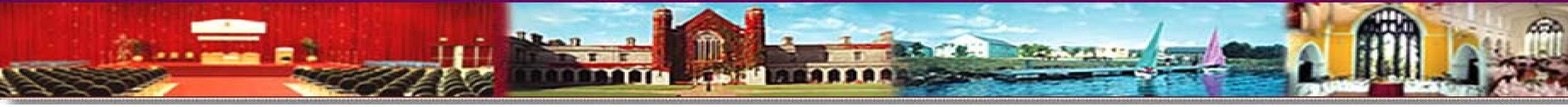
E_j^{qm} = air quality measure, visibility



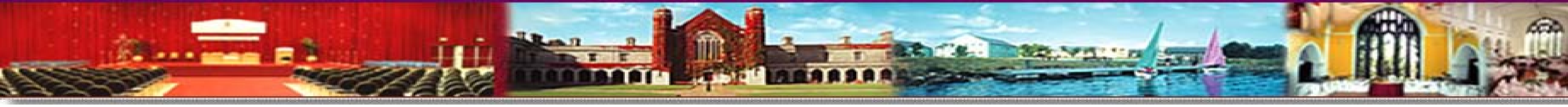
Mixed Model Estimation

Programs differ on the models fitted, the algorithms, the paradigm, and the extensions provided to the simplest RPM, $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{w}_i$.

- WinBUGS:
  - MCMC
  - User specifies the model – constructs the Gibbs Sampler/Metropolis Hastings
- MLWin:
  - Linear and some nonlinear – logit, Poisson, etc.
  - Uses MCMC for MLE (noninformative priors)
- SAS: Proc Mixed.
  - Classical
  - Uses primarily a kind of GLS/GMM (method of moments algorithm for loglinear models)
- Stata: Classical
  - Several loglinear models – GLAMM. Mixing done by quadrature.
  - Maximum simulated likelihood for multinomial choice (Arne Hole, user provided)
- LIMDEP/NLOGIT
  - Classical
  - Mixing done by Monte Carlo integration – maximum simulated likelihood
  - Numerous linear, nonlinear, loglinear models
- Ken Train’s Gauss Code
  - Monte Carlo integration
  - Mixed Logit (mixed multinomial logit) model only (but free!)
- Biogeme
  - Multinomial choice models
  - Many experimental models (developer’s hobby)



GEN 2.1 – RANDOM EFFECTS - DISCRETE



Heterogeneous Production Model

$$\text{Health}_{i,t} = \alpha_i + \beta_i \text{HEXP}_{i,t} + \gamma_i \text{EDUC}_{i,t} + \varepsilon_{i,t}$$

i = country, t = year

Health = health care outcome, e.g., life expectancy

HEXP = health care expenditure

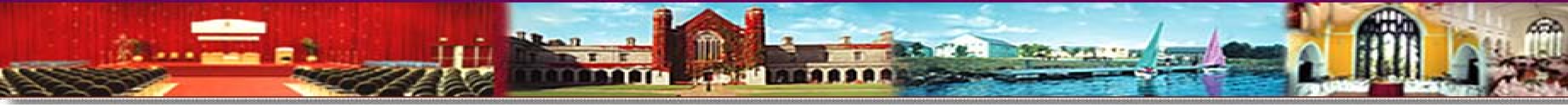
EDUC = education

Parameter heterogeneity:

Discrete? Aids dominated vs. QOL dominated

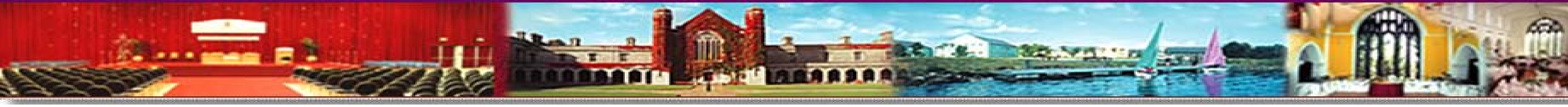
Continuous? Cross cultural heterogeneity

World Health Organization, "The 2000 World Health Report"



Parameter Heterogeneity

- Fixed and Random Effects Models
  - Latent common time invariant “effects”
  - Heterogeneity in level parameter – constant term – in the model
- **General Parameter Heterogeneity** in Models
  - Discrete: There is more than one time of individual in the population – parameters differ across types. Produces a **Latent Class Model**
  - Continuous; Parameters vary randomly across individuals: Produces a **Random Parameters Model** or a **Mixed Model**. (Synonyms)



Parameter Heterogeneity

(1) Regression model

$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta}_{i} + \varepsilon_{it}$

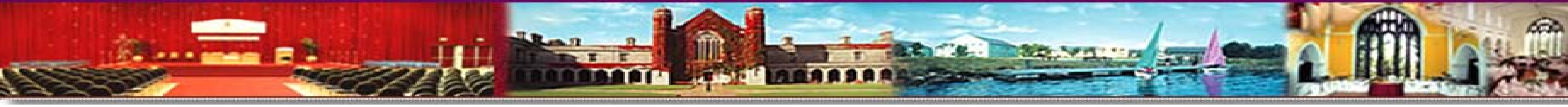
(2) Conditional probability model

$f(y_{it} \mid x_{i,t}, \boldsymbol{\beta}_{i})$

(3) Heterogeneity - how are parameters distributed across individuals?

(a) Discrete - the population contains a mixture of J types of individuals.

(b) Continuous. Parameters are part of the stochastic structure of the population.



Discrete Parameter Variation

The Latent Class Model

(1) Population is a (finite) mixture of Q types of individuals.

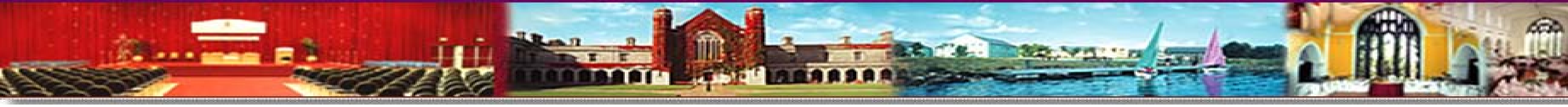
q = 1,...,Q. Q 'classes' differentiated by $(\boldsymbol{\beta}_q, \sigma_{\varepsilon,q})$

(a) Analyst does not know class memberships. ('latent.')

(b) 'Mixing probabilities' (from the point of view of the analyst) are $\pi_1, ..., \pi_Q$, with $\Sigma_{q=1}^{Q} \pi_q = 1$

(2) Conditional density is

$$P(y_{i,t} \mid \text{class} = q) = f(y_{it} \mid x_{i,t}, \boldsymbol{\beta}_q, \sigma_{\varepsilon,q})$$



Example: Mixture of Normals

Q normal populations each with a mean μ_q and standard deviation σ_q

For each individual in each class at each period,

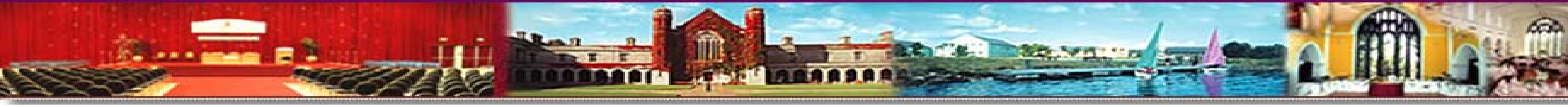
$$f(y_{it} \mid class = q) = \frac{1}{\sigma_q\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y_{it}-\mu_q}{\sigma_q}\right)^2\right] = \frac{1}{\sigma_j}\phi\left(\frac{y_{it}-\mu_q}{\sigma_q}\right).$$

Panel data, T observations on each individual i,

$$f(y_{i1},\ldots,y_{iT} \mid class = q) = \left(\frac{1}{\sigma_q\sqrt{2\pi}}\right)^T \exp\left[\Sigma_{t=1}^{T} -\frac{1}{2}\left(\frac{y_{it}-\mu_q}{\sigma_q}\right)^2\right]$$

Log Likelihood

$$\log L = \sum_{i=1}^{N} \log\left\{\sum_{q=1}^{Q} \pi_q \left\langle \left(\frac{1}{\sigma_q\sqrt{2\pi}}\right)^T \exp\left[\Sigma_{t=1}^{T} -\frac{1}{2}\left(\frac{y_{it}-\mu_q}{\sigma_q}\right)^2\right]\right\rangle\right\}$$



An Extended Latent Class Model

(1) There are Q classes, unobservable to the analyst

(2) Class specific model: $f(y_{it} \mid \mathbf{x}_{it}, \boldsymbol{\beta}, class = q) = g(y_{it}, \mathbf{x}_{it}, \;_q)$

(3) Conditional class probabilities π_q

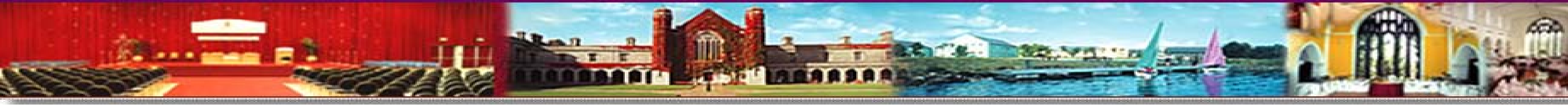
Common multinomial logit form for prior class probabilities

to constrain all probabilities to (0,1) and ensure $\sum_{q=1}^{Q} \pi_q = 1$;

multinomial logit form for class probabilities;

$$P(class=q|\boldsymbol{\delta}) = \pi_q = \frac{\exp(\delta_q)}{\sum_{j=1}^{J} \exp(\delta_q)}, \ \delta_Q = 0$$

Note, $\delta_q = \log(\pi_q / \pi_Q)$.



Log Likelihood for an LC Model

Conditional density for each observation is

$$P(y_{i,t} \mid \mathbf{x}_{i,t}, \text{class} = q) = f(y_{it} \mid \mathbf{x}_{i,t}, \boldsymbol{\beta}_q)$$

Joint conditional density for T_i observations is

$$f(y_{i1}, y_{i2}, \ldots, y_{i,T_i} \mid \mathbf{X}_i, \boldsymbol{\beta}_q) = \prod_{t=1}^{T_i} f(y_{it} \mid \mathbf{x}_{i,t}, \boldsymbol{\beta}_q)$$

(T_i may be 1. This is not only a 'panel data' model.)

Maximize this for each class if the classes are known.

They aren't. Unconditional density for individual i is

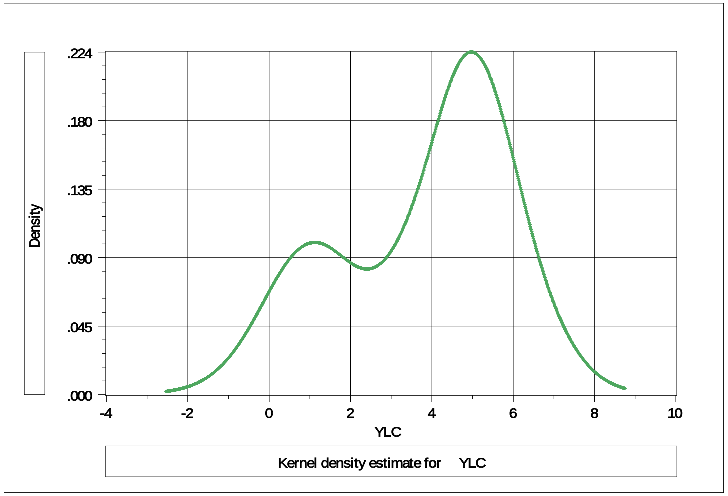
$$f(y_{i1}, y_{i2}, \ldots, y_{i,T_i} \mid \mathbf{X}_i) = \sum_{q=1}^{Q} \pi_q \left(\prod_{t=1}^{T_i} f(y_{it} \mid \mathbf{x}_{i,t}, \boldsymbol{\beta}_q) \right)$$

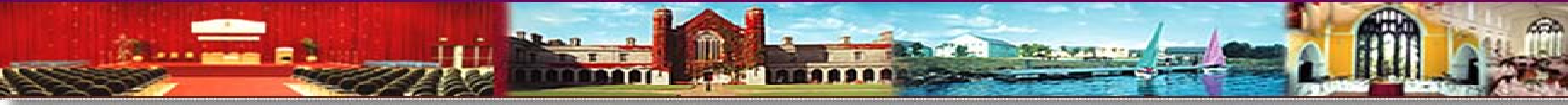
Log Likelihood

$$\text{LogL}(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_Q, \delta_1, \ldots, \delta_Q) = \sum_{i=1}^{N} \log \sum_{q=1}^{Q} \pi_q \prod_{t=1}^{T_i} f(y_{it} \mid \mathbf{x}_{i,t}, \boldsymbol{\beta}_q)$$

Unmixing a Mixed Sample N[1,1] and N[5,1]

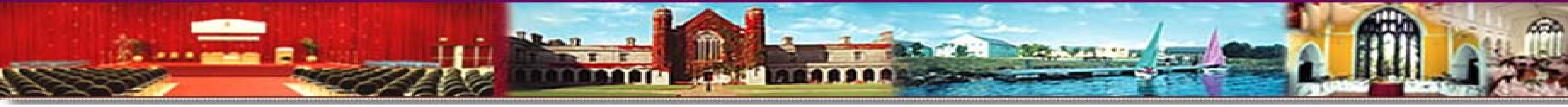
```
Sample  ; 1 – 1000$
Calc     ; Ran(123457)$
Create   ; lc1=rnn(1,1) ; lc2=rnn(5,1)$
Create   ; class=rnu(0,1)$
Create   ; if(class<.3)ylc=lc1 ; (else)ylc=lc2$
Kernel   ; rhs=ylc $
Regress  ; lhs=ylc;rhs=one;lcm;pts=2;pds=1$
```





Mixture of Normals

```
+----------------------------------------------+
| Latent Class / Panel LinearRg Model          |
| Dependent variable                  YLC      |
| Number of observations             1000      |
| Log likelihood function       -1960.443      |
| Info. Criterion: AIC =          3.93089      |
| LINEAR regression model                      |
| Model fit with  2 latent classes.            |
+----------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
+--------+Model parameters for latent class 1                         |
|Constant|    4.97029***       .04511814   110.162   .0000             |
|Sigma   |    1.00214***       .03317650    30.206   .0000             |
+--------+Model parameters for latent class 2                         |
|Constant|    1.05522***       .07347646    14.361   .0000             |
|Sigma   |     .95746***       .05456724    17.546   .0000             |
+--------+Estimated prior probabilities for class membership          |
|Class1Pr|     .70003***       .01659777    42.176   .0000             |
|Class2Pr|     .29997***       .01659777    18.073   .0000             |
+--------+------------------------------------------------------------+
| Note: ***, **, * = Significance at 1%, 5%, 10% level.               |
+---------------------------------------------------------------------+
```



Estimating Which Class

Prior probability $\text{Prob}[\text{class}=q]=\pi_q$

Joint conditional density for T_i observations is

$$P(y_{i1},y_{i2},\ldots,y_{i,T_i}\mid \text{class}=q)=\prod_{t=1}^{T_i} f(y_{it}\mid x_{i,t},\boldsymbol{\beta}_q,\sigma_{\varepsilon,q})$$

Joint density for data and class membership is

$$P(y_{i1},y_{i2},\ldots,y_{i,T_i},\text{class}=q)=\pi_q\prod_{t=1}^{T_i} f(y_{it}\mid x_{i,t},\boldsymbol{\beta}_q,\sigma_{\varepsilon,q})$$

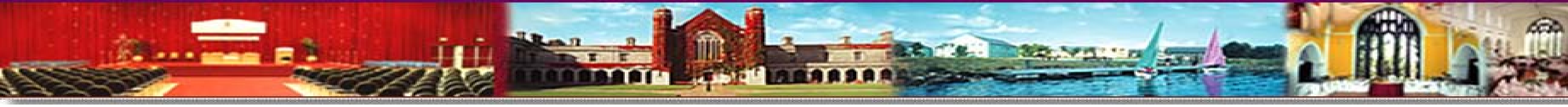
Posterior probability for class, given the data

$$P(\text{class}=q\mid y_{i1},y_{i2},\ldots,y_{i,T_i})=\frac{P(y_{i1},y_{i2},\ldots,y_{i,T_i},\text{class}=q)}{P(y_{i1},y_{i2},\ldots,y_{i,T_i})}=\frac{P(y_{i1},y_{i2},\ldots,y_{i,T_i},\text{class}=q)}{\sum_{j=1}^{J}P(y_{i1},y_{i2},\ldots,y_{i,T_i},\text{class}=q)}$$

Use Bayes Theorem to compute the **posterior probability**

$$w(q\mid \text{data}_i)=P(\text{class}=q\mid y_{i1},y_{i2},\ldots,y_{i,T_i})=\frac{\pi_q\prod_{t=1}^{T_i} f(y_{it}\mid x_{i,t},\boldsymbol{\beta}_q,\sigma_{\varepsilon,q})}{\sum_{q=1}^{Q}\pi_q\prod_{t=1}^{T_i} f(y_{it}\mid x_{i,t},\boldsymbol{\beta}_q,\sigma_{\varepsilon,q})}$$

Best guess = the class with the largest posterior probability.



Posterior for Normal Mixture

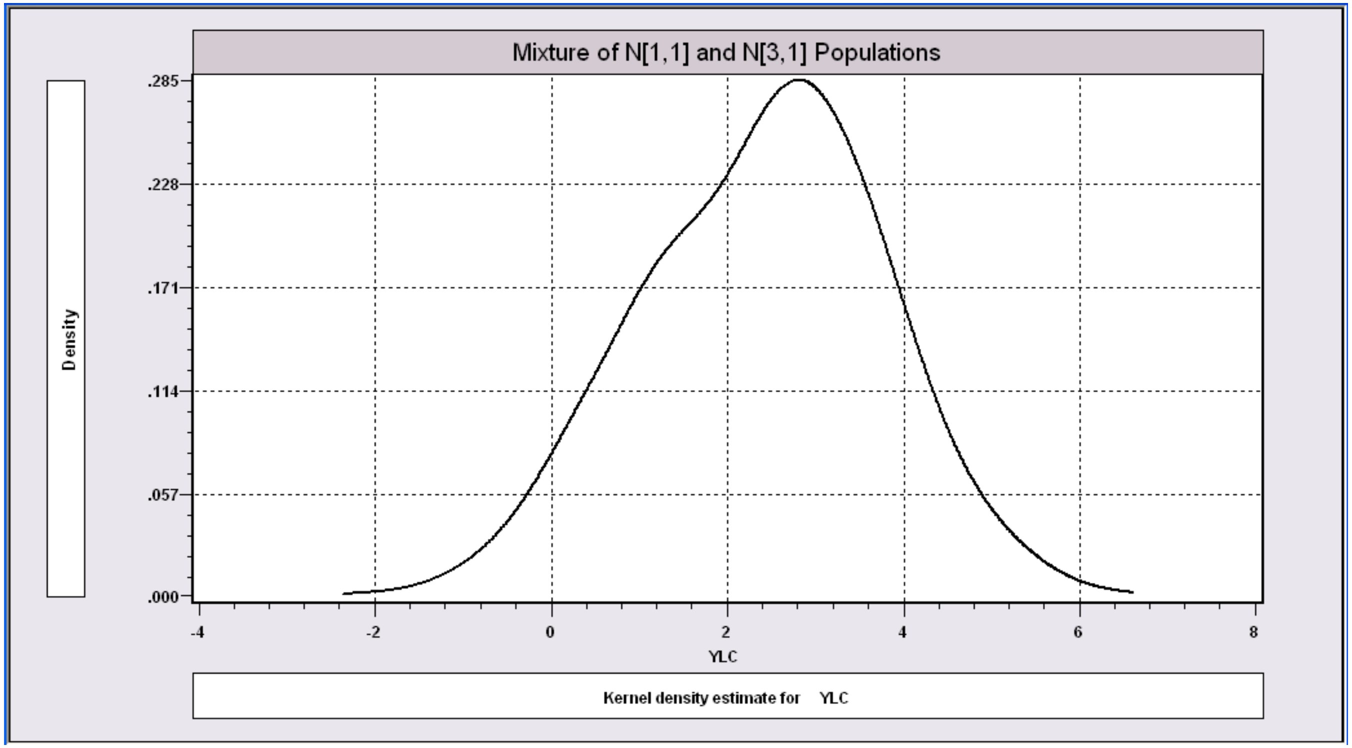
$$\hat{w}(q\,|\,\text{data}_i) = \hat{w}(q\,|\,i) = \frac{\hat{\pi}_q \left[\prod_{t=1}^{T_i} \frac{1}{\hat{\sigma}_q} \phi\left(\frac{y_{it} - \hat{\mu}_q}{\hat{\sigma}_q} \right) \right]}{\sum_{q=1}^{Q} \hat{\pi}_q \left[\prod_{t=1}^{T_i} \frac{1}{\hat{\sigma}_q} \phi\left(\frac{y_{it} - \hat{\mu}_q}{\hat{\sigma}_q} \right) \right]}$$

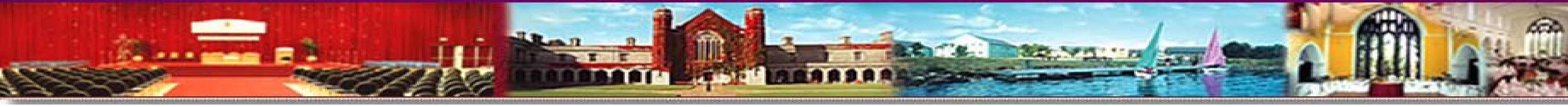
$$= \frac{c_{iq}\hat{\pi}_q}{\sum_{q=1}^{Q} c_{iq}\hat{\pi}_q}$$

Estimated Posterior Probabilities

G1	G2	YLCM
0.000380815	0.999619	0.749454
0.999998	2.4586e-006	6.00763
0.914497	0.0855032	3.3601
0.999983	1.65714e-005	5.53157
0.999998	2.1888e-006	6.03659
0.99946	0.000540485	4.65866
0.997469	0.00253128	4.27006
0.999999	1.40053e-006	6.14779
0.999863	0.000136592	5.00377
0.996985	0.00301525	4.2259
0.999983	1.72564e-005	5.52145
0.746364	0.253636	3.03334
0.860929	0.139071	3.22172
0.999932	6.81582e-005	5.1779
0.000111912	0.999888	0.434583
0.999933	6.66186e-005	5.18362
6.72545e-005	0.999933	0.303494
1	2.90138e-007	6.53927
0.985588	0.0144116	3.8288
0.279522	0.720478	2.51907

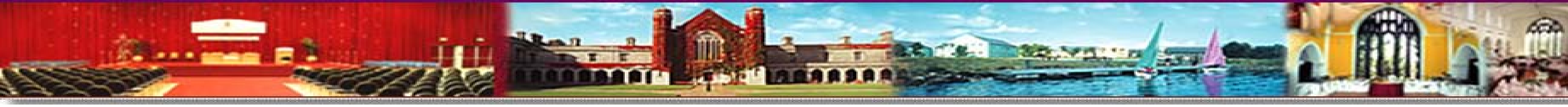
More Difficult When the Populations are Close Together





The Technique Still Works

```
----------------------------------------------------------------------
Latent Class / Panel LinearRg Model
Dependent variable                    YLC
Sample is  1 pds and   1000 individuals
LINEAR regression model
Model fit with  2 latent classes.
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
        |Model parameters for latent class 1
Constant|    2.93611***       .15813       18.568   .0000
   Sigma|    1.00326***       .07370       13.613   .0000
        |Model parameters for latent class 2
Constant|     .90156***       .28767        3.134   .0017
   Sigma|     .86951***       .10808        8.045   .0000
        |Estimated prior probabilities for class membership
Class1Pr|     .73447***       .09076        8.092   .0000
Class2Pr|     .26553***       .09076        2.926   .0034
--------+-------------------------------------------------------------
```

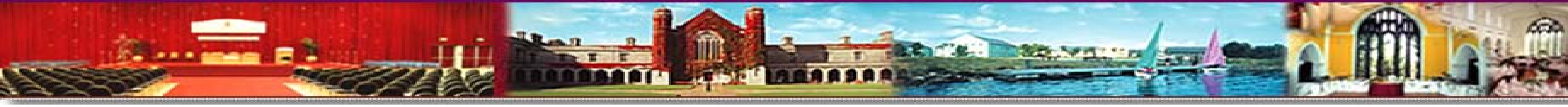


Estimating $E[\boldsymbol{\beta}_i | X_i, y_i, \boldsymbol{\beta}_1 ..., \boldsymbol{\beta}_Q]$

(1) Use $\hat{\boldsymbol{\beta}}_q$ from the class with the largest estimated probability

(2) Probabilistic

$$\hat{\boldsymbol{\beta}}_{\mathbf{i}} = \sum_{q=1}^{Q} \text{Posterior Prob}[\text{class}=q|\text{data}_i]\, \hat{\boldsymbol{\beta}}_q$$



How Many Classes?

(1) Q is not a 'parameter' - can't 'estimate' Q with π and $\boldsymbol{\beta}$

(2) Can't 'test' down or 'up' to Q by comparing log likelihoods. Degrees of freedom for Q+1 vs. Q classes is not well defined.

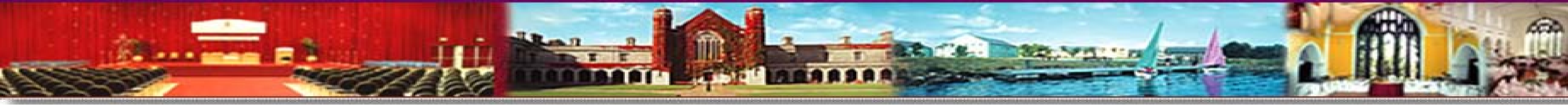
(3) Use AKAIKE IC; AIC = $-2 \times \log L + 2\#\text{Parameters}$.

For our mixture of normals problem,

$\text{AIC}_1 = 10827.88$

$\text{AIC}_2 = 9954.268$ <===

$\text{AIC}_3 = 9958.756$



Latent Class Regression

Assume normally distributed disturbances

$$f(y_{it} \mid \text{class} = q) = \frac{1}{\sigma_{\varepsilon,q}} \phi\left(\frac{y_{it} - \mathbf{x}_{it}'\boldsymbol{\beta}_q}{\sigma_{\varepsilon,q}}\right)$$

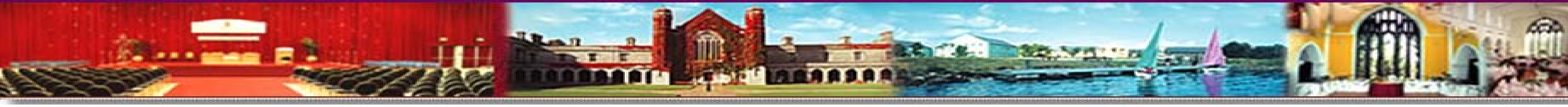
Mixture of normals sets $\mathbf{x}_{it}'\boldsymbol{\beta}_q = \boldsymbol{\mu}_{itq}$.

Baltagi and Griffin's Gasoline Data

World Gasoline Demand Data, 18 OECD Countries, 19 years
Variables in the file are

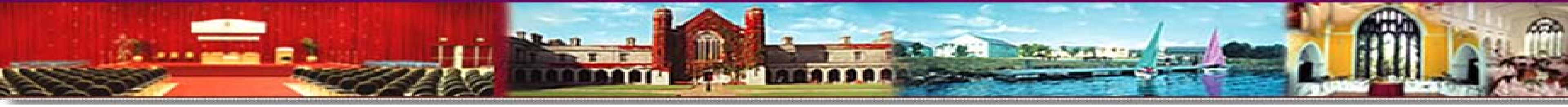
COUNTRY = name of country
YEAR = year, 1960-1978
LGASPCAR = log of consumption per car
LINCOMEP = log of per capita income
LRPMG = log of real price of gasoline
LCARPCAP = log of per capita number of cars

See Baltagi (2001, p. 24) for analysis of these data. The article on which the analysis is based is Baltagi, B. and Griffin, J., "Gasoline Demand in the OECD: An Application of Pooling and Testing Procedures," European Economic Review, 22, 1983, pp. 117-137. The data were downloaded from the website for Baltagi's text.



3 Class Linear Gasoline Model

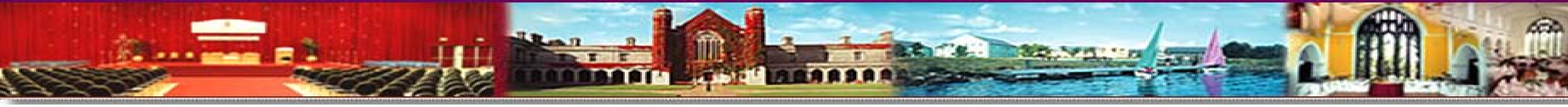
```
-----------------------------------------------------------------------------
Latent Class / Panel LinearRg Model
Dependent variable               LGASPCAR
Log likelihood function         342.49207
Restricted log likelihood          .00000
Chi squared [  13 d.f.]         684.98415
Significance level                 .00000
Estimation based on N =     342, K =   17
Inf.Cr.AIC  =    -651.0 AIC/N =    -1.903
Model estimated: Mar 19, 2012, 17:53:05
Sample is 19 pds and     18 individuals
LINEAR regression model
--------+--------------------------------------------------------------------
        |                 Standard            Prob.       95% Confidence
LGASPCAR|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Model parameters for latent class 1
Constant|     4.39532        2.87948     1.53  .1269     -1.24836  10.03899
LINCOMEP|      .79540*        .47356     1.68  .0930      -.13277   1.72357
   LRPMG|     -.53423        1.41047     -.38  .7049     -3.29870   2.23024
LCARPCAP|     -.53101***      .13124    -4.05  .0001      -.78822   -.27379
   Sigma|      .06884         .26250      .26  .7931      -.44564    .58332
        |Model parameters for latent class 2
Constant|     1.94357***      .25719     7.56  .0000      1.43948   2.44767
LINCOMEP|      .36731***      .14016     2.62  .0088       .09260    .64203
   LRPMG|     -.48399***      .05617    -8.62  .0000      -.59408   -.37391
LCARPCAP|     -.45973***      .06694    -6.87  .0000      -.59093   -.32852
   Sigma|      .07447***      .00676    11.01  .0000       .06122    .08772
        |Model parameters for latent class 3
Constant|     1.12698***      .41975     2.68  .0073       .30429   1.94967
LINCOMEP|      .43120***      .16420     2.63  .0086       .10937    .75302
   LRPMG|     -.41000***      .07130    -5.75  .0000      -.54975   -.27026
LCARPCAP|     -.61625***      .05971   -10.32  .0000      -.73327   -.49922
   Sigma|      .10660***      .01333     8.00  .0000       .08047    .13273
        |Estimated prior probabilities for class membership
Class1Pr|      .29098         .20048     1.45  .1467      -.10195    .68391
Class2Pr|      .31720**       .13142     2.41  .0158       .05963    .57478
Class3Pr|      .39182***      .15139     2.59  .0097       .09509    .68854
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```



Estimated Parameters
LCM vs. Gen1 RPM

	1	2	3	4
1	1.94356	0.367315	-0.483992	-0.459727
2	1.94357	0.367314	-0.483993	-0.459725
3	4.39532	0.795398	-0.534231	-0.531008
4	1.12698	0.431198	-0.410002	-0.616246
5	1.94357	0.367314	-0.483992	-0.459726
6	1.94357	0.367314	-0.483993	-0.459725
7	4.39532	0.795398	-0.534231	-0.531008
8	4.39532	0.795398	-0.534231	-0.531008
9	1.94357	0.367314	-0.483993	-0.459725
10	1.12698	0.431198	-0.410002	-0.616246
11	1.12698	0.431198	-0.410002	-0.616246
12	1.12698	0.431198	-0.410002	-0.616246
13	1.94357	0.367314	-0.483993	-0.459725
14	4.39147	0.79497	-0.534084	-0.531109
15	1.12702	0.431195	-0.410006	-0.616238
16	1.12698	0.431198	-0.410002	-0.616246
17	1.127	0.431197	-0.410004	-0.616242
18	4.39532	0.795398	-0.534231	-0.531008

	1	2	3	4
1	3.7712	0.55187	-0.485435	-0.387265
2	2.55238	0.714802	-0.201465	-0.635849
3	3.28567	0.418864	-0.290662	-0.446393
4	0.291564	0.122538	-0.145417	-0.530267
5	3.1753	1.00926	-0.202796	-0.770054
6	4.13698	0.431209	-0.216905	-0.254492
7	3.58449	0.595853	-0.272624	-0.484243
8	4.16349	0.32981	-0.201541	-0.23423
9	1.30752	0.126914	-0.356194	-0.35955
10	-1.09617	-0.00292814	-0.148817	-0.576652
11	0.914642	0.383942	-0.258079	-0.60601
12	2.68978	0.684036	-0.245217	-0.60324
13	0.23914	-0.146933	-0.155894	-0.313708
14	0.823016	-0.0308741	-0.404606	-0.223692
15	4.60208	0.933709	-0.339252	-0.56949
16	0.601735	0.289237	-0.217959	-0.577492
17	4.03822	0.441826	-0.116262	-0.298403
18	4.21776	0.223542	-0.238852	-0.196738



An Extended Latent Class Model

Class probabilities relate to observable variables (usually demographic factors such as age and sex).

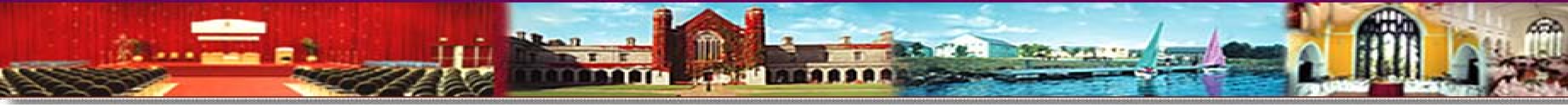
(1) There are Q classes, unobservable to the analyst

(2) Class specific model: $f(y_{it} \mid \mathbf{x}_{it}, \text{class} = q) = g(y_{it}, \mathbf{x}_{it}, \boldsymbol{\beta}_q)$

(3) Conditional class probabilities given some information, $\mathbf{z}_i$)

Common multinomial logit form for prior class probabilities

$$P(\text{class}=q \mid \mathbf{z}_i, \boldsymbol{\delta}) = \pi_{iq} = \frac{\exp(\mathbf{z}_i'\boldsymbol{\delta}_q)}{\sum_{q=1}^{Q} \exp(\mathbf{z}_i'\boldsymbol{\delta}_q)}, \quad \boldsymbol{\delta}_Q = \mathbf{0}$$



LC Poisson Regression for Doctor Visits

```
----------------------------------------------------------------------------
Latent Class / Panel Poisson  Model
Dependent variable                 DOCVIS
Log likelihood function    -71088.40014
Restricted log likelihood -213471.86049
Chi squared [  20 d.f.]    284766.92070
Significance level               .00000
Unbalanced panel has   7293 individuals
POISSON regression model
Model fit with  3 latent classes.
--------+-------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*          Interval
--------+-------------------------------------------------------------------
        |Model parameters for latent class 1
Constant|     3.17505***     .03833    82.84  .0000     3.09993    3.25017
     AGE|      .00734***     .00067    10.93  .0000      .00603     .00866
    HSAT|     -.16554***     .00271   -61.11  .0000     -.17085    -.16023
  INCOME|     -.49866***     .04988   -10.00  .0000     -.59642    -.40090
        |Model parameters for latent class 2
Constant|      .22803***     .07453     3.06  .0022      .08195     .37410
     AGE|      .02479***     .00127    19.49  .0000      .02230     .02728
    HSAT|     -.23278***     .00454   -51.32  .0000     -.24167    -.22389
  INCOME|      .08285        .06470     1.28  .2003     -.04396     .20966
        |Model parameters for latent class 3
Constant|     1.79613***     .03743    47.99  .0000     1.72278    1.86949
     AGE|      .01418***     .00063    22.48  .0000      .01294     .01542
    HSAT|     -.17543***     .00243   -72.17  .0000     -.18020    -.17067
  INCOME|     -.13569***     .03602    -3.77  .0002     -.20630    -.06509
        |Estimated prior probabilities for class membership
   ONE_1|    -1.17685***     .27028    -4.35  .0000    -1.70659    -.64710
FEMALE_1|      .25804***     .09562     2.70  .0070      .07063     .44544
MARRIE_1|     -.20921**      .09901    -2.11  .0346     -.40326    -.01516
  EDUC_1|     -.02913        .02113    -1.38  .1679     -.07054     .01228
   ONE_2|      .28202*       .16781     1.68  .0928     -.04688     .61093
FEMALE_2|     -.71203***     .05922   -12.02  .0000     -.82809    -.59597
MARRIE_2|      .02506        .06549      .38  .7020     -.10329     .15341
  EDUC_2|      .00599        .01290      .46  .6423     -.01929     .03127
   ONE_3|       0.0       .....(Fixed Parameter).....
FEMALE_3|       0.0       .....(Fixed Parameter).....
MARRIE_3|       0.0       .....(Fixed Parameter).....
  EDUC_3|       0.0       .....(Fixed Parameter).....
--------+-------------------------------------------------------------------
+--------------------------------------------------------------+
|  Prior class probabilities at data means for LCM variables  |
|   Class 1      Class 2      Class 3      Class 4      Class 5 |
|    .09543       .45860       .44596       .00000       .00000 |
+--------------------------------------------------------------+
```

Heckman and Singer's RE Model

- Random Effects Model
- Random Constants with Discrete Distribution

(1) There are Q classes, unobservable to the analyst

(2) Class specific model: $f(y_{it} \mid \mathbf{x}_{it}, \text{class} = q) = g(y_{it}, \mathbf{x}_{it}, \alpha_q, \boldsymbol{\beta})$

(3) Conditional class probabilities π_q

Common multinomial logit form for prior class probabilities

to constrain all probabilities to (0,1) and ensure $\sum_{q=1}^{Q} \pi_q = 1$;

multinomial logit form for class probabilities;

$$P(\text{class}=q|\boldsymbol{\delta}) = \pi_q = \frac{\exp(\delta_q)}{\sum_{j=1}^{J} \exp(\delta_q)}, \quad \delta_Q = 0$$

3 Class Heckman-Singer Form

Latent Class / Panel Poisson Model
Dependent variable DOCVIS
Log likelihood function -71303.62317
Restricted log likelihood -213471.86049
Chi squared [20 d.f.] 284336.47464
Significance level .00000

DOCVIS	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Model parameters for latent class 1					
Constant	2.94456***	.02913	101.09	.0000	2.88747	3.00165
AGE	.01146***	.00053	21.56	.0000	.01042	.01250
HSAT	-.17724***	.00177	-100.23	.0000	-.18071	-.17377
INCOME	-.21880***	.03045	-7.19	.0000	-.27848	-.15913
	Model parameters for latent class 2					
Constant	.60562***	.03277	18.48	.0000	.54140	.66984
AGE	.01146***	.00053	21.56	.0000	.01042	.01250
HSAT	-.17724***	.00177	-100.23	.0000	-.18071	-.17377
INCOME	-.21880***	.03045	-7.19	.0000	-.27848	-.15913
	Model parameters for latent class 3					
Constant	1.94821***	.03015	64.61	.0000	1.88911	2.00731
AGE	.01146***	.00053	21.56	.0000	.01042	.01250
HSAT	-.17724***	.00177	-100.23	.0000	-.18071	-.17377
INCOME	-.21880***	.03045	-7.19	.0000	-.27848	-.15913
	Estimated prior probabilities for class membership					
ONE_1	-1.02248***	.26972	-3.79	.0002	-1.55112	-.49383
FEMALE_1	.24984***	.09529	2.62	.0087	.06307	.43662
MARRIE_1	-.31995***	.09893	-3.23	.0012	-.51385	-.12605
EDUC_1	-.03069	.02108	-1.46	.1455	-.07200	.01063
ONE_2	.39483**	.17044	2.32	.0205	.06077	.72889
FEMALE_2	-.73345***	.06013	-12.20	.0000	-.85129	-.61560
MARRIE_2	-.11502*	.06663	-1.73	.0843	-.24562	.01558
EDUC_2	.00774	.01310	.59	.5546	-.01793	.03341
ONE_3	0.0	.....(Fixed Parameter).....				
FEMALE_3	0.0	.....(Fixed Parameter).....				
MARRIE_3	0.0	.....(Fixed Parameter).....				
EDUC_3	0.0	.....(Fixed Parameter).....				

Prior class probabilities at data means for LCM variables

Class 1	Class 2	Class 3	Class 4	Class 5
.09900	.46041	.44059	.00000	.00000

The EM Algorithm

Latent Class is a '**missing data**' model

$d_{i,q} = 1$ if individual i is a member of class q

If $d_{i,q}$ were observed, the complete data log likelihood would be

$$\log L_c = \sum_{i=1}^{N} \log\left\{\sum_{q=1}^{Q} d_{i,q}\left[\prod_{t=1}^{T_i} f(y_{i,t} \mid \text{data}_{i,t}, \text{class} = q)\right]\right\}$$

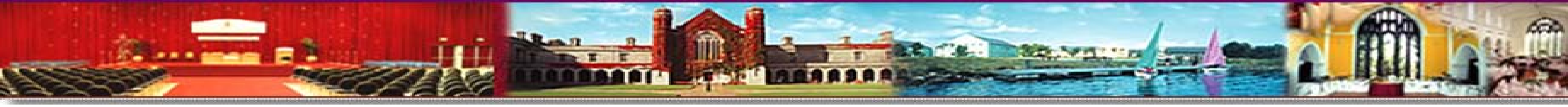
(Only one of the Q terms would be nonzero.)

Expectation - Maximization algorithm has two steps

(1) Expectation Step: Form the 'Expected log likelihood' given the data and a prior guess of the parameters.

(2) Maximize the expected log likelihood to obtain a new guess for the model parameters.

(E.g., http://crow.ee.washington.edu/people/bulyko/papers/em.pdf)



Implementing EM for LC Models

Given initial guesses $\pi_q^0 = \pi_1^0, \pi_2^0, \ldots, \pi_Q^0$, $\boldsymbol{\beta}_q^0 = \boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0, \ldots, \boldsymbol{\beta}_Q^0$

E.g., use 1/Q for each π_q and the MLE of $\boldsymbol{\beta}$ from a one class model. (Must perturb each one slightly, as if all π_q are equal and all $\boldsymbol{\beta}_q$ are the same, the model will satisfy the FOC.)

(1) Compute $\hat{F}(q|i)$ = posterior class probabilities, using $\hat{\boldsymbol{\beta}}^0, \hat{\boldsymbol{\delta}}^0$

Reestimate each $\boldsymbol{\beta}_q$ using a weighted log likelihood

Maximize wrt $\boldsymbol{\beta}_q$ $\sum_{i=1}^{N} \hat{F}_{iq} \sum_{t=1}^{T_i} \log f(y_{it} \mid \mathbf{x}_{it}\boldsymbol{\beta}_q)$

(2) Reestimate π_q by reestimating $\boldsymbol{\delta}$

$\hat{\pi}_q = (1/N)\Sigma_{i=1}^{N}\hat{F}(q|i)$ using old $\hat{\pi}$ and new $\boldsymbol{\beta}$

Now, return to step 1.

Iterate until convergence.