Lectures in panel data econometrics

University of Ferrara May 2013

Antonio Musolesi, GAEL INRA and University of Grenoble

Topics covered:

1) Unobserved heterogeneity and endogeneity

2) Models with heterogeneous slopes

3) Unit roots (and co-integration) with panel

4) Dealing with cross sectional correlation

Day 1.

Introduction, the role of unobserved heterogeneity and endogeneity: Fixed and random effects.

Application: production function.

Day 2.

Using the panel structure of the data to deal with a special kind of endogeneity: selection bias. Policy evaluation in a non randomized framework. Presentation of a research paper.

Models with heterogeneous slopes. Swamy, SURE, Mean Group, Pooled Mean Group, Bayesian approach.

Day 3.

Models with heterogeneous slopes: application to a consumption function

Unit roots (and co-integration) with panel.

Day 4.

Dealing with cross sectional correlation: spatial models, factor models (CCE).

Presentation of a research paper dealing with panel unit roots testing, cross section correlation and heterogeneous slopes.

Lectures in Panel Data Econometrics

Antonio Musolesi

Modelling heterogeneity with panel data

Introduction, the role of unobserved heterogeneity and endogeneity: Fixed and random effects. Application: production function.

Models with heterogeneous slopes. Swamy, SURE, Mean Group, Pooled Mean Group, Bayesian approach.

1 Introduction

- References: Matyas and Sevestre (2008, Pesaran, Hsiao, Heckman, etc), Hsiao (2003), Wooldridge (2010), Baltagi (2008), Arellano (2003)

- Panel data: N individuals and T periods

- Advantages with respect to time series: lower collinearity, more degrees of freedom (thanks to the CS dimension)

- Advantages with respect to cross section: dynamic models (AR), heterogeneous behaviours (thanks to the TS dimension)

2 The central role of heterogeneity in panel data econometrics

- Heterogeneity is a central notion for panel data econometrics
- General model vs pooled model

$$
\begin{aligned}
y_{it} &= \alpha_{it} + x_{it}'\beta_{it} + u_{it}; \\
y_{it} &= \alpha + x_{it}'\beta + u_{it}; \\
x_{it}' &= (x_{1t}, ..., x_{kt})
\end{aligned}
$$

- Example: production function:

$$\begin{aligned} Y_{it} &= AK_{it}^{\beta_1} L_{it}^{\beta_2} \exp(u_{it}) \\ \text{in logs} &: \\ y_{it} &= \alpha + \beta_1 k_{it} + \beta_2 l_{it} + u_{it} \end{aligned}$$

- "the technological progress", α; varying over individuals and/or in time?

- "the production process", β_1 and β_2 varying over individuals and/or in time?

3 Different kinds of heterogeneity

- No Heterogeneity (Pooled model: both intercept and slope coefficients are constant):

$$
\begin{aligned}
y_{it} &= \alpha + x_{it}'\beta + u_{it}, \\
x_{it}' &= (x_{1t}, ..., x_{kt})
\end{aligned}
$$

- Slope coefficients are constant and the intercept varies over individuals (and/or time):

$$
\begin{aligned}
y_{it} &= \alpha_i + x_{it}'\beta + u_{it}, \\
y_{it} &= \lambda_t + x_{it}'\beta + u_{it}, \\
y_{it} &= \alpha_i + \lambda_t + x_{it}'\beta + u_{it}. \text{ Note that } \alpha_{it} = \alpha_i + \lambda_t
\end{aligned}
$$

If α_i (λ_t) are fixed constants$\Rightarrow$ FE model

If α_i (λ_t) are random variables with a probability distribution $\Rightarrow$ RE (error component) / Bayesian model

$\longrightarrow$FE models allow to take into account to omitteed variability, i.e. omitted variables that are time-invariant (sex, ability) for the model with individual effects and individual-invariant (price, optimisme) for the the model with time effects.

Fixed individual effects: $y_{it} = \alpha_i + x'_{it}\beta + u_{it}$;

$\widehat{\beta}$ Within is obtained runnining OLS estimator on:

$$\begin{aligned}(y_{it} - y_{i.}) &= (\alpha_i - \alpha_i) + \left(x'_{it} - x'_{i.}\right)\beta + (u_{it} - u_{i.}) \\ &= \left(x'_{it} - x'_{i.}\right)\beta + (u_{it} - u_{i.})\end{aligned}$$

($\widehat{\alpha}_i$ is obtained as: $\widehat{\alpha}_i = y_{i.} - x'_{i.}\beta$)

$\widehat{\beta}$ first-differences, FD, is otained runnining OLS on:

$$\begin{aligned}(y_{it} - y_{it-1}) &= (\alpha_i - \alpha_i) + \left(x'_{it} - x'_{it-1}\right)\beta + (u_{it} - u_{i.}) \\ &= \left(x'_{it} - x'_{it-1}\right)\beta + (u_{it} - u_{it-1})\end{aligned}$$

for T=2$\Longrightarrow$Within and FD give identical estimates and inference

Within is more efficient when u_{it} are serially uncorrelatted

FD is more efficient when u_{it} follows a random walks

Within gives the same estimates and inference than running OLS on level equation with indidvidual constant terms

According to Woldridge (2010) this is a "coincidence". Recall: exepet when $T \longrightarrow \infty$ we cannot estimate consistently the parameters associated to the individual fixed effects. In many cases (e.g. non linear panel) when estimating a model containing both the fixed effects α_i and the β produces inconsistent estimates of β too: **Incidental parameter problem.**

Greene has shown that this problem does not appear for other kinds of models such as SFA and tobit.

$>>>$ Most pepole use within

Fixed individual effects with **time invariant variables**: $y_{it} = \alpha_i + x'_{it}\beta + z'_i\gamma + u_{it}$; $\widehat{\beta}$ is obtained runnining OLS estimator on:

$$\begin{aligned}(y_{it} - y_{i.}) &= (\alpha_i - \alpha_i) + \left(x'_{it} - x'_{i.}\right)\beta + \left(z'_i - z'_i\right)\gamma \\ &\quad + (u_{it} - u_{i.}) \\ &= \left(x'_{it} - x'_{i.}\right)\beta + (u_{it} - u_{i.})\end{aligned}$$

Consequences:

Positive: The individual fixed effects specification capture also unobserved variables time-invariant ($\widehat{\beta}$ is not biased due to the omission of relevant time-invariant variables)

Negative: when z is observed we can not estimate γ

- All coefficients vary over individuals (time): $y_{it} = \alpha_i + x'_{it}\beta_i + u_{it}$; $y_{it} = \alpha_t + x'_{it}\beta_t + u_{it}$;

Coefficients can be random (Random coefficient model) or fixed (SURE)

- All coefficients vary over individuals and time $y_{it} = \alpha_{it} + x'_{it}\beta_{it} + u_{it}$

Coefficients can be random (Random coefficient model) or fixed. Let consider the case of **fixed coefficients:**

$$
\begin{aligned}
y_{it} &= \alpha_{it} + x'_{it}\beta_{it} + u_{it}, \\
i &= 1, ..., N; \quad t = 1, ..., T
\end{aligned}
$$

$x'_{it} = (x_{1it}, ..., x_{kit})$ and $E(u_{it}) = 0$; $Var(u_{it}) = \sigma^2, \alpha_{it} \in \mathbb{R}$

Identification problem: NT+kNT= NT(k+1) parameters to be estimated but only NT observations: this model ca not be estimated. If we consider the parameters as random variables with a probability distribution with constant means and variance:

$$
\begin{aligned}
\beta_{it} &= \beta + \varepsilon_{it} \\
\alpha_{it} &= \alpha + \upsilon_{it}
\end{aligned}
$$

where ε_{it} and υ_{it} are random variables with with constant means and variances. There is not now an identification problem and the model can be estimated.

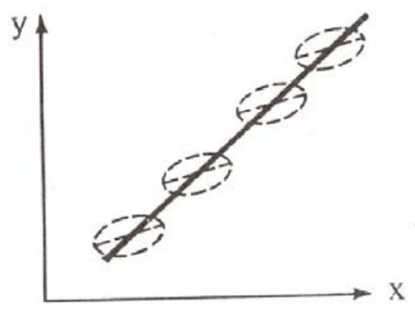
4 Heterogeneity biases

- Trade-off between to much parameters to be estimated (estimation not much accurate) and not enough parameters to be estimated (estimation bias).

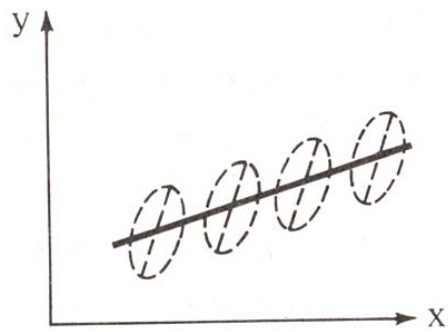
- Case 1:

Estimate pooled model: $y_{it} = \alpha + x'_{it}\beta + u_{it}$

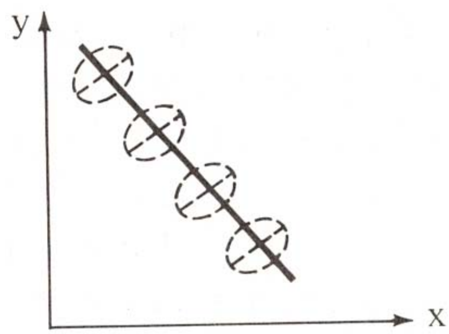
True model: Slope coefficients are constant and the intercept varies over individuals (and/or time) $y_{it} = \alpha_i + x'_{it}\beta + u_{it}$



$\widehat{\beta} > \beta$ (they have the same sign)



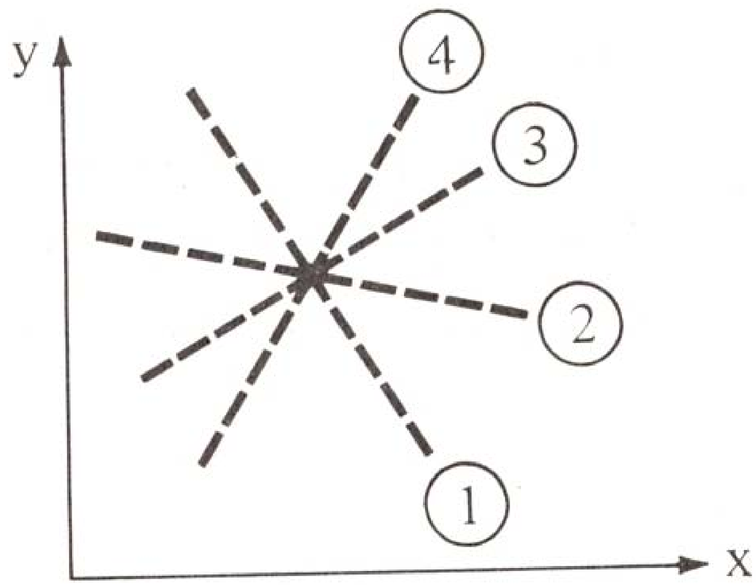
$\widehat{\beta} < \beta$ (same sign)



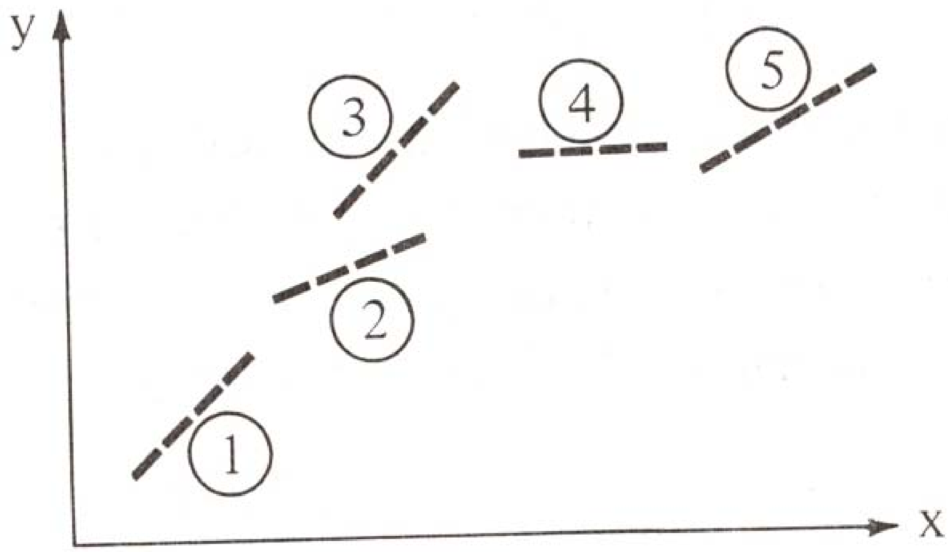
$\widehat{\beta} < \beta$ (different signs)

Remark 1 $\widehat{\beta} - \beta$ *has the same sign that* $Corr\left(\alpha_i, x'_{i.}\right)$

- Case 2: both slope coefficients and intercepts vary over individuals (and/or time): $y_{it} = \alpha_i + x'_{it}\beta_i + u_{it}$



Estimating a pooled model or a model with individual interceps give nonsensical results



Estimating a pooled model give a curvilinear (polynomial) relationship

5 fixed effects, random effects and unobservable

model with individual effects:

$$y_{it} = \alpha_i + x'_{it}\beta + u_{it}$$

FE: $\alpha_i \in \mathbb{R}$

RE: α_i : radom variable such that $E\left(\alpha_i \mid x'_{it}\right) = 0$

- Kind of data. exhaustive data$\longrightarrow$ FE; survey$\longrightarrow$ RE (Conditional Vs unconditional inference)

- Number of parameters. EF: N+K parameters. EA: 2 +K parameters

- statistic /economic arguments. if $E\left(\alpha_i \mid x'_{it}\right) \neq 0$ then $\widehat{\beta}_{FGLS}$ is biased

example: Mundlak (1978) criticised the random effect formulation because it nelects the correlation between α_i and x'_{it}. Production function. y is the output and can be affected by **unobserved manegiarial ability** α_i. Firm with more efficient manegiarial ability tend to produce more and use more inputs x such that $E\left(\alpha_i \mid x'_{it}\right) > 0$ giving a bias of $\widehat{\beta}_{FGLS}$.

α_i are unobservable and Mundlak (1978) suggests to approximate $\mathbf{E}\left(\alpha_i \mid x'_{it}\right)$ by a linear function:

$$\begin{aligned} \alpha_i &= x'_{i.}\gamma + \omega_i \\ with &: \mathbf{E}\left(\omega_i \mid x'_{i.}\right) = 0 \end{aligned}$$

Mundlak test:

$$
\begin{aligned}
H_0 &: \gamma = 0 \\
H_0 &: \gamma \neq 0
\end{aligned}
$$

See also Hausman test

Mundlak test is restrictive imposing linearity

Under Mundlak's hypothesis we can write the **unconstrainded model (u)**:

$$y_{it} = x'_{it}\beta + x'_{i.}\gamma + \omega_i + u_{it}$$

estimation with FGLS$\longrightarrow \widehat{\beta}, \widehat{\gamma}$

and the **constrainded model (c) ($\gamma = 0$):**

$$y_{it} = x'_{it}\beta + \omega_i + u_{it}$$

we can easily test the null $H_0 : \gamma = 0$ using the Ficher's statistic for nested hypotheses:

$$F = \frac{\left(RSQ_c - RSQ_u\right)/k}{RSQ_u/NT - 2k}$$

df(u)=NT-2k; df(c)=NT-k

Mundlak also shows that the GLS estimator of β of the unconstrained model is equal to the within estimator:

$$\widehat{\beta}_{GLSU} = \widehat{\beta}_{WC}$$

in the framework of unconstrained model (fixed effects correlated with x) the BLUE of β is the within estimator (on the constrained model) which correspond to apply GLS to unconstrained model.

However applying GLS to the constrained model yelds a biased estimator

6 Example

- Hoch (1962) estimates a production function using panel data. He estimates 3 specifications with the Cobb-Douglas formulation:

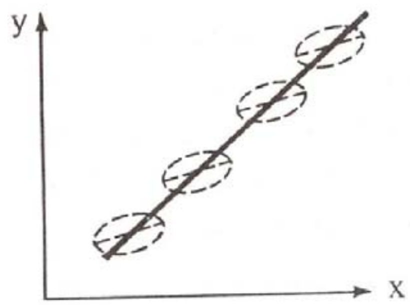
$$
\begin{aligned}
y_{it} &= \alpha + x_{it}'\beta + u_{it} \\
y_{it} &= \lambda_t + x_{it}'\beta + u_{it}, \quad \lambda_t \in \mathbb{R} \\
y_{it} &= \alpha_i + \lambda_t + x_{it}'\beta + u_{it}, \quad \lambda_t, \alpha_i \in \mathbb{R}
\end{aligned}
$$

variables are in log.

Least Squares Estimates of Elasticity Based on Alternative Assumptions

Estimate of elasticity: $\hat{a}_p$	Assumption		
	k_i and k_t^* identical to zero, all i and t (ordinary least squares)	k_i only identical to zero, all i	k_i and k_t^* different from zero (analysis of covariance)
Variable set 1:			
$\hat{a}_1$, labor	.256	.166	.043
$\hat{a}_2$, real estate	.135	.230	.199
$\hat{a}_3$, machinery	.163	.261	.194
$\hat{a}_4$, feed, etc.	.349	.311	.289
Sum $\hat{a}_p$	.904	.967	.726

Look labor input: selection bias?



CHARACTERISTICS OF FIRMS GROUPED ON THE BASIS OF THE FIRM CONSTANT

Characteristic	All firms	Firms classified by value of $\hat{K}_i$				
		$\hat{K}_i$ below .85	$\hat{K}_i$ .85 up to .95	$\hat{K}_i$ .95 up to 1.05	$\hat{K}_i$ 1.05 up to 1.15	$\hat{K}_i$ 1.15 or over
Number of firms in group	63	6	17	19	14	7
Average value of:						
$\hat{K}_i$, firm constant	1.00	0.81	0.92	1.00	1.11	1.26
Y_0, output (dollars)	15,602	10,000	15,570	14,690	16,500	24,140
Y_1, labor (dollars)	3,468	2,662	3,570	3.346	3,538	4,280
Y_4, feed and fertilizer (dollars)	3,217	2,457	3,681	3,064	2,621	5,014
Y_5, current expenses (dollars)	2,425	1,538	2,704	2,359	2,533	2,715
Y_6, fixed capital (dollars)	3,398	2,852	3,712	3,067	3,484	3,996
Profit (dollars)	3,094	491	1,903	2,854	4,324	8,135
Profit/output	0.20	0.05	0.12	0.19	0.26	0.33

The estimated individual fixed effects, $\widehat{\alpha}_i$, $(\widehat{K_i})$ are positively related with inputs, for example labor input:

$$Corr\left(\widehat{\alpha}_i, L_i\right) > 0$$

we can write the Mundlak auxilary regression $\alpha_i = x_{i.}^{\prime}\gamma + \omega_i$ and test $H_0 : \gamma = 0$. we reject H$_0$ and choose FE formulation.

7 Choosing between individual and time effects

$$
\begin{aligned}
y_{it} &= \alpha + x'_{it}\beta + u_{it} && (1)\\
y_{it} &= \alpha_i + x'_{it}\beta + u_{it}, && (2)\\
y_{it} &= \lambda_t + x'_{it}\beta + u_{it}, && (3)\\
y_{it} &= \alpha_i + \lambda_t + x'_{it}\beta + u_{it}, && (4)
\end{aligned}
$$

$1 \subset 2$,$1\subset 3$, $1 \subset 4$,$2\subset 4$,$3\subset 4$

2 (3) is not nested in 3 (2)

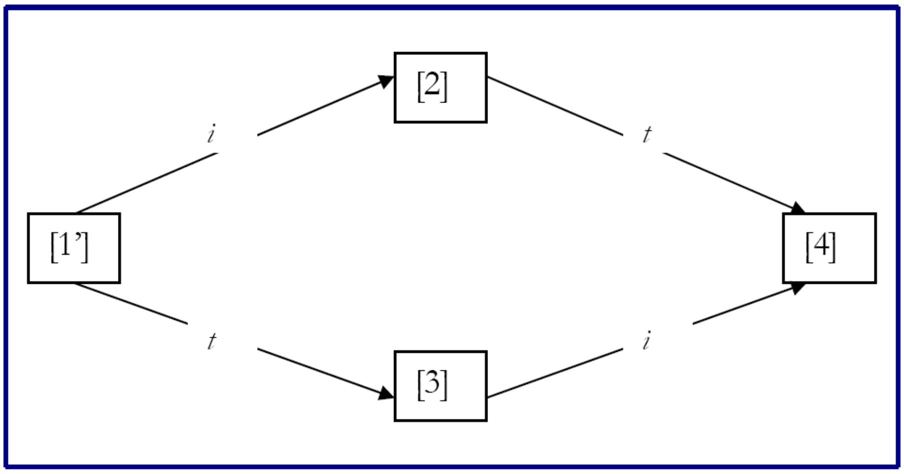
For nested models we can use the Fisher test:

$$F = \frac{(RSQ_c - RSQ_u)\,/df_c - df_u}{RSQ_u/df_u} \sim F\,(df_c - df_u, df_u)$$

for testing 1 vs 2 (H$_0$: $\alpha_1 = \alpha_2 = ... = \alpha_N = \alpha$), we can use:

$$F = \frac{(RSQ_1 - RSQ_2) / [NT - k - 1 - (NT - k - N)]}{RSQ_2 / (NT - k - 1)}$$

We can use the following strategy:



8 Translog

in logs

$y_{it} = \alpha_i + \beta_k k_{it} + \beta_l l_{it} + \beta_{kk}\frac{1}{2}k_{it}^2 + \beta_{ll}\frac{1}{2}l_{it}^2 + \beta_{kl}k_{it}l_{it} + \varepsilon_{it}$

$E_{Y,K} = \frac{\partial y_{it}}{\partial k_{it}} = \beta_k + 2\frac{1}{2}\beta_{kk}k_{it} + \beta_{kl}l_{it}$

$\frac{\partial y_{it}}{\partial k_{it}}$ varies across cross sections

average of the "individual elasticities"

mean, median, etc, Kernel density estimation

average elasticity

$\left(\overline{\frac{\partial y_{it}}{\partial k_{it}}}\right) = \beta_k + 2\frac{1}{2}\beta_{kk}\overline{k} + \beta_{kl}\bar{l}$

9 Variable slopes

Let consider the **general model** :$y_{it} = x'_{it}\beta_{it} + u_{it}$, $i = 1, ..., N$; $t = 1, ..., T$. If parameters are **fixed** there are NTK parameters to be estimated and only NT observations. this model can not be estimated. For easy of esposition we consider the following model in which **coefficients vary across cross sections but are constant over time:**

$$y_{it} = x'_{it}\beta_i + u_{it},$$

β_i can be treated as fixed or random. We can also rewrite the model as :

$$y_{it} = x'_{it}(\beta + \alpha_i) + u_{it}$$

$\beta = (\beta_1, ..., \beta_k)'$ can be viewed as the common mean coefficient vector and $\alpha_i = (\alpha_{i1}, \alpha_{i2}, ..., \alpha_{ik})'$ are the individual deviations from the common mean β.If the interest is inthe performance of individual units from data, the α_i can be treated as fixed. If the population characteristics are of interests: α_i random.

10 Fixed-coefficient model (SURE)

Kronecker product

Definition: Two matrices A et B of dimension (m, n) et (p, q), the Kronecker product $A \otimes B$, is a matrix of dimension (mp, nq) defined by: $A \otimes B = (a_{ij}B)$ where a_{ij} are the elements of matrix A.

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 2 & 1 \end{pmatrix} \otimes \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b & 2a & 2b \\ c & d & 2c & 2d \\ 3a & 3b & 4a & 4b \\ 3c & 3d & 4c & 4d \\ 2a & 2b & a & b \\ 2c & 2d & c & d \end{pmatrix}.$$

1) $(A + B) \otimes C = (A \otimes C) + (B \otimes C)$.

2) $(A \otimes B).(C \otimes D) = A.C \otimes B.D$

4) $(A \otimes B)' = A' \otimes B'$.

SURE (Seemingly unrelated regressions equations)

proposed by Zellner (1962). The N equations are relied via the cross-section covariance of the error terms. In a model with N equations (each one with T observations) and k regressors we can write **the equation i as**: $y_i = X_i\beta_i + \varepsilon_i \qquad i = 1, ..., N$

where y_i ε_i are $T \times 1$ vectors, X_i is a $T \times k$ matrix and β_i is a $K \times 1$ vector. The System can be write as:

$$\begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{pmatrix} = \begin{pmatrix} X_1 & 0 & . & . & . & 0 \\ 0 & X_2 & . & . & . & 0 \\ & & & & . & \\ & & & & . & \\ & & & & . & \\ 0 & 0 & . & . & . & X_n \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{pmatrix}$$

which can be written compactly:

$$y = X\beta + \varepsilon$$

y is $NT \times 1$, $\qquad X$ is $NT \times kN$,

β is $kN \times 1$, $\qquad \varepsilon$ is $NT \times 1$.

$$y = X\beta + \varepsilon$$

The hypotheses are

1. $E(\varepsilon) = \mathbf{0}$

2. $E(\varepsilon\varepsilon') = \underset{NT\times NT}{\Omega} = \underset{N\times N}{\boldsymbol{\Sigma}} \otimes \underset{T\times T}{\mathbf{I}}$

$$= \begin{pmatrix} \sigma_{11}\mathbf{I} & \sigma_{12}\mathbf{I} & . & . & . & \sigma_{1n}\mathbf{I} \\ \sigma_{21}\mathbf{I} & \sigma_{22}\mathbf{I} & . & . & . & \sigma_{2n}\mathbf{I} \\ & & & & . & \\ & & & & . & \\ & & & & . & \\ \sigma_{n1}\mathbf{I} & \sigma_{n2}\mathbf{I} & . & . & . & \sigma_{nn}\mathbf{I} \end{pmatrix}$$

with, $\boldsymbol{\Sigma} = \left\{\sigma_{ij}\right\}$

3. $\left(X'\Omega^{-1}X\right)^{-1}$ is non singular,

$p\lim\left(\dfrac{X'\Omega^{-1}X}{T}\right)$ is finite and non singular

(Recall: an n-by-n (square) matrix $\mathbf{A}$ is called nonsingular if there exists an n-by-n matrix $\mathbf{B}$ such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$)

The GLS estimator:

$$\begin{aligned}\widetilde{\beta} &= \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}y \\ &= \left(X'\left(\mathbf{\Sigma}^{-1}\otimes\mathbf{I}\right)X\right)^{-1}X'\left(\mathbf{\Sigma}^{-1}\otimes\mathbf{I}\right)y\end{aligned}$$

is BLUE (Best linear Unbiased Estimator) while OLS is only consistent.

If the equations are not relied, $\sigma_{ij} = 0$ for $i \neq j \Rightarrow \widetilde{\beta} = \widehat{\beta}$.

To test the null Hp of no cross-section correlation (zero non diagonal elements of $\mathbf{\Sigma}$) $\longrightarrow$Breush and Pagan (1980) test

It is possible to constrain $\beta_i = \beta$ for all i with the exeption of constant term$\longrightarrow$model similar to FE

Theorem If $\sigma_{ij} = 0$ for $i \neq j$ then $\widetilde{\beta} = \widehat{\beta}$.

Proof. If $\sigma_{ij} = 0$ for $i \neq j$ then $\mathbf{\Sigma}$ is a diagonal matrix:$\mathbf{\Sigma} = diag\left(\sigma_{11}...\sigma_{nn}\right)$ and

$$\mathbf{\Sigma}^{-1} = diag\left(1/\sigma_{11}...1/\sigma_{nn}\right)$$

$$X'\left(\mathbf{\Sigma}^{-1} \otimes \mathbf{I}\right) X \;\; is\; a\; bloc - diagonal\; matrix :$$

$$X'\left(\mathbf{\Sigma}^{-1} \otimes \mathbf{I}\right) X = diag\left\{\left(1/\sigma_{11}\right) X_1' X_1 ... \left(1/\sigma_{nn}\right) X_n' X_n\right\}$$

and $\left(X'\left(\mathbf{\Sigma}^{-1} \otimes \mathbf{I}\right) X\right)^{-1} = diag\left\{\sigma_{11} X_1' X_1 ... \sigma_{nn} X_n' X_n\right\}$,
$X'\left(\mathbf{\Sigma}^{-1} \otimes \mathbf{I}\right) y$ is also a bloc vector with the i element eauqls to $1/\sigma_{ii} X_i' y_i$.
consequently , $\widetilde{\beta}$ is a bloc vector:

$$\begin{aligned}
\widetilde{\beta} &= \sigma_{ii}\left(X_i' X_i\right)^{-1}\left(1/\sigma_{ii}\right) X_i' y_i \\
&= \left(X_i' X_i\right)^{-1} X_i' y_i \\
&= \widehat{\beta}
\end{aligned}$$

■

11 Random-coefficient model

Random-coefficient model consider each coefficient as a random variable with a probability distribution. The coefficients are treated as having constant means and variance covariance. **Swamy (1970) formulation:**

$$\begin{aligned}
\underset{T\times1}{y_i} &= \underset{T\times K}{X_i}\underset{K\times1}{\beta_i} + \underset{T\times1}{u_i} \qquad i = 1, ..., N \\
\underset{K\times1}{\beta_i} &= \underset{K\times1}{\beta} + \underset{K\times1}{\alpha_i} \\
\mathbf{E}\left(\underset{K\times1}{\alpha_i}\right) &= 0, \\
\mathbf{E}\left(\underset{K\times K}{\alpha_i\alpha_i'}\right) &= \begin{cases} \Delta \text{ if i=j} \\ 0 \text{ if i} \neq j \end{cases} = \begin{pmatrix} \Delta & & \\ & \Delta & \\ & & \Delta \end{pmatrix} \\
&\Rightarrow \mathbf{E}(\beta_i) = \beta, \\
&\Rightarrow \mathbf{E}\left(\beta_i\beta_i'\right) = \begin{cases} \Delta \text{ if i=j} \\ 0 \text{ if i} \neq j \end{cases} \\
\mathbf{E}\left(\underset{K\times K}{\underset{K\times1}{x_{it}}\underset{1\times K}{\alpha_i'}}\right) &= 0
\end{aligned}$$

$$\mathbf{E}\left(\underset{T\times 1}{u_i}\right) = 0,$$

$$\mathbf{E}\left(\underset{T\times T}{u_i u_j'}\right) = \begin{cases} \sigma_i^2 \mathbf{I}_T \text{ if i=j} \\ 0 \text{ if i} \neq j \end{cases} = \begin{pmatrix} \sigma_1^2 & & \\ & \sigma_i^2 & \\ & & \sigma_N^2 \end{pmatrix}$$

Stacking all NT observations

$$\underset{NT\times 1}{y} = \underset{NT\times K}{X}\underset{K\times 1}{\beta} + \underset{NT\times NK}{\widetilde{X}}\underset{NK\times 1}{\alpha} + \underset{NT\times 1}{u}$$

$$
\begin{aligned}
\underset{NT\times 1}{y} &= \underset{NT\times K}{X}\underset{K\times 1}{\beta} + \underset{NT\times NK}{\widetilde{X}}\underset{NK\times 1}{\alpha} + \underset{NT\times 1}{u} \\
\underset{NT\times 1}{y} &= \left(y_1', ..., y_N'\right)' \\
\underset{NT\times K}{X} &= \begin{pmatrix} \underset{T\times k}{X_1} \\ . \\ X_N \end{pmatrix} \\
\underset{NT\times NK}{\widetilde{X}} &= \begin{pmatrix} \underset{T\times k}{X_1} & & 0 \\ & X_2 & \\ 0 & & X_N \end{pmatrix} = diag\left(X_1, ..., X_N
\end{aligned}
$$

$$
\begin{aligned}
\underset{NK\times 1}{\alpha} &= \left(\underset{K\times 1}{\alpha_1'}, ..., \alpha_N'\right)' \\
\underset{NT\times 1}{u} &= \left(\underset{T\times 1}{u_1'}, ..., u_N'\right)' \\
\text{let define} &: \quad \underset{NT\times 1}{v} = \widetilde{X}\alpha + u \\
\mathbf{E}\left(\underset{NT\times NT}{vv'}\right) &= \begin{pmatrix} \Phi_1 & & 0 \\ & \Phi_i & \\ 0 & & \Phi_N \end{pmatrix} \\
\text{with} &: \quad \Phi_i = X_i \mathbf{\Delta} X_i' + \sigma^2 \mathbf{I}_T
\end{aligned}
$$

12 Estimation of the Swamy's model

- Under Swamy's assumption, the regression of y on X will yeld an unbiased and consistent estimator of β (see Hsiao, 2003). It is however inefficient and usual LS computation of the variance-covariance matrix of the estimator is incorrect leading invalid inference. The BLUE estimator of β is the GLS estimator:

$$
\begin{aligned}
\widehat{\beta}_{GLS} &= \sum_{i=1}^{N} W_i \widehat{\beta}_i \\
with &: \widehat{\beta}_i = \left(X_i' X_i\right)^{-1} X_i' y_i \\
and &: W_i = \left\{ \sum_{i=1}^{N} \left[\Delta + \sigma_i^2 \left(X_i' X_i\right)^{-1} \right]^{-1} \right\}^{-1} \\
& \quad \times \left[\left[\Delta + \sigma_i^2 \left(X_i' X_i\right)^{-1} \right]^{-1} \right]
\end{aligned}
$$

- $\widehat{\beta}_{GLS}$ is a weigheted average of the LS estimators $\widehat{\beta}_i$ $(i = 1, ..., N)$

with th e weights W_i that are inversely proportional the their variances-covariance matrices (var $(\widehat{\beta}_i) = \sigma_i^2 \left(X_i' X_i\right)^{-1}$).

- The variance-covariance matrix of $\widehat{\beta}_{GLS}$ is

$$
\begin{aligned}
Var\left(\widehat{\beta}_{GLS}\right) &= \left(\sum_{i=1}^{N} X_i' \Phi_i^{-1} X_i\right)^{-1} \\
&= \left\{\sum_{i=1}^{N} \left[\Delta + \sigma_i^2 \left(X_i' X_i\right)^{-1}\right]^{-1}\right\}^{-1}
\end{aligned}
$$

- Problem: this estimator is impossible to use since Δ and σ_i^2 are unknown. Swamy proposes uses the least squares estimators $\widehat{\beta}_i$ and their residuals to obtain $\widehat{u}_i = y_i - X_i \widehat{\beta}_i$ to obtain unbiased estimators of Δ and σ_i^2.

- Substituting $\widehat{\Delta}$ and $\widehat{\sigma}_i^2$ for Δ and σ_i^2 yelds an asymptototically normal and efficient estimator of β

- Hsiao (1975) considers: $\beta_{it} = \beta + \epsilon_{it} = \beta + \alpha_i + \lambda_t$.

13 Testing for coefficient variation

We can test for coefficient variation by testing wheter or not the coefficient-vector β_i are all equal:

$$H_0 : \beta_1 = \beta_2 = ... = \beta_N = \beta$$

If the different cross-sectional units have the same variance, $\sigma_i^2 = \sigma^2$, the conventional F test statistics can be used (as for testing fixed effects):

If σ_i^2 differs across i (as assumed by Swamy), we can apply a modified test statistic:

$$F^{*}=\sum_{i=1}^{N}\frac{\left(\widehat{\beta}_i-\widehat{\beta}^{*}\right)' X_i' X_i \left(\widehat{\beta}_i-\widehat{\beta}^{*}\right)}{\widehat{\sigma}_i^2}$$

where

$$\widehat{\beta}^{*}=\left[\sum_{i=1}^{N}\frac{1}{\widehat{\sigma}_i^2}X_i' X_i\right]^{-1}\left[\sum_{i=1}^{N}\frac{1}{\widehat{\sigma}_i^2}X_i' y_i\right]$$

Under H_0 ,is asymptotically $\chi^2\left(K(N-1)\right)$

14 Fixed or random coefficients

- wheter we are making inference conditional on individual charactersitics: fixed coefficients. or making unconditional inferences on the population characteristics: random coefficients

- Extending his work on variable-coefficient model, Mundlak (1978b) has questioned the random coefficients formulation since **the variable coefficients can be correlated with the explanatory variables (CRC)**$\rightarrow$ GLS estimator of the mean coefficient vector β will b e biased.

  Wooldridge (2005), Hsiao (2011)$>$ FE "could be" consistent in that case

- Mundlak (1978b) suggests to introduce auxiliary regressions of the type

$$\begin{aligned} \alpha_i &= x'_{i.}\gamma + \omega_i \\ with &: \mathbf{E}\left(\omega_i \mid x'_{i.}\right) = 0 \end{aligned}$$

15 Dynamic heterogeneous panels

Let a dynamic (AR1) homogeneous panel data model:

$$y_{it} = \alpha_i + \rho y_{it-1} + x'_{it}\beta + u_{it}$$

the inclusion of y_{it-1} leads to inconsistent estimates both in the FE and in the RE framework beaucause of the correlation between the error term and the regressors (the inconsistency does not occur in the same way in FE and RE formulation)$\rightarrow$ IV approach $\rightarrow$ GMM (Arellano-Bond, Blundell-Bond etc)

The problem is more transparent in the RE formulation in which the lagged dependent variable is correlated with α_i

What appens if ρ is allowed to vary across individuals?

16 Alternative estimators of dynamic panels

- small N and Large T. "Time series literature"

If N=1 (Time Series): "traditional"ARDL Vs cointegration

N>1 : SURE. allows for cross-section correlation

- large N and small T. "Dynamic panel literature"

Arellano-Bond, Anderson-Hsiao, Arellano-Bover, Blundell-Bond... GMM approach

- Large N and Large T. "Heterogeneous dynamic panels"

Why?

For large T, Pesaran-Smith (1995) show showed that the GMM approach can produce inconsistent and very misleading estimates

Coice between fixed and random coefficients (Sampling Vs bayesian approach)

Fixed coefficients: Pesaran et al; 1999: PMG

Random coefficients (sampling): Pesaran-Smith, 1995: MG

Random coefficiens (Bayes): Hsiao-Tahmiscioglu (1997), Hsiao et al. (1999)

17 The Mean Group estimator

Pesaran-Smith, 1995: MG. Consider a dynamic model of the form:

$$y_{it} = \rho_i y_{it-1} + x_{it}' \beta_i + u_{it}$$

Let $\theta_i = (\rho_i, \beta_i')'$ and assume that θ_i is independently distributed across i with:

$$\begin{aligned} E(\theta_i) &= \theta \\ E\left((\theta_i - \theta)(\theta_i - \theta)'\right) &= \Delta \end{aligned}$$

Random coefficients: $\theta_i = \theta + a_i$. The Mean Group estimator is:

$$\widehat{\theta}_{MG} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\theta}_i$$

The $\widehat{\theta}_{MG}$ is consistent when both T and N$\longrightarrow \infty$. Pesearan-Smith make some quite strong assumptions: The group-specific parameters are distributed indipendently of the regressors and the regressors are strictly exogenous

18 The Pooled Mean Group estimator

Pesaran et al. (1999).

The PMG estimator is an "intermediate estimator: it allows the intercepts, short-run coefficients and error vriances to differ freely across cross-section while the long-run coefficients are constrained to be the same.

ARDL$(p, q, q, ..., q)$ with individual fe:

all quantities are scalars exept x'_{it} and δ that are K-vector

$$y_{it} = \sum_{j=1}^{p} \lambda_{ij} y_{it-j} + \sum_{j=0}^{q} x'_{it-j} \delta_{ij} + \alpha_i + \varepsilon_{it}$$

Re-parametrisation

$$\Delta y_{it} = \phi_i y_{it-1} + x'_{it} \beta_i + \sum_{j=1}^{p-1} \lambda^*_{ij} \Delta y_{it-j} + \sum_{j=0}^{q-1} \Delta x'_{it-j} \delta^*_{ij} + \alpha_i +$$

with

$$\begin{aligned}\phi_i &= -\left(1 - \sum_{j=1}^{p} \lambda_{ij}\right)\\ \beta_i &= \sum_{j=0}^{q} \delta_{ij}\\ \lambda^*_{ij} &= -\sum_{m=j+1}^{p} \lambda_{im} \qquad j = 1, 2, ..., p-1\\ \delta^*_{ij} &= -\sum_{m=j+1}^{q} \delta_{im} \qquad j = 1, 2, ..., q-1\end{aligned}$$

stack the time series observation for each group:

$$\Delta y_i = \phi_i y_{i-1} + X_i \beta_i + \sum_{j=1}^{p-1} \Delta y_{i-j} \lambda^*_{ij} + \sum_{j=0}^{q-1} \Delta X_{i-j} \delta^*_{ij} + \iota_T \alpha_i \tag{5}$$

y_i and ε_i are $T \times 1$, X_i is $T \times k$, β_i and δ^*_{ij} are $k \times 1$, ι_T is $T \times 1$ et α_i and λ^*_{ij} are scalars.

Hypotheses

1. ε_{it} are indipendently distributed across i: $\varepsilon_{it} i.i.d \left(0, \sigma_i^2\right)$, with $\sigma_i^2 > 0$ and $\varepsilon_{it} \perp x_{it}$.

$\varepsilon_{it} \perp x_{it}$ is needed for consistency ot the estimator of short-run parameters

2. $\phi_i < 0$ and hence there exists a long-run relationship defined by

$$y_{it} = -\left(\beta_i / \phi_i\right) x_{it} + \eta_{it}$$

where η_{it} is a stationary process.

3. The long run coefficients θ_i=-β_i / ϕ_i are homogeneous across individuals, i.e. $\theta_i = \theta$.

..

..

ML approach to the estimation

19 Bayesian Approach

$$
\begin{aligned}
\underset{T\times 1}{y_i} &= \underset{T\times K}{X_i}\,\underset{K\times 1}{\beta_i} + \underset{T\times 1}{u_i} \qquad i = 1, ..., N \\
\beta_i &= \beta + \alpha_i
\end{aligned}
$$

All quantities including the parameters are random variables

Prior probability distributions are introduced for the parameters (state of knowledge or ignorance about the parameters before obtaining the data)

α_i is normally distributed with mean 0 and covariance matrix $\boldsymbol{\Delta}$

. . . . . . SWAMY. . . . . .

This prior information is combined with the model and the data to revise the probability distribution of α_i, which is called **posterior distribution**

Inference is made from this posterior distribution

Empirical Bayes, Iterative empirical Bayes, Hierarchical Bayes

20 Swamy and Bayesian: some remarks

The Swamy formulation of β_i having mean β and covariance $\mathbf{\Delta}$ is equivalent to specifying an informative prior for the parameters β_i

Swamy: sampling approach predictor of $\beta_i = \beta + \alpha_i$ is the LS estimator. The Bayes predictor of β_i is a weighted average between the LS estimator of β_i and the overall mean $\beta \Rightarrow$ the individual estimates β_i shrinks towards the mean β

As $T \rightarrow \infty$ the Bayes estimator approaches the least squares estimator $\widehat{\beta_i}$

21 Some empirical studies

Baltagi-Bresson-Pirotte (2002)

Abstract

Maddala et al. [Journal of Business and Economic Statistics, 15 (1997) 90] obtained short-run and long-run elasticities of energy demand for each of 49 US states over the period 1970–1990. They showed that heterogeneous time series estimates for each state yield inaccurate signs for the coefficients, while panel data estimates are not valid because the hypothesis of homogeneity of the coefficients was rejected. Their preferred estimates are those obtained using the shrinkage estimator. This paper contrasts the out-of-sample forecast performance of heterogeneous, panel and shrinkage estimators using the Maddala et al. [Journal of Business and Economic Statistics 15 (1997) 90] electricity and natural gas data sets. Our results show that the homogeneous panel data estimates give the best out-of-sample forecasts.

Maddala et al. (1997) considered the following standard dynamic linear regression (DLR) model for energy demand

$$\begin{aligned} y_{i,t} = {} & \beta_{i,0} + \beta_{i,1} y_{i,t-1} + \beta_{i,2} x_{1,i,t} + \beta_{i,3} x_{1,i,t-1} + \beta_{i,4} x_{2,i,t} + \beta_{i,5} x_{2,i,t-1} \\ & + \beta_{i,6} x_{3,i,t} + \beta_{i,7} x_{4,i,t} + \beta_{i,8} x_{5,i,t} + u_{i,t} \end{aligned}$$

where $i = 1, 2, \ldots, 49$ (states) and $t = 2, 3, \ldots, 21$ (years) spanning the period 1970–1990. The variables for the electricity regression are:

- $y_{i,t}$, the logarithm of residential electricity per capita consumption;
- $x_{1,i,t}$, the logarithm of real per capita personal income;
- $x_{2,i,t}$, the logarithm of real residential electricity price;
- $x_{3,i,t}$, the logarithm of real residential natural-gas price;
- $x_{4,i,t}$, heating degree days; 0
- $x_{5,i,t}$, cooling degree days.

For the natural-gas regression, we have:

- $y_{i,t}$, the logarithm of residential natural-gas per capita consumption;
- $x_{1,i,t}$, the logarithm of real per capita personal income;
- $x_{2,i,t}$, the logarithm of real residential natural-gas price;
- $x_{3,i,t}$, the logarithm of real residential electricity price;

Comparison of forecast performance for US electricity demand

Ranking	1st year		5th year		5-year average	
	Estimator	RMSE[b]	Estimator	RMSE[b]	Estimator	RMSE[b]
1.	Empirical iterative Bayes	3.032	GLS	4.733	GLS	4.207
2.	Iterative Bayes	3.230	Within	5.092	Within	4.482
3.	GLS	3.252	2SLS-KR[a]	5.552	Within-2SLS	4.817
4.	Individual ML	3.352	2SLS	5.609	2SLS-KR[a]	4.892
5.	Within	3.355	OLS	5.676	Empirical iterative Bayes	4.981
6.	Within-2SLS	3.441	Within-2SLS	5.829	2SLS	5.019
7.	FD2SLS-KR[a]	3.603	Empirical iterative Bayes	6.544	OLS	5.020
8.	Empirical Bayes	3.622	Individual ML	8.420	Iterative Bayes	5.998
9.	FDGMM	3.624	Iterative Bayes	8.774	FDGMM	6.236
10	FD2SLS[a]	3.842	FDGMM	8.917	Individual ML	6.465
11.	2SLS-KR[a]	3.846	Empirical Bayes	9.407	FD2SLS[a]	7.045
12.	OLS	3.879	FD2SLS[a]	10.826	Empirical Bayes	7.275
13.	2SLS	3.905	GMM	10.899	GMM	8.084
14.	Individual OLS	4.886	FD2SLS-KR[a]	15.157	FD2SLS-KR[a]	8.598
15.	GMM	5.286	Individual OLS	15.594	Indivividual OLS	11.913
16.	Individual 2SLS	5.521	Swamy	18.933	Swamy	16.239
17.	Swamy	12.427	Average 2SLS	21.829	Individual 2SLS	16.397
18.	Average 2SLS	16.408	Average OLS	21.830	Average 2SLS	19.665
19.	Average OLS	17.208	Individual 2SLS	24.288	Average OLS	20.009

Comparison of forecast performance for US natural-gas demand

Ranking	1st year		5th year		5-year average	
	Estimator	RMSE[b]	Estimator	RMSE[b]	Estimator	RMSE[b]
1.	Within-2SLS	6.062	Within-2SLS	10.049	Within-2SLS	8.730
2.	OLS	6.460	OLS	10.517	OLS	9.060
3.	GLS	6.682	GLS	10.788	GLS	9.295
4.	Within	7.071	Within	11.132	Within	9.637
5.	2SLS	7.084	2SLS	12.564	2SLS	10.325
6.	Iterative Bayes	7.130	2SLS-KR[a]	14.440	2SLS-KR[a]	11.036
7.	2SLS-KR[a]	7.146	FDGMM	15.319	FDGMM	12.627
8.	Individual ML	7.706	Empirical iterative Bayes	17.919	Empirical iterative Bayes	14.189
9.	Empirical iterative Bayes	7.788	Iterative Bayes	21.493	Iterative Bayes	17.061
10	Empirical Bayes	8.045	Individual ML	21.845	Individual ML	17.717
11.	FDGMM	8.273	Empirical Bayes	24.096	Empirical Bayes	19.394
12.	FD2SLS[a]	9.608	FD2SLS[a]	40.215	FD2SLS[a]	23.897
13.	Individual OLS	9.628	GMM	40.943	Indivividual OLS	30.777
14.	Individual 2SLS	12.707	Individual OLS	41.110	GMM	33.840
15.	FD2SLS-KR[a]	14.109	Swamy	49.616	FD2SLS-KR[a]	40.204
16.	GMM	23.108	Average OLS	53.308	Swamy	47.826
17.	Swamy	44.197	Average 2SLS	53.828	Average 2SLS	50.424
18.	Average 2SLS	44.721	FD2SLS-KR[a]	69.431	Average OLS	52.906
19.	Average OLS	52.611	Individual 2SLS	100.281	Individual 2SLS	54.399

4. Summary and conclusions

This paper confirms again the value of panel data sets and the emphasis given to pooled estimators using two US panel data sets on residential electricity and natural-gas demand across 49 states over the period 1970–1990. Our results show that when the data is used to estimate heterogeneous models across states, individual estimates offer the worst out-of-sample forecasts. Despite the fact that shrinkage estimators outperform these individual estimates, they are outperformed by simple homogeneous panel data estimates in out-of-sample forecasts. Admittedly, this is another case study using US data, but it does add to the evidence that simplicity and parsimony in model estimation offer better forecasts.

Baltagi-bresson-Pirotte (2004)

Following Hsiao et al. (1999), we consider a simple dynamic version of the classical Tobin-q investment model:

$$\left(\frac{I}{K}\right)_{it} = a_i + b_i \left(\frac{I}{K}\right)_{it-1} + c_i q_{it} + u_{it}$$

where I_{it} denotes investment expenditures by firm i during period t , K_{it} is the replacement value of the capital stock and q_{it} is Tobin's q of the firm. Tobin's

q is the ratio of the market value of new investment goods to their replacement cost.

273 firms over a longer time period 1973–1992.

Table 2. Comparison of forecast performance of the q investment model

Ranking	1st year		5th year		5-year average	
	Estimator	RMSE[1]	Estimator	RMSE[1]	Estimator	RMSE[1]
1	Hierarchical Bayes	6.6781	OLS	10.0769	Hierarchical Bayes	8.5307
2	Individual ML	6.9151	2SLS-KR	10.0825	GLS	8.8064
3	Iterative Bayes	6.9651	2SLS	10.0915	Iterative empirical Bayes	8.8069
4	Iterative empirical Bayes	7.0024	Hierarchical Bayes	10.1428	Iterative Bayes	8.8464
5	GLS	7.0722	GLS	10.1968	OLS	8.8957
6	Empirical Bayes	7.0805	Iterative empirical Bayes	10.4385	2SLS-KR	8.9089
7	OLS	7.1541	Iterative Bayes	10.6349	2SLS	8.9239
8	2SLS-KR	7.1773	Within	10.9203	Individual ML	8.9909
9	2SLS	7.1970	Within-2SLS	10.9614	Empirical Bayes	9.2750
10	FD2SLS	7.4861	Individual ML	10.9756	Within	9.2786
11	FD2SLS- KR	7.5008	Empirical Bayes	11.4226	Within-2SLS	9.4586
12	Within	7.5030	FD2SLS- KR	11.9677	FD2SLS- KR	9.9345
13	FDGMM	7.6695	FD2SLS	12.0473	FD2SLS	9.9486
14	Individual OLS	7.7484	FDGMM	12.5747	FDGMM	10.2930
15	Within-2SLS	7.8644	Individual OLS	13.6907	Individual OLS	10.6765
16	Individual 2SLS	8.5933	Swamy	16.1467	Swamy	14.0715
17	Swamy	11.9773	Average OLS	19.8330	Individual 2SLS	14.1792
18	Pooled mean group	12.9823	Average 2SLS	21.8026	Average OLS	17.2825
19	Average OLS	14.9043	Individual 2SLS	21.8941	Pooled mean group	17.4408
20	Average 2SLS	15.5311	Pooled mean group	22.0320	Average 2SLS	18.6442

[1] RMSE $\times 10^{-2}$

consistent finding in all these studies including this one is that homogeneous panel data estimators perform well in forecast performance mostly due to their simplicity, their parsimonious representation and the stability of the parameter estimates. Average heterogeneous estimators perform badly due to parameter estimate instability caused by the estimation of several parameters with short time series. Shrinkage estimators did well for this application, especially iterative Bayes and iterative empirical Bayes. For this empirical example, the hierarchical Bayes estimator performs very well and gives the best RMSE forecast for the five year average.

Mazzanti and Musolesi (2013)

Environmental Kuznets Curve

$$C02_{it} = a_i + \beta_{1(i)} GDPPC_{it} + \beta_{2(i)} GDPPC_{it}^2 + ... + u_{it}$$

Heterogeneous estimators suggest a linear CO2-income relation

Heterogeneous estimators Bayesian or not provide a very robust picture (probably due to the large time series dimension, T=40)

Introducing unobserved common factors destroys the ECK relation. . .