Modelling heterogeneity and testing for units roots in panels with a fixed time-series dimension

Edith Madsen
Center for Applied Microeconometrics (CAM)
Institute of Economics
University of Copenhagen

January 2005

1 Introduction

Different types of panel data:

- Micro-panels where the time-series dimension is small relative to the cross-section dimension, (individuals, households, firms)
- Macro-panels where the time-series and cross-section dimensions are of similar magnitude, (regions, countries)

The properties of estimators are likely to depend on the type of panel data.

We consider inference in dynamic micro-panels and the most simple framework for this is a first-order autoregressive panel data model with individual-specific levels

$$y_{it} = \rho y_{it-1} + \alpha_i + \varepsilon_{it} \qquad \text{for } i = 1, ..., N \text{ and } t = 1, ..., T \qquad (1)$$

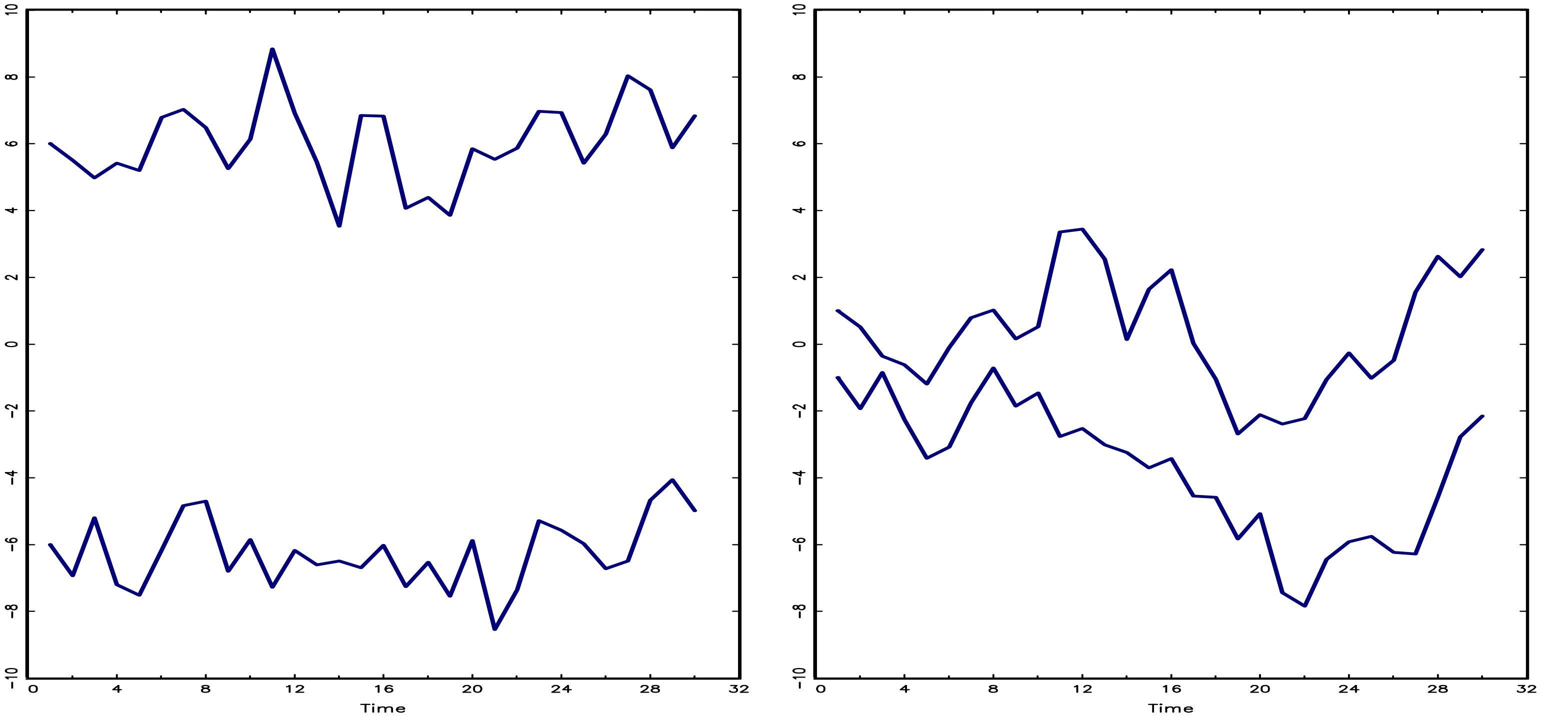
This model is widely used in applied econometrics. The reason for this is that the persistency measured by cross-sectional autocorrelation coefficients (the correlation between a variable and its lagged value) is often high. An example is income at the individual level.

The persistency can be caused by:

- Units having very different levels
- Strong correlation in the time-series processes (common for all units)

The first-order autogressive model allows for both types of persistency and inference on the parameters in the model provides information about the relative importance of the two sources of persistency.

Figure 1: The two types of persistency



2 Heterogeneity

Consider the following more general first-order autoregressive panel data model

$$y_{it} = \rho_i y_{it-1} + \alpha_i + \varepsilon_{it} \qquad \text{for } i = 1, ..., N \text{ and } t = 1, ..., T \qquad (2)$$

where ε_{it} is iid across i, t with $E(\varepsilon_{it}) = 0$ and $E\left(\varepsilon_{it}^2\right) = \sigma_{i\varepsilon}^2$.

Two extreme versions of the model:

a. The AR-parameter is the same for all units

b. The AR-parameter varies completely at random across units

In version (a) of the model it is possible to make inference on the AR-parameter. Even though this model is widely used it can be argued that it does not allow for enough heterogeneity between cross-section units.

Version (b) of the model allows for so much heterogeneity that there is very little panel data structure in it. Inference on the individual-specific AR-parameters or the mean AR-parameter is problematic in micro-panels, see Robertson & Symons (1992) and Pesaran & Smith (1995).

A compromise between the two extreme versions of the model:
Let the AR-parameter vary according to some specific parametric distribution (impose panel data structure) and then estimate the parameters describing this distribution. This approach used by Alvarez, Browning & Ejrnæs (2002) and Hsiao, Pesaran & Tahmiscioglu (1999).

The approach suggested in this paper:
The AR-parameter and the variance of the error term vary across units according to a discrete distribution with only two possible values. This is formulated in terms of a traditional mixture model. This type of model is well-known from the analysis of non-linear time-series (switching regression model, unobserved regime model).

3 The mixture model

For every cross-section unit $i = 1, ..., N$:

$$
\begin{aligned}
y_{it} &= \rho_1 y_{it-1} + \alpha_{1i} + \varepsilon_{1,it} \quad \text{for } t = 1, ..., T \qquad \text{with prob. } p_i \\
y_{it} &= \rho_2 y_{it-1} + \alpha_{2i} + \varepsilon_{2,it} \quad \text{for } t = 1, ..., T \qquad \text{with prob. } 1 - p_i
\end{aligned}
\tag{3}
$$

where $\varepsilon_{1,it}$ and $\varepsilon_{2,it}$ are iid Normal across i, t with means zero and variances σ_1^2 and σ_2^2 and they are both independent of y_{i0}, α_{1i} and α_{2i}.

The mixing weights are on logistic form

$$
p_i = \frac{\exp\left(\gamma' D_i\right)}{1 + \exp\left(\gamma' D_i\right)} \tag{4}
$$

where D_i is a vector of observed individual-specific characteristics.

Important property:
The mixing is done only in the cross-section dimension not in the time-series dimension. This means that for a given cross-section unit, the parameters are constant over time.

The individual-specific intercepts:

α_{1i} and α_{2i} come from the same underlying parametric distribution. Here we assume a distribution which allows for correlation between the individual-specific effects and the initial value, (the Chamberlain-approach):

$$
\begin{aligned}
(\alpha_{1i} | \, y_{i0}, D_i) &\sim N\left(\alpha_1 y_{i0}, \sigma^2_{1\alpha}\right) \qquad \text{where } \sigma^2_{1\alpha} \geq 0 \\
(\alpha_{2i} | \, y_{i0}, D_i) &\sim N\left(\alpha_2 y_{i0}, \sigma^2_{2\alpha}\right) \qquad \text{where } \sigma^2_{2\alpha} \geq 0
\end{aligned}
$$

Under this assumption, the component density in the mixture model is the same as in a standard random effects model where the initial value is included as a regressor in all T equations.

4 Maximum-likelihood estimation

Usually in this type of mixture model the likelihood function tends to infinity on the boundary of the parameter space when the variances tend to zero. The singularity problem can be avoided if there are enough observations from each component of the mixture model.

Theoretical result:
When the two components are different and there are enough observations over time (at least 4 observations over time), then the usual result about the existence and asymptotic properties of the ML estimator holds, (as $N \to \infty$).

In practice the ML estimator is obtained by using the EM-algorithm as this simplifies the optimization problem.

E-step:
Compute the probability that unit i conditional on the observed values $y_{i0}, ..., y_{iT}$ belongs to the first component p_i^*

M-step:

- The parameters in the first component density $\left(\rho_1, \alpha_1, \sigma^2_{1\alpha}, \sigma^2_{1\varepsilon}\right)$ are obtained as weighted random effects ML estimators with weights p_i^*.
- The parameters in the second component density $\left(\rho_2, \alpha_2, \sigma^2_{2\alpha}, \sigma^2_{2\varepsilon}\right)$ are obtained as weighted random effects ML estimators with weights $(1 - p_i^*)$.
- The parameters in the mixing weights γ are obtained by a logistic regression of p_i^* on the variables describing the mixing weights.

5 Testing for unit roots

The model is the following:

$$\begin{aligned} y_{it} &= \rho_1 y_{it-1} + \alpha_1 y_{i0} + v_{1,it} \quad \text{with prob. } p_i \\ y_{it} &= \rho_2 y_{it-1} + \alpha_2 y_{i0} + v_{2,it} \quad \text{with prob. } 1 - p_i \end{aligned}$$

where

$$E\left(v_{k,it} v_{k,is}\right) = \begin{cases} \sigma^2_{k\varepsilon} + \sigma^2_{k\alpha} & \text{for } t = s \\ \sigma^2_{k\alpha} & \text{for } t \neq s \end{cases} \qquad k = 1, 2$$

Test the hypothesis that a group of the cross-section units has an AR-parameter of unity

$$\begin{aligned} H_{01} &: \rho_2 = 1 \quad \alpha_2 = 0 \quad \sigma^2_{2\alpha} = 0 \\ H_{A1} &: \rho_2 \neq 1 \quad \text{or} \quad \alpha_2 \neq 0 \quad \text{or} \quad \sigma^2_{2\alpha} \neq 0 \end{aligned}$$

Time-series interpretation:
The stationary AR without drift against a non-stationary AR without drift

Cross-section interpretation:
The high cross-sectional correlation between y_{it} and y_{it-1} is coming from the autoregressive mechanism

The hypothesis is tested by using a LR test statistic which is asymptotically $\chi^2(3)$.

6 Existing unit root tests

Tests based on pooled estimators of the AR-parameter:

Breitung & Meyer (1994), Harris & Tzavalis (1999), Breitung (1997)

Model: $y_{it} = \rho y_{it-1} + (1-\rho)\,\alpha_i + \varepsilon_{it}$

Testing $H_0 : \rho = 1$ against $H_A : \rho < 1$

Tests based on individual-specific statistics:

Im, Pesaran & Shin (2002), Maddala & Wu (1999)

Model: $y_{it} = \rho_i y_{it-1} + (1-\rho_i)\,\alpha_i + \varepsilon_{it}$

Testing $H_0 : \rho_i = 1$ against $H_A : \rho_i \leq 1$ with $\rho_i < 1$ for a share of units

These tests provide an answer to the question:

Is the AR-parameter equal to unity for all cross-section units?

If the answer is NO there are at least two possible explanations:

- No units with an AR-parameter of unity
- Only a share of units with an AR-parameter of unity

The tests can not distinguish between these two explanations.

7 Application: US annual income

Data: sample drawn from the PSID covering the period 1969-93 according to the following criteria:

- Males aged 25-55, head of the household
- Positive earnings
- Constant level of education
- Log real earnings between 8.5 and 12.5

Total of 562 individuals observed for at least 15 adjoining years.

Many studies of income/earnings dynamics based on the PSID.
E.g. McCurdy (1982) and Abowd & Card (1989) where an AR-parameter of unity is assumed as a starting point.

$$y_{it} = \log(\text{annual income}_{it})$$

$$x_{it} = \text{constant},\ \text{age}_{it},\ \text{age}_{it}^2,\ \text{education-dummies}\,'$$

The mixing weights do not depend on education level or year of birth, therefore they are the same for all individuals, i.e. $p_i = p$ for $i = 1, ..., N$.

Some estimates from the mixture model:

Parameter	Group 1	Group 2
ρ	0.5043 (0.0145)	0.7060 (0.0150)
α	0.1853 (0.0240)	0.2016 (0.0174)
Constant	2.3185 (0.9674)	0.5709 (0.2582)
σ_ε^2	0.1074 (0.0025)	0.0170 (0.0005)
$\sigma_\alpha^2/\sigma_\varepsilon^2$	0.1885 (0.0263)	0.2593 (0.0406)
Weight	0.4885 (0.0230)	0.5115 (0.0230)
Log-like:	684.69	

Standard errors are in brackets

Mixture model under unit root hypothesis:

Parameter	Group 1	Group 2
ρ	0.6616 (0.0169)	1
α	0.2075 (0.0166)	0
Constant	0.6189 (0.2671)	-0.9122 (0.4306)
σ_ε^2	0.0212 (0.0011)	0.1608 (0.0060)
$\sigma_\alpha^2/\sigma_\varepsilon^2$	0.3035 (0.0430)	0
Weight	0.5795 (0.0292)	0.4205 (0.0292)
Log-like:	198.16	

Conclusion:

- The processes describing the two groups of individuals are very different, (high levels are associated with low variation and vice versa)
- There is no evidence of unit roots

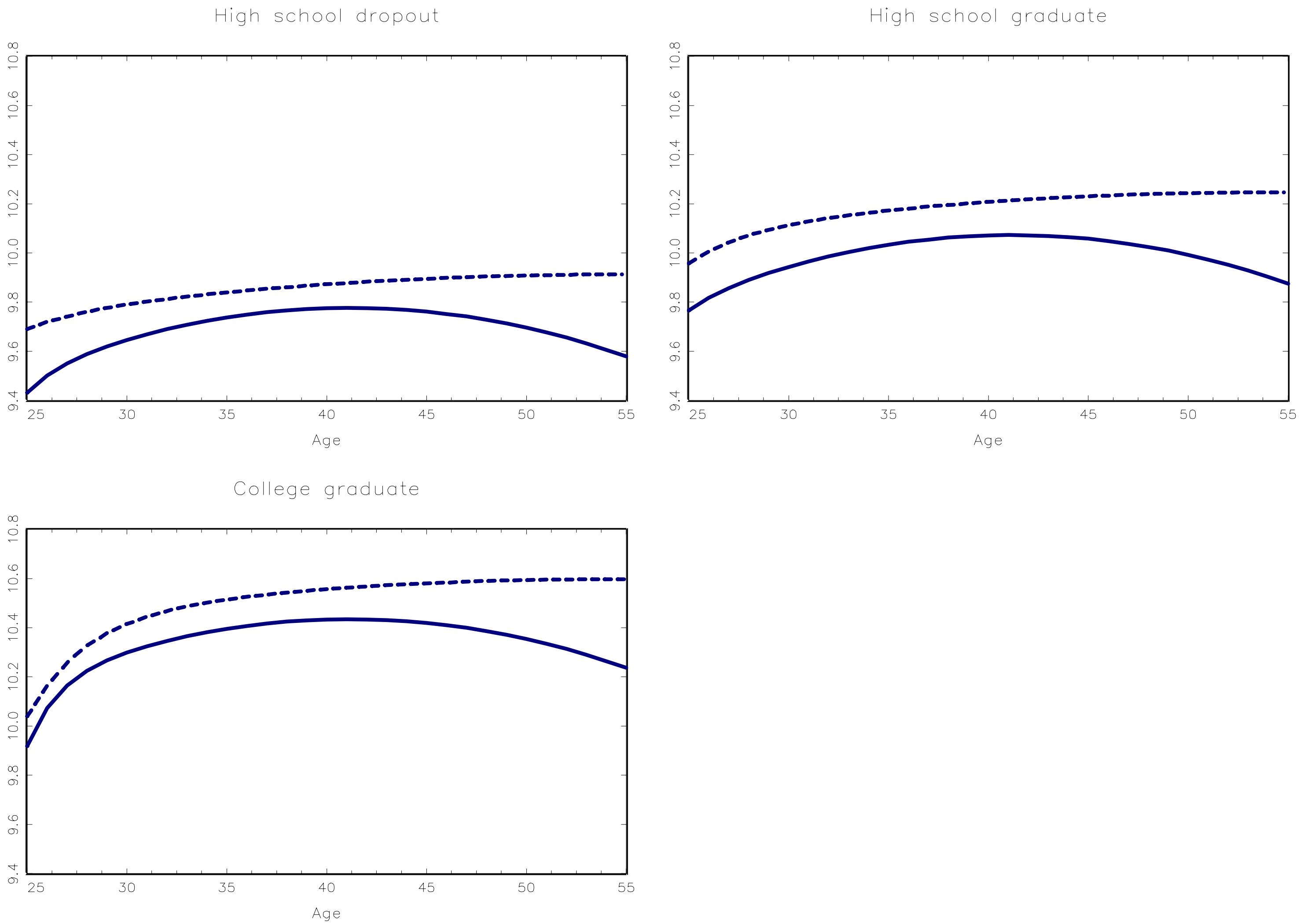


Figure 2: Log earnings for group 1 (solid line) and group 2 (dashed line)

8 Summary

- The paper suggests a method which can be used to model cross-section heterogeneity in micro-panels. The method is very flexible since the mixture model easily can be extended to contain more than two components. (This raises the issue of how to determine the number of components).
- The model makes it possible to distinguish between unit root hypotheses which can not be distinguished by any of the existing test procedures. More specifically, it is possible to test the hypothesis that a group of units have unit root processes.