ading for the

α_i and γ_i.

Large Heterogeneous Panel Data Models

28.1 Introduction

Panel data models introduced in the previous two chapters, 26 and 27, deal with panels where the time dimension (T) is fixed, and assumes that conditional on a number of observable characteristics, any remaining heterogeneity over the cross-sectional units can be modelled through an additive intercept (assuming either fixed or random), and possibly heteroskedastic errors. This chapter extends the analysis of panels to linear panel data models with slope heterogeneity. It discusses how neglecting such heterogeneities affects the consistency of the estimates and inferences based upon them, and introduces models that explicitly allow for slope heterogeneity both in the case of static and dynamic panel data models. To deal with slope heterogeneity, particularly in the case of dynamic models, it is often necessary to assume that the number of time series observations, T, is relatively large, so that individual equations can be estimated for each unit separately. Models, estimation and inference procedures developed in this and subsequent chapters are more suited to large N and T panels. Such panel data sets are becoming increasingly available and cover countries, regions, industries, and markets over relatively long time periods.

Despite the slope heterogeneity, the cross-sectional units could nevertheless share common features of interest. For example, it is possible for different countries or geographical regions to have different dynamics of adjustments towards equilibrium, due to their historical and cultural differences, but they could all converge to the same economic equilibrium in the very long run, due to forces of arbitrage and interconnections through international trade and cultural exchanges. Other examples include cases where slope coefficients can be viewed as random draws from a distribution with a number of parameters that are bounded in N. Large number of panel data sets fit within this setup, where the cross-sectional units might be industries, regions, or countries, and we wish to identify common patterns of responses across otherwise heterogeneous units. The parameters of interest may be intercepts, short-run coefficients, long-run coefficients or error variances.

This chapter deals with panels with stationary variables. The econometric analyses of panels with unit roots and cointegration is covered in Chapter 31.

28.2 Heterogeneous panels with strictly exogenous regressors

Suppose that the variable y_{it} for the i^{th} unit at time t is specified as a linear function of k strictly exogenous variables, x_{kit}, $k = 1, 2, \ldots, k$, in the form

$$\begin{aligned} y_{it} &= \sum_{k=1}^{k} \beta_{kit} x_{kit} + u_{it}, \\ &= \boldsymbol{\beta}'_{it}\mathbf{x}_{it} + u_{it},\ i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T, \end{aligned} \tag{28.1}$$

where u_{it} denotes the random error term, $\mathbf{x}_{it}$ is a $k \times 1$ vector of exogenous variables and $\boldsymbol{\beta}_{it}$ is the $k \times 1$ vector of coefficients. The above specification is very general and allows the coefficients to vary both across time and over individual units. As it is specified it is too general. It simply states that each individual unit has its own coefficients that are specific to each time period. However, as pointed out by Balestra (1996), this general formulation is, at most, descriptive. It lacks any explanatory power and it is not useful for prediction. Furthermore, it is not estimable, as the number of parameters to be estimated exceeds the number of observations. For a model to become interesting and to acquire explanatory and predictive power, it is essential that some structure is imposed on its parameters.

One way to reduce the number of parameters in (28.1) is to adopt a random coefficient approach, which assumes that the coefficients $\boldsymbol{\beta}_{it}$ are draws from probability distributions with a fixed number of parameters that do not vary with N and/or T. Depending on the type of assumption about the parameter variation, we can further classify the models into one of two categories: stationary and non-stationary random-coefficient models.

The stationary random-coefficient models view the coefficients as having constant means and variance-covariances. Namely, the $k \times 1$ vector $\boldsymbol{\beta}_{it}$ is specified as

$$\boldsymbol{\beta}_{it} = \boldsymbol{\beta} + \boldsymbol{\eta}_{it},\ \ i = 1, 2, \ldots, N,\ \ t = 1, 2, \ldots, T, \tag{28.2}$$

where $\boldsymbol{\beta}$ is a $k \times 1$ vector of constants, and $\boldsymbol{\eta}_{it}$ is a $k \times 1$ vector of stationary random variables with zero means and constant variance-covariances. One widely used random coefficient specification is the Swamy (1970) model, which assumes that the randomness is time-invariant

$$\boldsymbol{\beta}_{it} = \boldsymbol{\beta} + \boldsymbol{\eta}_{i},\ \ i = 1, 2, \ldots, N,\ \ t = 1, 2, \ldots, T, \tag{28.3}$$

and

$$E(\boldsymbol{\eta}_i) = \mathbf{0},\ E(\boldsymbol{\eta}_i\mathbf{x}'_{it}) = \mathbf{0}, \tag{28.4}$$

$$E(\boldsymbol{\eta}_i\boldsymbol{\eta}_j) = \begin{cases} \boldsymbol{\Omega}_{\eta}, & \text{if} \quad i = j, \\ \mathbf{0}, & \text{if} \quad i \neq j. \end{cases} \tag{28.5}$$

Estimation and inference in the above specification are discussed in Section 28.4.

Hsiao (1974, 1975) consider the following model

$$\begin{aligned}\boldsymbol{\beta}_{it} &= \boldsymbol{\beta} + \boldsymbol{\xi}_{it} \\ &= \boldsymbol{\beta} + \boldsymbol{\eta}_i + \boldsymbol{\lambda}_t, \quad i = 1, 2, \ldots, N,\ t = 1, 2, \ldots, T,\end{aligned} \tag{28.6}$$

and assume

$$\begin{aligned} E(\boldsymbol{\eta}_i) = E(\boldsymbol{\lambda}_t) &= \mathbf{0}, E\left(\boldsymbol{\eta}_i \boldsymbol{\lambda}_t'\right) = \mathbf{0}, \\ E\left(\boldsymbol{\eta}_i \mathbf{x}_{it}'\right) &= \mathbf{0}, E\left(\boldsymbol{\lambda}_t \mathbf{x}_{it}'\right) = \mathbf{0}, \\ E\left(\boldsymbol{\eta}_i \boldsymbol{\eta}_j'\right) &= \begin{cases} \boldsymbol{\Omega}_{\eta}, & \text{if } \quad i = j, \\ \mathbf{0}, & \text{if } \quad i \neq j, \end{cases} \\ E(\boldsymbol{\lambda}_i \boldsymbol{\lambda}_j') &= \begin{cases} \boldsymbol{\Omega}_{\Lambda}, & \text{if } \quad i = j, \\ \mathbf{0}, & \text{if } \quad i \neq j, \end{cases} \end{aligned} \tag{28.7}$$

Alternatively, a time varying parameter model may be treated as realizations of a stationary stochastic process, thus $\boldsymbol{\beta}_{it}$ can be written in the form,

$$\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_t = \mathbf{H}\boldsymbol{\beta}_{t-1} + \boldsymbol{\delta}_t, \tag{28.8}$$

where all eigenvalues of $\mathbf{H}$ lie inside the unit circle, and $\boldsymbol{\delta}_t$ is a stationary random variable with mean $\boldsymbol{\mu}$. Hence, letting $\mathbf{H} = \mathbf{0}$ and $\boldsymbol{\delta}_t$ be IID we obtain the model proposed by Hildreth and Houck (1968), while for the Pagan (1980) model, $\mathbf{H} = \mathbf{0}$ and

$$\boldsymbol{\delta}_t - \boldsymbol{\mu} = \boldsymbol{\delta}_t - \overline{\boldsymbol{\beta}} = \boldsymbol{A}(L)\boldsymbol{\epsilon}_t, \tag{28.9}$$

where $\overline{\boldsymbol{\beta}}$ is the mean of $\boldsymbol{\beta}_t$ and $\boldsymbol{A}(L)$ is a matrix polynomial in the lag operator L (with $L\boldsymbol{\epsilon}_t = \boldsymbol{\epsilon}_{t-1}$), and $\boldsymbol{\epsilon}_t$ is independent normal. The Rosenberg (1972), Rosenberg (1973) return-to-normality model assumes that the absolute value of the characteristic roots of $\mathbf{H}$ be less than 1, with $\boldsymbol{\eta}_t$ independently normally distributed with mean $\boldsymbol{\mu} = (\mathbf{I}_k - \mathbf{H})\overline{\boldsymbol{\beta}}$.

The non-stationary random coefficients models do not regard the coefficient vector as having constant mean or variances. Changes in coefficients from one observation to the next can be the result of the realization of a nonstationary stochastic process or can be a function of exogenous variables. When the coefficients are realizations of a nonstationary stochastic process, we may again use (28.8) to represent such a process. For instance, the Cooley and Prescott (1976) model can be obtained by letting $\mathbf{H} = \mathbf{I}_k$ and $\boldsymbol{\mu} = \mathbf{0}$. When the coefficients $\boldsymbol{\beta}_{it}$ are functions of individual characteristics or time variables (e.g. see Amemiya (1978), Boskin and Lau (1990)), we can let

$$\boldsymbol{\beta}_{it} = \boldsymbol{\Gamma}\mathbf{q}_{it} + \boldsymbol{\eta}_{it}. \tag{28.10}$$

While the detailed formulation and estimation of the random coefficients model depends on the specific assumptions about the parameter variation, many types of random coefficients models can be conveniently represented using a mixed fixed and random coefficients framework of the form (see, for example, Hsiao, Appelbe, and Dineen (1992))

$$y_{it} = \mathbf{z}'_{it}\overline{\boldsymbol{\gamma}} + \mathbf{w}'_{it}\boldsymbol{\alpha}_{it} + u_{it},\ i = 1, 2, \ldots, N,\ t = 1, 2, \ldots, T, \tag{28.11}$$

where $\mathbf{z}_{it}$ and $\mathbf{w}_{it}$ are vectors of exogenous variables with dimensions ℓ and p respectively, $\overline{\boldsymbol{\gamma}}$ is an $\ell \times 1$ vector of constants, $\boldsymbol{\alpha}_{it}$ is a $p \times 1$ vector of random variables, and u_{it} is the error term. For instance, the Swamy type model, (28.3), can be obtained from (28.11) by letting $\mathbf{z}_{it} = \mathbf{w}_{it} = \mathbf{x}_{it}$, $\overline{\boldsymbol{\gamma}} = \overline{\boldsymbol{\beta}}$, and $\boldsymbol{\alpha}_{it} = \boldsymbol{\eta}_i$; the Hsiao type model (28.6) and (28.7) is obtained by letting $\mathbf{z}_{it} = \mathbf{w}_{it} = \mathbf{x}_{it}$, $\overline{\boldsymbol{\gamma}} = \overline{\boldsymbol{\beta}}$, and $\boldsymbol{\alpha}_{it} = \boldsymbol{\eta}_i + \boldsymbol{\lambda}_t$; the stochastic time varying parameter model (28.8) is obtained by letting $\mathbf{z}_{it} = \mathbf{x}_{it}$, $\mathbf{w}'_{it} = \mathbf{x}'_{it}\left(\mathbf{H}, \mathbf{I}_k\right)$, $\overline{\boldsymbol{\gamma}} = \boldsymbol{\mu}$, and $\boldsymbol{\alpha}'_{it} = \boldsymbol{\lambda}'_t = [\boldsymbol{\beta}'_{t-1}, (\boldsymbol{\delta}_t - \boldsymbol{\mu})']$; and the model where $\boldsymbol{\beta}_{it}$ is a function of other variables is obtained by letting $\mathbf{z}'_{it} = \mathbf{x}'_{it} \otimes \mathbf{q}'_{it}$, $\overline{\boldsymbol{\gamma}}' = vec(\boldsymbol{\Gamma})$, $\mathbf{w}_{it} = \mathbf{x}_{it}$, $\boldsymbol{\alpha}_{it} = \boldsymbol{\eta}_{it}$, etc.

In this chapter we focus on models with time-invariant slope coefficients that vary randomly or freely over the cross-sectional units. We begin by considering the implications of neglecting such heterogeneity on the consistency and efficiency of the homogenous slope type estimators such as fixed and random effects models.

28.3 Properties of pooled estimators in heterogeneous panels

To understand the consequences of erroneously ignoring slope heterogeneity, consider the following panel data model, where, for simplicity of exposition, we set $k = 1$

$$y_{it} = \mu_i + \beta_i x_{it} + u_{it}, \tag{28.12}$$

$u_{it} \sim IID\left(0, \sigma_u^2\right)$, and μ_i are unknown fixed parameters. The coefficients, β_i, are allowed to vary freely across units but are otherwise assumed to be fixed (over time). It proves useful to decompose β_i into a common component, β, and a remainder term, η_i, that varies across units:

$$\beta_i = \beta + \eta_i. \tag{28.13}$$

The nature of the slope heterogeneity can now be characterized in terms of the properties of η_i, in particular where there is systematic dependence between η_i and the regressors x_{it} and an additional regressor z_{it}.

Consider an investigator that ignores the heterogeneity of the slope coefficients in (28.12), and instead estimates the model

$$y_{it} = \alpha_i + \delta_x x_{it} + \delta_z z_{it} + v_{it}, \tag{28.14}$$

where z_{it} is an additional regressor spuriously thought to be important by the researcher.

To simplify the derivations we make the following assumptions:

Assumption H.1: u_{it} is serially uncorrelated and distributed independently of u_{jt} for all $i \neq j$, with variance $0 < \sigma_i^2 < K$.

Assumption H.2: $\mathbf{w}_{it} = (x_{it}, z_{it})'$ is distributed independently of $u_{it'}$, for all i, t and t'.

Assumption H.3: $\mathbf{w}_{it}$ follows a covariance stationary process with the covariance matrix, $\boldsymbol{\Omega}_i$,

$$\boldsymbol{\Omega}_i = \begin{pmatrix} \omega_{ixx} & \omega_{ixz} \\ \omega_{izx} & \omega_{izz} \end{pmatrix}, \tag{28.15}$$

such that

$$E(\boldsymbol{\Omega}_i) = \lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Omega}_i\right), \tag{28.16}$$

is a positive definite matrix.

Assumption H.4: For each t, $\mathbf{w}_{it}$ is distributed independently across i.

Note that not all the above assumptions are necessary when both N and T are sufficiently large. For example, assumption H.4 is not needed when T is sufficiently large. Assumptions H.1 and H.3 can be relaxed when T is small. It is also worth noting that assumption H.3 does not require the correlation matrix of the regressors for all i to be nonsingular, only that the 'pooled' covariance matrix, $E(\boldsymbol{\Omega}_i)$, defined by (28.16), should be nonsingular.

In matrix notation, (28.12) and (28.14) can be written as

$$\mathbf{y}_i = \mu_i \boldsymbol{\tau}_T + \beta_i \mathbf{x}_i + \mathbf{u}_i, \tag{28.17}$$

and

$$\mathbf{y}_i = \alpha_i \boldsymbol{\tau}_T + \mathbf{W}_i \boldsymbol{\delta} + \mathbf{v}_i, \tag{28.18}$$

respectively, where

$$\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})', \boldsymbol{\tau}_T = (1, 1, \ldots, 1)', \mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iT})',$$
$$\mathbf{u}_i = (u_{i1}, u_{i2}, \ldots, u_{iT})', \boldsymbol{\delta} = (\delta_x, \delta_z)',$$

and

$$\mathbf{W}_i = \begin{pmatrix} x_{i1} & z_{i1} \\ x_{i2} & z_{i2} \\ \vdots & \vdots \\ x_{iT} & z_{iT} \end{pmatrix}, \ \mathbf{v}_i = \begin{pmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{iT} \end{pmatrix}.$$

The fixed-effects (FE) estimators of the slope coefficients in (28.18) can be written as

$$\hat{\boldsymbol{\delta}}_{FE} = \begin{pmatrix} \hat{\delta}_{x,FE} \\ \hat{\delta}_{z,FE} \end{pmatrix} = \left(\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{M}_T\mathbf{W}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{M}_T\mathbf{y}_i\right), \tag{28.19}$$

where $\mathbf{M}_T = \mathbf{I}_T - \boldsymbol{\tau}_T(\boldsymbol{\tau}_T'\boldsymbol{\tau}_T)^{-1}\boldsymbol{\tau}_T'$.[1] Under (28.17) we have

$$\hat{\boldsymbol{\delta}}_{FE} = \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{M}_T\mathbf{W}_i\right)^{-1}\left[\frac{1}{NT}\sum_{i=1}^{N}\left(\mathbf{W}_i'\mathbf{M}_T\mathbf{x}_i\right)\beta_i + \frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{M}_T\mathbf{u}_i\right]. \tag{28.20}$$

[1] The fixed-effects estimator in (28.19) assumes a balanced panel. But the results readily extend to unbalanced panels.

It is now easily seen that under Assumptions H.1–H.4 and for N and/or T sufficiently large

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{W}_i'\mathbf{M}_T\mathbf{u}_i}{T}\right)\xrightarrow{p}0, \tag{28.21}$$

where $\xrightarrow{p}$ denotes convergence in probability. To see this, note that since u_{it} are cross-sectionally independent and $\mathbf{w}_{it}$ are strictly exogenous, then we have

$$Var\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{W}_i'\mathbf{M}_T\mathbf{u}_i}{T}\right)\right]=\frac{1}{TN}\left[\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 E\left(\frac{\mathbf{W}_i'\mathbf{M}_T\mathbf{W}_i}{T}\right)\right].$$

Also, under Assumptions H.1 and H.3, σ_i^2 and $E\left(T^{-1}\mathbf{W}_i'\mathbf{M}_T\mathbf{W}_i\right)$ are bounded and as a result

$$Var\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{W}_i'\mathbf{M}_T\mathbf{u}_i}{T}\right)\right]\to\mathbf{0},$$

if N and/or $T\to\infty$. Also, under strict exogeneity of $\mathbf{w}_{it}$, $E\left(T^{-1}\mathbf{W}_i'\mathbf{M}_T\mathbf{u}_i\right)=\mathbf{0}$, for all i, and the desired result in (28.21) follows.

Using (28.21) in (28.20) we now have

$$Plim_{N,T\to\infty}(\hat{\delta}_{FE})=\left[Plim\left(\sum_{i=1}^{N}\frac{\mathbf{W}_i'\mathbf{M}_T\mathbf{W}_i}{NT}\right)\right]^{-1}Plim\left[\sum_{i=1}^{N}\left(\frac{\mathbf{W}_i'\mathbf{M}_T\mathbf{x}_i}{NT}\right)\beta_i\right]. \tag{28.22}$$

In the case where the slopes are homogenous, namely $\beta_i=\beta$, we have

$$Plim(\hat{\delta}_{FE})=\begin{pmatrix}\beta\\0\end{pmatrix}. \tag{28.23}$$

Consider now the case where the slopes are heterogenous. Using the above results, it is now easily seen that the consistency result in (28.23) will follow if and only if

$$\sum_{i=1}^{N}\left(\frac{\mathbf{W}_i'\mathbf{M}_T\mathbf{x}_i}{NT}\right)\eta_i=\begin{pmatrix}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{M}_T\mathbf{x}_i\eta_i\\\frac{1}{NT}\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{M}_T\mathbf{x}_i\eta_i\end{pmatrix}\xrightarrow{p}\mathbf{0}. \tag{28.24}$$

This condition holds under the random coefficient specification where it is assumed that η_i's are distributed independently of $\mathbf{w}_{it}$ for all i and t. (See below and Swamy (1970)). Under Assumption H.3 and as $T\to\infty$ we have

$$\frac{1}{NT}\sum_{i=1}^{N}\left(\mathbf{x}_i'\mathbf{M}_T\mathbf{x}_i\right)\eta_i-\frac{1}{N}\sum_{i=1}^{N}\omega_{ixx}\eta_i\xrightarrow{p}0,$$

and

$$\frac{1}{NT}\sum_{i=1}^{N}\left(\mathbf{z}_i'\mathbf{M}_T\mathbf{x}_i\right)\eta_i - \frac{1}{N}\sum_{i=1}^{N}\omega_{izx}\eta_i \xrightarrow{p} 0.$$

Therefore, for the fixed-effects estimator, (28.19), to be consistent we must have

$$\frac{1}{N}\sum_{i=1}^{N}\omega_{ixx}\eta_i \xrightarrow{p} 0, \text{ and } \frac{1}{N}\sum_{i=1}^{N}\omega_{izx}\eta_i \xrightarrow{p} 0, \tag{28.25}$$

as $N \to \infty$. Namely, any systematic dependence between β_i and the second-order moments of the steady state distribution of the regressors $\mathbf{w}_{it}$ must also be ruled out. When the conditions in (28.25) are not satisfied, the inconsistencies of the fixed-effects estimators (for T and N sufficiently large) are given by[2]

$$Plim(\hat{\delta}_{x,FE} - \beta) = \frac{Cov(\omega_{ixx},\eta_i)E(\omega_{izz}) - E(\omega_{ixz})Cov(\omega_{ixz},\eta_i)}{E(\omega_{ixx})E(\omega_{izz}) - [E(\omega_{ixz})]^2}, \tag{28.26}$$

$$Plim(\hat{\delta}_{z,FE}) = \frac{Cov(\omega_{ixz},\eta_i)E(\omega_{ixx}) - E(\omega_{ixz})Cov(\omega_{ixx},\eta_i)}{E(\omega_{ixx})E(\omega_{izz}) - [E(\omega_{ixz})]^2}, \tag{28.27}$$

where

$$Cov(\omega_{ixx},\eta_i) = Plim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\omega_{ixx}\eta_i\right),\ Cov(\omega_{ixz},\eta_i) = Plim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\omega_{ixz}\eta_i\right),$$

$$E(\omega_{ixx}) = \lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\omega_{ixx}\right),\ E(\omega_{izz}) = \lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\omega_{izz}\right), \tag{28.28}$$

$$E(\omega_{ixz}) = \lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\omega_{ixz}\right).$$

The above results have a number of interesting implications:

1. The FE estimators, $\hat{\delta}_{x,FE}$ and $\hat{\delta}_{z,FE}$, are both consistent if

$$Cov(\omega_{ixz},\eta_i) = Cov(\omega_{ixx},\eta_i) = 0. \tag{28.29}$$

Clearly, these conditions are met under slope homogeneity. In the present application where the regressors are assumed to be strictly exogenous, the fixed-effects estimators converge to their true values under the random coefficient model (RCM) where the slope coefficients and the regressors are assumed to be independently distributed. Notice, however, that since the β_i's are assumed to be fixed over time, then any systematic depen-

[2] Notice that under slope heterogeneity the fixed-effects estimators are inconsistent when N is finite and only $T \to \infty$.

dence of η_i on $\mathbf{w}_{it}$ over time is already ruled out under model (28.12). The random coefficients assumption imposes further restrictions on the joint distribution of η_i and the cross-sectional distribution of $\mathbf{w}_{it}$.

2. The *FE* estimator of δ_z is robust to slope heterogeneity if the incorrectly included regressors, z_{it}, are on average orthogonal to x_{it}, namely when $E(\omega_{ixz}) = 0$, and if $Cov(\omega_{ixz}, \eta_i) = 0$. However, in the presence of slope heterogeneity, the FE estimator of δ_x continues to be inconsistent even if z_{it} and x_{it} are on average orthogonal. The direction of the asymptotic bias of $\hat{\delta}_{x,FE}$ depends on the sign of $Cov(\omega_{ixx}, \eta_i)$. The bias of $\hat{\delta}_{x,FE}$ is positive when $Cov(\omega_{ixx}, \eta_i) > 0$ and *vice versa*.[3]
3. In general, where $E(\omega_{ixz}) \neq 0$ and $Cov(\omega_{ixz}, \eta_i) \neq 0$ and/or $Cov(\omega_{ixx}, \eta_i) \neq 0$, the fixed-effects estimators, $\hat{\delta}_{x,FE}$ and $\hat{\delta}_{z,FE}$, are both inconsistent.

In short, if the slope coefficients are fixed but vary systematically across the groups, the application of the general-to-specific methodology to standard panel data models can lead to misleading results (spurious inference). An important example is provided by the case when attempts are made to check for the presence of nonlinearities by testing the significance of quadratic terms in static panel data models using fixed-effects estimators. In the context of our simple specification, this would involve setting $z_{it} = x_{it}^2$, and a test of the significance of z_{it} in (28.14) will yield sensible results only if the conditions defined by (28.29) are met. In general, it is possible to falsely reject the linearity hypothesis when there are systematic relations between the slope coefficients and the cross-sectional distribution of the regressors. Therefore, results from nonlinearity tests in panel data models should be interpreted with care. The linearity hypothesis may be rejected not because of the existence of a genuine nonlinear relationship between y_{it} and x_{it}, but due to slope heterogeneity.

Finally, it is worth noting that since the β_i's are fixed for each i, the nonlinear specification

$$y_{it} = \alpha_i + \delta_x x_{it} + \delta_z x_{it}^2 + v_{it}, \tag{28.30}$$

cannot be reconciled with (28.29), unless it is assumed that β_i varies proportionately with x_{it}. Clearly, it is possible to allow the slopes, β_i, to vary systematically with some aspect of the cross-sectional distribution of x_{it} without requiring β_i to be proportional to x_{it}, and hence time-varying. For example, it could be that

$$\beta_i = \gamma_0 + \gamma_1 \bar{x}_i, \tag{28.31}$$

where $\bar{x}_i = T^{-1} \sum_{t=1}^{T} x_{it}$. This specification retains the linearity of (28.29) for each i, but can still yield a statistically significant effect for x_{it}^2 in (28.30) if slope heterogeneity is ignored and fixed-effects estimates of (28.30) are used for inference. This feature of fixed-effects regressions under heterogeneous slopes is illustrated in Figure 28.1. The figure shows scatter points and associated regression lines for three countries with slopes that differ systematically with $\bar{x}_i$. It is clear that the pooled regression based on the scatter points from all three countries will exhibit strong nonlinearities, although the country-specific regressions are linear.

[3] Notice that $E(\omega_{ixx})E(\omega_{izz}) - (E(\omega_{ixz}))^2 > 0$, unless x_{it} and z_{it} are perfectly collinear for all i, which we rule out.

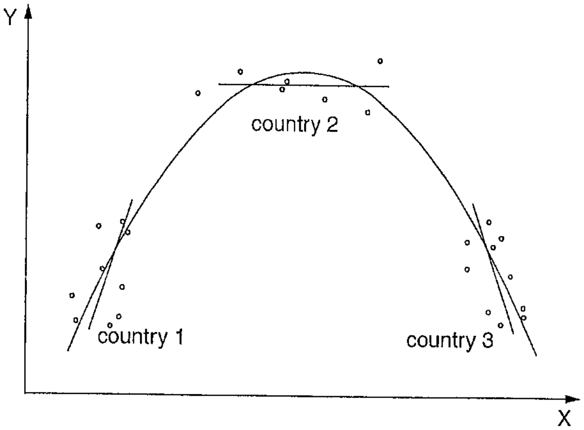

Figure 28.1 Fixed-effects and pooled estimators.

Example 63 *One interesting study illustrating the importance of slope heterogeneity in cross country analysis is the analysis by Haque, Pesaran, and Sharma (2000) on the determinants of cross-country private savings rates, using a subset of data from Masson, Bayoumi, and Samiei (1998) (MBS), on 21 OECD countries over 1971–1993. MBS ran FE regressions of*

PSAV	:	*the private savings rate, defined as the ratio of aggregate private savings to GDP;*

on the explanatory variables

SUR	:	*the ratio of general government budget surplus to GDP;*
GCUR	:	*the ratio of the general government current expenditure to GDP;*
GI	:	*the ratio of the general government investment to GDP;*
GR	:	*GDP growth rate;*
RINT	:	*real interest rate;*
INF	:	*inflation rate;*
PCTT	:	*percentage change in terms of trade;*
YRUS	:	*per capita GDP relative to the U.S.;*
DEP	:	*dependency ratio, defined as the ratio of those under 20, 65 and over to those aged 20–64;*
W	:	*ratio of private wealth (measured as the cumulative sum of past nominal private savings) to GDP.*

Table 28.1 contains the FE regression for the industrial countries. We refer to this specification as model M_0. The estimates under 'model M_0' in Table 28.1 are identical to those reported in column 1 of Table 3 in MBS (1998), except for a few typos. Apart from the coefficient of the GDP growth

rate (GR), all the estimated coefficients are statistically (some very highly) significant, and in particular suggest a strong quadratic relationship between saving and per-capita income. However, the validity of these estimates and the inferences based on them critically depend on the extent to which slope coefficients differ across countries, and in the case of static models, whether these differences are systematic. As shown above, one important implication of neglected slope heterogeneity is the possibility of obtaining spurious nonlinear effects. This possibility is explored by adding quadratic terms in W, INF, PCTT, and DEP to the regressors already included in model M_0. *Estimation results, reported under 'model* M_1*' in Table 28.1, show that the quadratic terms are all statistically highly significant. While there may be some a priori argument for a nonlinear wealth effect in the savings equation, the rationale for nonlinear effects in the case of the other three variables seems less clear. The quadratic relationships between the private savings rate and the variables W, PCTT, and DEP are in fact much stronger than the quadratic relationship between savings and per capita income that MBS focus on. The* $\bar{R}^2$ *of the augmented model, 0.801, is also appreciably larger than that obtained for model* M_0, *0.766. A similar conclusion is reached using other model selection criteria such as the Akaike information criterion (AIC) and the Schwarz Bayesian criterion (SBC) also reported in Table 28.1. As an alternative to the quadratic specifications used in model* M_1, *the authors investigate the possibility that the slope coefficients in each country are fixed over time, but are allowed to vary across countries linearly with the sample means of their wealth to GDP ratio or their per-capita income. More specifically, denote the vector of slope coefficients for country i by* $\boldsymbol{\beta}_i$, *and define*

$$\overline{W}_i = T^{-1}\sum_{t=1}^{T} W_{it}, \text{ and } \overline{YRUS}_i = T^{-1}\sum_{t=1}^{T} YRUS_{it}.$$

Then, slope heterogeneity is modelled by

$$\boldsymbol{\beta}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_{01}\overline{W}_i + \boldsymbol{\beta}_{02}\overline{YRUS}_i. \quad (28.32)$$

Substituting the above expression for $\boldsymbol{\beta}_i$ *in the FE specification, yields*

$$y_{it} = \mu_i + \boldsymbol{\beta}_0'\mathbf{x}_{it} + \boldsymbol{\beta}_{01}'(\mathbf{x}_{it}\overline{W}_i) + \boldsymbol{\beta}_{02}'(\mathbf{x}_{it}\overline{YRUS}_i) + u_{it},$$

where $y_{it} = PSAV_{it}$,

$$\mathbf{x}_{it} = (SUR_{it}, GCUR_{it}, GI_{it}, GR_{it}, RINT_{it}, W_{it}, INF_{it}, PCTT_{it}, YRUS_{it}, DEP_{it})'.$$

The estimated elements of $\boldsymbol{\beta}_0, \boldsymbol{\beta}_{01}$, *and* $\boldsymbol{\beta}_{02}$ *together with their t-ratios are given in Table 28.2. Apart from the coefficient of the SUR variable, all the other coefficients show systematic variation across countries. The coefficient of the SUR variable seems to be least affected by slope heterogeneity, and the hypothesis of slope homogeneity cannot be rejected in the case of this variable. However, none of the other estimates is directly comparable to the FE estimates given in Table 28.1. In particular, the coefficients of output growth variables (*GR_{it} *and* $GR_{it} \times \overline{W}_i$*) are both statistically significant, while this was not so in the case of the FE estimates in Table 28.1. Care must also be exercised when interpreting these estimates. For example, the results suggest that the effect of real output growth on the savings rate is likely to be higher in a country with a high wealth–GDP ratio. Similarly, inflation*

Table 28.1 Fixed-effects estimates of static private saving equations, models M_0 and M_1 (21 OECD countries, 1971–1993)

	Model M_0		Model M_1	
Regressors	Linear Terms	Quadratic Terms	Linear Terms	Quadratic Terms
SUR	−0.574	–	−0.58	–
	(−9.39)	–	(−10.30)	–
GCUR	−0.467	–	−0.521	–
	(−11.30)	–	(−13.39)	–
GI	−0.603	–	−0.701	–
	(−5.71)	–	(−6.92)	–
GR	−0.060	–	−0.065	–
	(−1.14)	–	(−1.33)	
RINT	0.212	–	0.281	
	(4.40)	–	(5.90)	
W	0.023	–	0.175	−0.00025
	(5.11)	–	(8.38)	(−7.69)
INF	0.180	–	−0.041	0.011
	(4.63)		(−0.53)	(3.29)
PCTT	0.047		0.063	−0.0013
	(3.07)		(4.11)	(−2.81)
YRUS	0.586	−0.0048	0.286	−0.0026
	(3.41)	(−3.90)	(1.70)	(−2.15)
DEP	−0.118	–	−1.201	0.0073
	(−4.12)		(−5.25)	(4.85)
$\bar{R}^2$	0.766		0.801	
$\hat{\sigma}$	2.325		2.145	
LL	−1076.4		−1035.3	
AIC	−1108.4		−1071.3	
SBC	−1165.3		−1146.5	

*The dependent variable (PSAV) is the ratio of private savings to GNP. Model M_0 is the specification estimated by Masson et al. (1998), see column 1 of Table 3 in that paper. The figures in brackets are t-ratios. $\bar{R}$ is the adjusted multiple correlation coefficient, $\hat{\sigma}$ is the standard error of the regression; LL is the maximazed value of the log-likehood function; AIC is the Akaike information criterion, and SBC is the Schwarz Bayesian criterion.

effects on the savings rate are estimated to be higher in countries with higher wealth to GDP ratios. However, these results do not predict, for instance, that an individual country's savings rate will necessarily rise with output growth.

For further discussion on the consequences of ignoring parameter heterogeneity see, for example, Robertson and Symons (1992) and Haque, Pesaran, and Sharma (2000).

28.4 The Swamy estimator

Consider the panel data model

$$y_{it} = \boldsymbol{\beta}_i' \mathbf{x}_{it} + u_{it}, \tag{28.33}$$

Table 28.2 Fixed-effects estimates of private savings equations with cross-sectionally varying slopes, (Model M2), (21 OECD countries, 1971–1993)

Regressors	$\hat{\beta}_0$	$\hat{\beta}_{01}$	$\hat{\beta}_{02}$
SUR	−0.625	–	–
	(−12.10)		
GCUR	−1.146	0.0022	–
	(−6.91)	(4.26)	
GI	−1.891	0.0039	–
	(−2.44)	(1.60)	
GR	−0.744	0.0023	
	(−2.69)	(2.71)	–
RINT	0.417	–	−0.0052
	(4.36)	–	(−3.53)
W	0.119	−0.00033	–
	(5.28)	(−4.70)	–
INF	−0.860	0.0031	–
	(−5.29)	(6.29)	
PCTT	−0.214	0.00083	–
	(−1.88)	(2.30)	
YRUS	1.435	−0.0046	–
	(6.31)	(−6.72)	
DEP	0.502	−0.0021	–
	(2.54)	(−3.39)	
$\bar{R}^2$	0.838		
$\hat{\sigma}$	1.934		
LL	−982.9		
AIC	−1022.9		
SBC	−1106.5		

*See the notes to Table 28.1

under the Swamy (1970) random coefficient scheme (28.3), where η_i satisfies assumptions (28.4)–(28.5). For simplicity, we also assume that u_{it} is *independently* distributed across i and over t with zero mean and $Var(u_{it}) = \sigma_i^2$. Substituting $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\eta}_i$ into (28.33) we obtain, using stacked form notation,

$$\mathbf{y}_{i.} = \mathbf{X}_{i.}\boldsymbol{\beta} + \mathbf{v}_{i.},$$

where the composite error, $\mathbf{v}_{i.}$, is given by

$$\mathbf{v}_{i.} = \mathbf{X}_{i.}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_{i.}.$$

Stacking the regression equations by cross-sectional units we now have

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v},$$

where

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_{1.} \\ \mathbf{y}_{2.} \\ \vdots \\ \mathbf{y}_{N.} \end{pmatrix}, \mathbf{X} = \begin{pmatrix} \mathbf{X}_{1.} \\ \mathbf{X}_{2.} \\ \vdots \\ \mathbf{X}_{N.} \end{pmatrix}, \text{ and } \mathbf{v} = \begin{pmatrix} \mathbf{v}_{1.} \\ \mathbf{v}_{2.} \\ \vdots \\ \mathbf{v}_{N.} \end{pmatrix}.$$

Suppose we are interested in estimating the mean coefficient vector, $\boldsymbol{\beta}$, and the covariance matrix of $\mathbf{v}$, $\boldsymbol{\Sigma}$, given by

$$\boldsymbol{\Sigma} = E\left(\mathbf{v}\mathbf{v}'\right) = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Sigma}_N \end{pmatrix},$$

where

$$\boldsymbol{\Sigma}_i = Var\left(\mathbf{v}_{i.}\right) = \sigma_i^2 \mathbf{I}_T + \mathbf{X}_{i.}\boldsymbol{\Omega}_\eta \mathbf{X}'_{i.}.$$

For known values of $\boldsymbol{\Omega}_\eta$ and σ_i^2, the best linear unbiased estimator of $\boldsymbol{\beta}$ is given by the generalized least squares (GLS) estimator, known in this case as the Swamy estimator

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{SW} &= \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}, \\ &= \left(\sum_{i=1}^N \mathbf{X}'_{i.}\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_{i.}\right)^{-1} \sum_{i=1}^N \mathbf{X}'_{i.}\boldsymbol{\Sigma}_i^{-1}\mathbf{y}_{i.}. \end{aligned}$$

It is easily seen that (under the assumption that $\boldsymbol{\Omega}_\eta$ is nonsingular) (see property (A.9) in Appendix A)

$$\boldsymbol{\Sigma}_i^{-1} = \frac{\mathbf{I}_T}{\sigma_i^2} - \frac{\mathbf{X}_{i.}}{\sigma_i^2}\left(\frac{\mathbf{X}'_{i.}\mathbf{X}_{i.}}{\sigma_i^2} + \boldsymbol{\Omega}_\eta^{-1}\right)^{-1} \frac{\mathbf{X}'_{i.}}{\sigma_i^2}.$$

Note that $\boldsymbol{\Sigma}_i^{-1}$ exists even if $\boldsymbol{\Omega}_\eta$ is singular. In general we can write[4]

$$\boldsymbol{\Sigma}_i^{-1} = \frac{\mathbf{I}_T}{\sigma_i^2} - \frac{\mathbf{X}_{i.}\boldsymbol{\Omega}_\eta}{\sigma_i^2}\left(\mathbf{I}_k + \frac{\mathbf{X}'_{i.}\mathbf{X}_{i.}}{\sigma_i^2}\boldsymbol{\Omega}_\eta\right)^{-1} \frac{\mathbf{X}'_{i.}}{\sigma_i^2},$$

which is valid irrespective of whether $\boldsymbol{\Omega}_\eta$ is singular or not. Let

$$\mathbf{Q}_{iT} = \frac{\mathbf{X}'_{i.}\mathbf{X}_{i.}}{T\sigma_i^2},\ \mathbf{q}_{iT} = \frac{\mathbf{X}'_{i.}\mathbf{y}_{i.}}{T\sigma_i^2},$$

$$\mathbf{H}_{iT} = \mathbf{Q}_{iT} + \frac{1}{T}\boldsymbol{\Omega}_\eta^{-1}.$$

[4] In formula (A.9), let $\mathbf{X} = \mathbf{X}_i\boldsymbol{\Omega}_\eta$, $\mathbf{Y} = \mathbf{X}'_i$, $\mathbf{C} = \frac{\mathbf{I}_T}{\sigma_i^2}$, and $\mathbf{D} = \mathbf{I}_k$, then the desired result follows.

Then

$$\frac{\mathbf{X}_{i.}'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_{i.}}{T} = \mathbf{Q}_{iT} - \mathbf{Q}_{iT}\mathbf{H}_{iT}^{-1}\mathbf{Q}_{iT},$$

and

$$\frac{\mathbf{X}_{i.}'\boldsymbol{\Sigma}_i^{-1}\mathbf{y}_{i.}}{T} = \mathbf{q}_{iT} - \mathbf{Q}_{iT}\mathbf{H}_{iT}^{-1}\mathbf{q}_{iT}.$$

It follows that the Swamy estimator can also be written as

$$\hat{\boldsymbol{\beta}}_{SW} = \left[\sum_{i=1}^{N}\left(\mathbf{Q}_{iT} - \mathbf{Q}_{iT}\mathbf{H}_{iT}^{-1}\mathbf{Q}_{iT}\right)\right]^{-1}\sum_{i=1}^{N}\left(\mathbf{q}_{iT} - \mathbf{Q}_{iT}\mathbf{H}_{iT}^{-1}\mathbf{q}_{iT}\right). \tag{28.34}$$

By repeatedly utilizing the identity relation (A.9) in Appendix A, we obtain

$$\hat{\bar{\boldsymbol{\beta}}}_{SW} = \sum_{i=1}^{N}\mathbf{R}_i\hat{\boldsymbol{\beta}}_i,$$

where

$$\mathbf{R}_i = \left[\sum_{i=1}^{N}\left(\boldsymbol{\Omega}_\eta + \boldsymbol{\Sigma}_{\hat{\beta}_i}\right)^{-1}\right]^{-1}\left(\boldsymbol{\Omega}_\eta + \boldsymbol{\Sigma}_{\hat{\beta}_i}\right)^{-1}, \tag{28.35}$$

and

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{X}_{i.}'\mathbf{X}_{i.})^{-1}\mathbf{X}_{i.}'\mathbf{y}_{i.},\ \boldsymbol{\Sigma}_{\hat{\beta}_i} = Var\left(\hat{\boldsymbol{\beta}}_i\right) = \sigma_i^2(\mathbf{X}_{i.}'\mathbf{X}_{i.})^{-1}. \tag{28.36}$$

The expression (28.34) shows that the Swamy estimator is a matrix weighted average of the least squares estimator for each cross-sectional unit (28.36), with the weights inversely proportional to their covariance matrices. It also shows that the *GLS* estimator requires only a matrix inversion of order k, and so it is not much more complicated to compute than the sample least squares estimator.

The covariance matrix of the SW estimator is

$$Var\left(\hat{\bar{\boldsymbol{\beta}}}_{SW}\right) = \left(\sum_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right)^{-1} = \left[\sum_{i=1}^{N}\left(\boldsymbol{\Omega}_\eta + \boldsymbol{\Sigma}_{\hat{\beta}_i}\right)^{-1}\right]^{-1}. \tag{28.37}$$

If errors u_{it} and η_i are normally distributed, the SW estimator is the same as the maximum likelihood (*ML*) estimator of $\boldsymbol{\beta}$ conditional on $\boldsymbol{\Omega}_\eta$ and σ_i^2. Without knowledge of $\boldsymbol{\Omega}_\eta$ and σ_i^2, we can estimate $\boldsymbol{\beta}$, $\boldsymbol{\Omega}_\eta$ and σ_i^2, $i = 1, 2, \ldots, N$ simultaneously by the *ML* method. However, it can be computationally tedious. A natural alternative is to first estimate $\boldsymbol{\Sigma}_i$, then substitute the estimated $\boldsymbol{\Sigma}_i$ into (28.37).

Swamy proposes using the least squares estimator of $\boldsymbol{\beta}_i$, $\hat{\boldsymbol{\beta}}_i = (\mathbf{X}'_{i.}\mathbf{X}_{i.})^{-1}\mathbf{X}'_{i.}\mathbf{y}_{i.}$ and the residuals $\hat{\mathbf{u}}_{i.} = \mathbf{y}_{i.} - \mathbf{X}_{i.}\hat{\boldsymbol{\beta}}_i$ to obtain consistent estimators of σ_i^2, for $i = 1, \ldots, N$, and $\boldsymbol{\Omega}_\eta$. Noting that

$$\hat{\mathbf{u}}_{i.} = [\mathbf{I}_T - \mathbf{X}_{i.}(\mathbf{X}'_{i.}\mathbf{X}_{i.})^{-1}\mathbf{X}'_{i.}]\mathbf{u}_{i.}, \tag{28.38}$$

and

$$\hat{\boldsymbol{\beta}}_i = \boldsymbol{\beta}_i + (\mathbf{X}'_{i.}\mathbf{X}_{i.})^{-1}\mathbf{X}'_{i.}\mathbf{u}_{i.}, \tag{28.39}$$

we obtain the unbiased estimators of σ_i^2 and $\boldsymbol{\Omega}_\eta$ as

$$\begin{aligned}\hat{\sigma}_i^2 &= \frac{\hat{\mathbf{u}}'_{i.}\hat{\mathbf{u}}_{i.}}{T-k}, \qquad (28.40)\\ &= \frac{1}{T-k}\mathbf{y}'_i[\mathbf{I}_T - \mathbf{X}_{i.}(\mathbf{X}'_{i.}\mathbf{X}_{i.})^{-1}\mathbf{X}'_{i.}]\mathbf{y}_{i.},\\ \hat{\boldsymbol{\Omega}}_\eta &= \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\beta}}_i - N^{-1}\sum_{j=1}^{N}\hat{\boldsymbol{\beta}}_j\right)\left(\hat{\boldsymbol{\beta}}_i - N^{-1}\sum_{j=1}^{N}\hat{\boldsymbol{\beta}}_j\right)'\\ &\quad - \frac{1}{TN}\sum_{i=1}^{N}\hat{\sigma}_i^2\left(\frac{\mathbf{X}'_{i.}\mathbf{X}_{i.}}{T}\right)^{-1}. \qquad (28.41)\end{aligned}$$

Just as in the error-components model, the estimator (28.41) is not necessarily non-negative definite. In this situation, Swamy has suggested replacing (28.41) by

$$\hat{\boldsymbol{\Omega}}_\eta^* = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\beta}}_i - N^{-1}\sum_{j=1}^{N}\hat{\boldsymbol{\beta}}_j\right)\left(\hat{\boldsymbol{\beta}}_i - N^{-1}\sum_{j=1}^{N}\hat{\boldsymbol{\beta}}_j\right)'. \tag{28.42}$$

This estimator, although biased, is nonnegative definite and consistent when T tends to infinity.

For further discussion on the above estimator see Swamy (1970), and Hsiao and Pesaran (2008).

28.5 The mean group estimator (*MGE*)

One alternative to Swamy's estimator of $\boldsymbol{\beta}$ in equation (28.33) is the mean group (*MG*) estimator, proposed by Pesaran and Smith (1995) for estimation of dynamic random coefficient models. The *MG* estimator is defined as the simple average of the *OLS* estimators, $\hat{\boldsymbol{\beta}}_i$

$$\hat{\boldsymbol{\beta}}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i,$$

where

$$\hat{\beta}_i = \left(\mathbf{X}_{i.}'\mathbf{X}_{i.}\right)^{-1}\mathbf{X}_{i.}'\mathbf{y}_{i.}.$$

MG estimation is possible when both *T* and *N* are sufficiently large, and is applicable irrespective of whether the slope coefficients are random (in Swamy's sense), or fixed in the sense that the diversity in the slope coefficients across cross sectional units cannot be captured by means of a finite parameter probability distribution. To compute the variance of the *MG* estimator, first note that

$$\hat{\beta}_i = \beta + \eta_i + \xi_{i.},$$

where

$$\xi_{i.} = \left(\mathbf{X}_{i.}'\mathbf{X}_{i.}\right)^{-1}\mathbf{X}_{i.}'\mathbf{u}_{i.},$$

$$\hat{\beta}_{MG} = \beta + \overline{\eta} + \overline{\xi}, \tag{28.43}$$

and

$$\overline{\eta} = \frac{1}{N}\sum_{i=1}^{N}\eta_i, \quad \overline{\xi} = \frac{1}{N}\sum_{i=1}^{N}\xi_{i.}.$$

Hence, when the regressors are strictly exogenous and the errors, u_{it}, are independently distributed, the variance of $\hat{\beta}_{MG}$ is

$$\begin{aligned} Var\left(\hat{\beta}_{MG}\right) &= Var\left(\overline{\eta}\right) + Var\left(\overline{\xi}\right) \\ &= \frac{1}{N}\boldsymbol{\Omega}_\eta + \frac{1}{N^2}\sum_{i=1}^{N}\sigma_i^2 E\left[\left(\frac{\mathbf{X}_{i.}'\mathbf{X}_{i.}}{T}\right)^{-1}\right]. \end{aligned}$$

An unbiased estimator of the covariance matrix of $\hat{\beta}_{MG}$ can be computed as

$$\widehat{Var}\left(\hat{\beta}_{MG}\right) = \frac{1}{N(N-1)}\sum_{i=1}^{N}\left(\hat{\beta}_i - \hat{\beta}_{MG}\right)\left(\hat{\beta}_i - \hat{\beta}_{MG}\right)'.$$

For a proof, first note that

$$\hat{\beta}_i - \hat{\beta}_{MG} = \left(\eta_i - \overline{\eta}\right) + \left(\xi_{i.} - \overline{\xi}\right),$$

$$\begin{aligned} \left(\hat{\beta}_i - \hat{\beta}_{MG}\right)\left(\hat{\beta}_i - \hat{\beta}_{MG}\right)' &= \left(\eta_i - \overline{\eta}\right)\left(\eta_i - \overline{\eta}\right)' + \left(\xi_{i.} - \overline{\xi}\right)\left(\xi_{i.} - \overline{\xi}\right)' \\ &\quad + \left(\eta_i - \overline{\eta}\right)\left(\xi_{i.} - \overline{\xi}\right)' + \left(\xi_{i.} - \overline{\xi}\right)\left(\eta_i - \overline{\eta}\right)', \end{aligned}$$

and

$$\sum_{i=1}^{N} E\left[\left(\hat{\beta}_i - \hat{\beta}_{MG}\right)\left(\hat{\beta}_i - \hat{\beta}_{MG}\right)'\right] = (N-1)\mathbf{\Omega}_\eta + \left(1 - \frac{1}{N}\right)\sum_{i=1}^{N}\sigma_i^2 E\left[\left(\mathbf{X}_{i.}'\mathbf{X}_{i.}\right)^{-1}\right].$$

Using the above results it is now easily seen that

$$E\left[\widehat{Var}\left(\hat{\beta}_{MG}\right)\right] = Var\left(\hat{\beta}_{MG}\right),$$

as required. For a further discussion of the mean group estimator, see Pesaran and Smith (1995), and Hsiao and Pesaran (2008).

Example 64 *Continuing from Example 63, Haque, Pesaran, and Sharma (2000) further investigate the determinants of cross country private savings rates by carrying out a country-specific analysis. The FE regression in Table 28.2 assumes that the slope coefficients across countries are exact linear functions of* $\overline{W}_i$ *and/or* $\overline{YRUS}_i$ *(see equation 28.32), and that the error variances,* $Var(u_{it}) = \sigma_i^2$*, are the same across countries. Clearly, these are rather restrictive assumptions, and the consequences of incorrectly imposing them on the parameters of interest need to be examined. Under the alternative assumption of unrestricted slope and error variance heterogeneity, MG estimates can be computed as simple averages of country-specific estimates from country-specific regressions and can then be used to make inferences about* $E(\boldsymbol{\beta}_i) = \boldsymbol{\beta}$*. Results on country-specific estimates and MG estimates are summarized in Table 28.3. The estimated slope coefficients differ considerably across countries, both in terms of their magnitude and their statistical significance. Some of the coefficients are statistically significant only in the case of 3 or 4 countries and in general are very poorly estimated. This is true of the coefficients of GI, GR, W, PCTT, and YRUS. Also the sign of these estimated coefficients varies quite widely across countries. The coefficients of RINT and INF are better estimated, but still differ significantly both in magnitude and in sign across the countries. Only the coefficients of SUR and GCUR tend to be similar across countries. The coefficient of SUR is estimated to be negative in 19 of the 20 countries, and 13 of these are statistically significant. The positive estimate obtained for New Zealand is very small and not statistically significant. Similarly, 17 out of 20 coefficients estimated for the GCUR variable have a negative sign, with 7 of the 17 negative coefficients statistically significant. None of the three positive coefficients estimated for GCUR is statistically significant. The MG estimates based on the individual country regressions in Table 28.3 support these general conclusions. Only the MG estimates of the SUR and the GCUR variables are statistically significant (see the last two rows of Table 28.3). At −0.671, the MGE of the SUR variable is only marginally higher than the corresponding FE estimate in Table 28.2 that allows for some slope heterogeneity.*

28.5.1 Relationship between Swamy's and *MG* estimators

The Swamy and *MG* estimators are algebraically equivalent when T is sufficiently large. To see this, consider $\hat{\boldsymbol{\beta}}_{SW}$ in equation (28.34), and note that

$$\mathbf{H}_{iT}^{-1} = \left(\mathbf{Q}_{iT} + \frac{1}{T}\mathbf{\Omega}_\eta^{-1}\right)^{-1} = \mathbf{Q}_{iT}^{-1}\left(\mathbf{I}_k + \frac{1}{T}\mathbf{\Omega}_\eta^{-1}\mathbf{Q}_{iT}^{-1}\right)^{-1}.$$

Table 28.3 Country-specific estimates of 'static' private saving equations (20 OECD countries, 1972–1993)

Country	SUR	CCUR	GI	GR	RINT	W	INF	PCTT	YRUS	DEP
Australia	−0.81	−0.18	−1.00	0.08	0.18	0.06	0.27	0.04	0.42	0.46
	[0.18]	[0.27]	[0.41]	[0.08]	[0.08]	[0.02]	[0.09]	[0.03]	[0.17]	[0.22]
Austria	−0.48	−0.42	0.35	0.06	0.24	0004	0.09	0.11	−0.10	−0.03
	[0.56]	[0.40]	[0.84]	[0.32]	[0.32]	[0.05]	[0.54]	[0.16]	[0.24]	[0.21]
Belgium	−0.68	−0.53	−2.47	0.09	−0.04	−0.02	−0.10	−0.00	0.17	−0.22
	[0.23]	[0.15]	[1.51]	[0.11]	[0.14]	[0.03]	[0.13]	[0.02]	[0.09]	[0.35]
Canada	−1.31	−0.56	1.01	0.24	0.10	−0.03	0.29	0.17	0.07	−0.17
	[0.10]	[0.14]	[1.03]	[0.09]	[0.08]	[0.04]	[0.09]	[0.05]	[0.12]	[0.12]
Denmark	−1.08	−0.64	0.36	0.03	−0.20	−0.01	0.10	0.02	0.14	−1.17
	[0.15]	[0.22]	[0.80]	[0.25]	[0.20]	[0.03]	[0.29]	[0.05]	[0.24]	[0.36]
Finland	−0.70	−0.35	0.87	0.14	0.40	0.03	0.52	0.01	0.02	−0.39
	[0.16]	[0.21]	[1.59]	[0.20]	[0.18]	[0.03]	[0.22]	[0.02]	[0.19]	[0.52]
France	−1.45	−0.78	−3.13	0.10	−0.16	−0.04	−0.22	−0.06	0.12	−0.16
	[0.51]	[0.52]	[2.00]	[0.23]	[0.18]	[0.10]	[0.24]	[0.05]	[0.12]	[0.43]
Germany	−0.80	−0.54	−0.18	0.19	−0.06	0.00	0.02	−0.01	−0.10	−0.28
	[0.35]	[0.28]	[0.71]	[0.18]	[0.17]	[0.03]	[0.25]	[0.05]	[0.20]	[0.11]
Greece	−0.69	−0.29	−1.13	0.15	1.23	0.10	1.05	−0.49	−0.87	1.52
	[0.45]	[0.71]	[1.65]	[0.34]	[0.58]	[0.05]	[0.63]	[0.27]	[1.29]	[1.24]
Ireland	−0.48	−0.50	1.33	−0.08	−0.71	−0.13	−0.88	0.32	0.79	1.14
	[0.29]	[0.14]	[1.18]	[0.14]	[0.28]	[0.06]	[0.22]	[0.11]	[0.24]	[0.35]
Italy	−0.46	0.05	−0.16	0.13	0.12	−0.00	0.09	−0.00	−0.12	0.32
	[0.18]	[0.21]	[0.48]	[0.15]	[0.11]	[0.03]	[0.13]	[0.04]	[0.15]	[0.19]
Japan	−0.58	−0.79	−0.98	−0.14	−0.05	0.04	0.01	0.04	−0.06	0.22
	[0.21]	[0.31]	[0.50]	[0.12]	[0.16]	[0.03]	[0.09]	[0.01]	[0.08]	[0.32]
Netherlands	−0.75	−0.43	−1.50	−0.05	0.09	0.12	−0.37	0.06	0.30	0.22
	[0.33]	[0.33]	[2.64]	[0.20]	[0.28]	[0.05]	[0.27]	[0.15]	[0.26]	[0.39]
New Zealand	[0.02]	−0.54	−1.22	−0.12	−0.07	0.02	−0.20	0.07	−0.46	0.24
	[0.29]	[0.45]	[0.78]	[0.22]	[0.20]	[0.03]	[0.19]	[0.07]	[0.33]	[0.18]
Norway	−0.22	0.13	−0.15	−0.06	0.02	−0.07	−0.04	0.23	0.12	−0.16
	[0.51]	[0.66]	[0.61]	[0.46]	[0.51]	[0.05]	[0.60]	[0.07]	[0.31]	[0.64]
Portugal	−1.00	−0.57	2.91	0.60	0.47	−0.07	0.64	0.21	−0.72	0.16
	[0.20]	[0.32]	[1.64]	[0.24]	[0.20]	[0.05]	[0.19]	[0.13]	[0.37]	[0.41]
Spain	−0.18	−0.06	1.36	−0.01	0.07	−0.09	0.11	0.18	−0.78	−0.28
	[0.55]	[0.59]	[1.58]	[0.31]	[0.38]	[0.05]	[0.42]	[0.12]	[0.32]	[0.57]
Sweden	−0.84	−0.96	−2.54	−0.53	0.24	0.05	−0.02	0.09	0.00	0.22
	[0.11]	[0.20]	[1.49]	[0.30]	[0.23]	[0.05]	[0.23]	[0.10]	[0.22]	[0.81]
Switzerland	−0.22	−0.09	0.36	−0.26	0.02	0.06	0.21	−0.04	−0.06	−0.59
	[0.50]	[0.16]	[0.76]	[0.13]	[0.14]	[0.03]	[0.11]	[0.5]	[0.12]	[0.09]
UK	−0.72	0.03	−0.79	0.37	0.18	−0.04	0.21	0.01	−0.25	0.34
	[0.12]	[0.10]	[0.34]	[0.09]	[0.08]	[0.03]	[0.08]	[0.04]	[0.15]	[0.15]
Average	−0.671	−0.401	−0.335	0.046	0.104	0.001	0.089	0.048	−0.069	0.080
Standard error	[.083]	[.067]	[.332]	[.052]	[.081]	[.014]	[.088]	[.036]	[.127]	[.090]

Table 28.3 Continued

	$\hat{\sigma}^2$	$\chi^2_{SC}(1)$	$\chi^2_{FF}(1)$	$\chi^2_N(2)$	$\chi^2_H(1)$	$\bar{R}^2$	LL
Australia	0.573	0.83	0.29	0.70	1.24	0.90	−11.36
Austria	1.210	0.05	2.69	1.56	0.10	0.28	−27.78
Belgium	0.693	18.10	2.84	1.03	0.81	0.69	−15.51
Canada	0.518	0.01	1.27	0.18	0.22	0.76	−9.10
Denmark	1.197	1.63	0.20	1.56	2.32	0.49	−27.55
Finland	1.079	8.32	2.44	1.78	0.38	0.70	−25.27
France	0.689	1.78	12.51	1.80	1.13	0.54	−15.40
Germany	0.817	10.02	0.00	0.76	0.48	0.16	−19.15
Greece	2.439	6.25	0.09	1.05	0.32	0.53	−43.21
Ireland	1.469	3.04	0.59	1.49	0.01	0.77	−32.06
Italy	0.606	2.80	5.09	0.74	2.76	0.72	−12.57
Japan	0.399	0.39	1.59	4.97	0.12	0.77	−3.37
Netherlands	1.052	3.40	1.57	0.20	2.02	0.52	−24.70
New Zealand	1.743	12.38	8.26	0.68	10.45	0.70	−35.82
Norway	1.622	8.47	2.18	0.81	0.53	0.39	−34.23
Portugal	2.042	0.00	1.67	0.80	0.69	0.86	−39.30
Spain	1.319	8.68	5.47	1.20	0.40	0.58	−29.68
Swedan	1.194	6.97	0.76	0.15	1.67	0.68	−27.49
Switzerland	0.535	3.13	3.40	0.70	7.79	0.44	−9.83
UK	0.541	1.63	2.20	1.38	0.50	0.81	−10.07

** $\hat{\sigma}$ is the standard error of the country specific regressions, $\chi^2_{SC}(1)$, $\chi^2_{FF}(1)$, $\chi^2_N(2)$ and $\chi^2_H(1)$ are chi-aquared statistics for tests of residual serial correlation, functional form mis-specification, non-normal errors and heteroskedasticity. The figures in brackets are their degrees of freedom. $\bar{R}$ is the adjusted multiple correlation coefficient, and LL is the maximized log-likelihood value of the country-specific regressions.

Write

$$\mathbf{\Omega}_\eta^{-1} = \lambda \mathbf{A},$$

where λ represents an overall index of parameter heterogeneity, such that

$$\lambda \to 0, \text{ highest degree of heterogeneity,}$$
$$\lambda \to \infty, \text{ homogeneity.}$$

Then $\hat{\boldsymbol{\beta}}_{SW}$ can be written as

$$\hat{\boldsymbol{\beta}}_{SW} = \left\{ \sum_{i=1}^{N} \left[\mathbf{Q}_{iT} - \left(\mathbf{I}_k + \frac{\lambda}{T} \mathbf{A} \mathbf{Q}_{iT}^{-1} \right)^{-1} \mathbf{Q}_{iT} \right] \right\}^{-1} \sum_{i=1}^{N} \left[\mathbf{q}_{iT} - \left(\mathbf{I}_k + \frac{\lambda}{T} \mathbf{A} \mathbf{Q}_{iT}^{-1} \right)^{-1} \mathbf{q}_{iT} \right].$$

For a fixed N and T, and for a sufficiently small λ

$$\left(\mathbf{I}_k + \frac{\lambda}{T}\mathbf{A}\mathbf{Q}_{iT}^{-1}\right)^{-1} = \mathbf{I}_k - \frac{\lambda}{T}\mathbf{G}_{iT} + \left(\frac{\lambda}{T}\right)^2 \mathbf{G}_{iT}^2 - \ldots$$

where $\mathbf{G}_{iT} = \mathbf{A}\mathbf{Q}_{iT}^{-1}$. Therefore,

$$\begin{aligned}
\hat{\beta}_{SW} &= \left\{\sum_{i=1}^{N}\left[\mathbf{Q}_{iT} - \left(\mathbf{I}_k - \frac{\lambda}{T}\mathbf{G}_{iT} + \left(\frac{\lambda}{T}\right)^2 \mathbf{G}_{iT}^2 + \ldots\right)\mathbf{Q}_{iT}\right]\right\}^{-1} \\
&\quad \sum_{i=1}^{N}\left[\mathbf{q}_{iT} - \left(\mathbf{I}_k - \frac{\lambda}{T}\mathbf{G}_{iT} + \left(\frac{\lambda}{T}\right)^2 \mathbf{G}_{iT}^2 + \ldots\right)\mathbf{q}_{iT}\right]. \\
&= \left\{\sum_{i=1}^{N}\mathbf{G}_{iT}\mathbf{Q}_{iT} - \frac{\lambda}{T}\sum_{i=1}^{N}\mathbf{G}_{iT}^2\mathbf{Q}_{iT} + O\left[\left(\frac{\lambda}{T}\right)^2\right]\right\}^{-1} \\
&\quad \left\{\sum_{i=1}^{N}\mathbf{G}_{iT}\mathbf{q}_{iT} - \frac{\lambda}{T}\sum_{i=1}^{N}\mathbf{G}_{iT}^2\mathbf{q}_{iT} + O\left[\left(\frac{\lambda}{T}\right)^2\right]\right\}.
\end{aligned}$$

Hence for any fixed $T > k$ and for any N, as $\lambda \to 0$,

$$\hat{\beta}_{SW} \to \left(\sum_{i=1}^{N}\mathbf{G}_{iT}\mathbf{Q}_{iT}\right)^{-1}\sum_{i=1}^{N}\mathbf{G}_{iT}\mathbf{q}_{iT}.$$

However, note that

$$\begin{aligned}
\left(\sum_{i=1}^{N}\mathbf{G}_{iT}\mathbf{Q}_{iT}\right)^{-1}\sum_{i=1}^{N}\mathbf{G}_{iT}\mathbf{q}_{iT} &= \left(\sum_{i=1}^{N}\mathbf{A}\mathbf{Q}_{iT}^{-1}\mathbf{Q}_{iT}\right)^{-1}\sum_{i=1}^{N}\mathbf{A}\mathbf{Q}_{iT}^{-1}\mathbf{q}_{iT} \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbf{Q}_{iT}^{-1}\mathbf{q}_{iT} = \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i = \hat{\beta}_{MG},
\end{aligned}$$

From which it follows that

$$\lim_{\lambda\to 0}\hat{\beta}_{SW}(\lambda) = \hat{\beta}_{MG},$$

and, for all values of N and $\lambda > 0$,

$$\lim_{T\to\infty}\left(\hat{\beta}_{SW}(\lambda) - \hat{\beta}_{MG}\right) = \mathbf{0}.$$

28.6 Dynamic heterogeneous panels

Consider the $ARDL(p, \underbrace{q, q, \ldots, q}_{k\text{-times}})$ model (see Chapter 6 for an introduction to *ARDL* models)

$$y_{it} = \alpha_i + \sum_{j=1}^{p} \lambda_{ij} y_{i,t-j} + \sum_{j=0}^{q} \delta'_{ij} \mathbf{x}_{i,t-j} + u_{it}, \text{ for } i = 1, 2, \ldots, N, \tag{28.44}$$

where $\mathbf{x}_{it}$ is a k-dimensional vector of explanatory variables for group i; α_i represent the fixed-effects; the coefficients of the lagged dependent variables, λ_{ij}, are scalars; and δ_{ij} are k-dimensional coefficient vectors. In the following, we assume that the disturbances u_{it}, $i = 1, 2, \ldots, N$; $t = 1, 2, \ldots, T$, are independently distributed across i and t, with zero means, variances σ_i^2, and are distributed independently of the regressors $\mathbf{x}_{it}$.

The error correction representation of the above *ARDL* model is:

$$\Delta y_{it} = \alpha_i + \phi_i y_{i,t-1} + \boldsymbol{\beta}'_i \mathbf{x}_{it} + \sum_{j=1}^{p-1} \lambda^*_{ij} \Delta y_{i,t-j} + \sum_{j=0}^{q-1} \delta^{*\prime}_{ij} \Delta \mathbf{x}_{i,t-j} + u_{it}, \tag{28.45}$$

where

$$\phi_i = -(1 - \sum_{j=1}^{p} \lambda_{ij}), \quad \boldsymbol{\beta}_i = \sum_{j=0}^{q} \delta_{ij},$$

$$\lambda^*_{ij} = -\sum_{m=j+1}^{p} \lambda_{im}, \; j = 1, 2, \ldots, p-1,$$

$$\delta^*_{ij} = -\sum_{m=j+1}^{q} \delta_{im}, \; j = 1, 2, \ldots, q-1.$$

If we stack the time series observations for each group, (28.45) can be written as

$$\Delta \mathbf{y}_{i.} = \alpha_i \boldsymbol{\tau}_T + \phi_i \mathbf{y}_{i.,-1} + \mathbf{X}_{i.} \boldsymbol{\beta}_i + \sum_{j=1}^{p-1} \lambda^*_{ij} \Delta \mathbf{y}_{i.,-j} + \sum_{j=0}^{q-1} \Delta \mathbf{X}_{i.,-j} \delta^*_{ij} + \mathbf{u}_{i.},$$

for $i = 1, 2, \ldots, N$, where $\boldsymbol{\tau}_T$ is a $T \times 1$ vector of ones, $\mathbf{y}_{i.,-j}$ and $\mathbf{X}_{i.,-j}$ are j-period lagged values of $\mathbf{y}_{i.}$ and $\mathbf{X}_{i.}$, $\Delta \mathbf{y}_{i.} = \mathbf{y}_{i.} - \mathbf{y}_{i.,-1}$, $\Delta \mathbf{X}_{i.} = \mathbf{X}_{i.} - \mathbf{X}_{i.,-1}$, $\Delta \mathbf{y}_{i.,-j}$ and $\Delta \mathbf{X}_{i.,-j}$ are j-period lagged values of $\Delta \mathbf{y}_{i.}$ and $\Delta \mathbf{X}_{i.}$.

If the roots of the polynomial

$$f_i(z) = 1 - \sum_{j=1}^{p} \lambda_{ij} z^j = 0,$$

for $i = 1, 2, \ldots, N$, fall outside the unit circle, then the $ARDL(p, q, q, \ldots, q)$ model is stable. In this chapter we will take up this assumption, while the non-stationary case will be discussed in Chapter 31. This condition ensures that $\phi_i < 0$, and that there exists a long-run relationship between y_{it} and $\mathbf{x}_{it}$ defined by (see Sections 6.5 and 22.2)

$$y_{it} = \boldsymbol{\theta}_i' \mathbf{x}_{it} + \eta_{it},$$

for each $i = 1, 2, \ldots, N$, where η_{it} is $I(0)$, and $\boldsymbol{\theta}_i$ are the long-run coefficients on $\mathbf{X}_{i,}$, $\boldsymbol{\theta}_i = -\boldsymbol{\beta}_i/\phi_i$.

28.7 Large sample bias of pooled estimators in dynamic heterogeneous models

Traditional procedures for estimation of pooled models, such as the FE estimator or the IV/GMM approaches reviewed in Chapter 27, can produce inconsistent and potentially misleading estimates of the average value of the parameters in dynamic panel data models unless the slope coefficients are in fact homogeneous. To see this, consider the simple dynamic panel data model ($ARDL(1, 0)$)

$$y_{it} = \alpha_i + \lambda_i y_{i,t-1} + \beta_i x_{it} + u_{it}, \tag{28.46}$$

where the slopes, λ_i and β_i, as well as the intercepts, α_i, are allowed to vary across cross-sectional units (groups). Here, for simplicity, x_{it} is a scalar random variable but the analysis can be extended to the case of more than one regressor. We assume that x_{it} is strictly exogenous. Let $\theta_i = \beta_i / (1 - \lambda_i)$ be the long-run coefficient of x_{it} for the i^{th} group and rewrite (28.46) as

$$\Delta y_{it} = \alpha_i - (1 - \lambda_i)\left(y_{i,t-1} - \theta_i x_{it}\right) + u_{it},$$

or

$$\Delta y_{it} = \alpha_i - \phi_i \left(y_{i,t-1} - \theta_i x_{it}\right) + u_{it}.$$

Consider now the random coefficient model

$$\phi_i = \phi + \eta_{i1}, \tag{28.47}$$
$$\theta_i = \theta + \eta_{i2}. \tag{28.48}$$

Hence

$$\beta_i = \theta_i \phi_i = \theta\phi + \eta_{i3}, \tag{28.49}$$

where

$$\eta_{i3} = \phi\eta_{i2} + \theta\eta_{i1} + \eta_{i1}\eta_{i2}, \tag{28.50}$$
$$\begin{pmatrix} \eta_{i1} \\ \eta_{i2} \end{pmatrix} \sim IID\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix}\right],$$

and

$$\omega_{33} = Var\left(\eta_{i3}\right) = Var(\phi\eta_{i2} + \theta\eta_{i1} + \eta_{i1}\eta_{i2}).$$

Letting $\lambda = 1 - \phi$ and $\beta = \theta\phi$, and using the above in (28.46) we have

$$y_{it} = \alpha_i + \lambda y_{i,t-1} + \beta x_{it} + v_{it}, \tag{28.51}$$

$$v_{it} = u_{it} - \eta_{i1} y_{i,t-1} + \eta_{i3} x_{it}. \tag{28.52}$$

It is now clear that v_{it} and $y_{i,t-1}$ are correlated and the FE or RE estimators will not be consistent. This is not a surprising result in the case where T is small. In Chapter 27 we saw that the FE (and RE) estimators are inconsistent when T is finite and N large when the slopes λ_i and β_i are homogeneous, that is, $\eta_{i1} = \eta_{i3} = 0$. The significant result here is that the inconsistency of the FE and RE estimators *will not disappear* even when both $T \to \infty$ and $N \to \infty$, if the slopes λ_i and/or β_i are heterogenous across groups. In fact, in the relatively simple case where

$$\lambda_i = \lambda, \qquad \left(\text{or } \eta_{i1} = 0\right),$$
$$\beta_i = \beta + \eta_{i3},$$

namely only the coefficients of x_{it} vary across groups, and

$$x_{it} = \mu_i \left(1 - \rho\right) + \rho x_{i,t-1} + \nu_{it},$$
$$|\rho| < 1, \quad E\left(x_{it}\right) = \mu_i,$$
$$\nu_{it} \sim IID\left(0, \tau^2\right), \tag{28.53}$$

we have[5]

$$\underset{N,T\to\infty}{Plim}\left(\hat{\lambda}_{FE}\right) - \lambda = \frac{\rho\left(1 - \lambda\rho\right)\left(1 - \lambda^2\right)\omega_{33}}{\Psi_1}, \tag{28.54}$$

$$\underset{N,T\to\infty}{Plim}\left(\hat{\beta}_{FE}\right) - \beta = -\frac{\beta\rho^2\left(1 - \lambda^2\right)\omega_{33}}{\Psi_1},$$

where

$$\Psi_1 = \left(\frac{\sigma^2}{\tau^2}\right)\left(1 - \rho^2\right)\left(1 - \lambda\rho\right)^2 + \left(1 - \lambda^2\rho^2\right)\omega_{33} + \left(1 - \rho^2\right)\beta^2 > 0,$$

and $\omega_{33} = Var\left(\eta_{i3}\right) = Var\left(\beta_i\right)$ measures the degree of heterogeneity in β_i. It is now clear that when $\rho > 0$,

$$Plim\left(\hat{\lambda}_{FE}\right) > \lambda, \ Plim\left(\hat{\beta}_{FE}\right) < \beta.$$

[5] It is interesting that when $\rho > 0$ the heterogeneity bias, given by (28.54), is in the opposite direction to the Nickell bias defined by (27.3).

The bias of the FE estimator of the long-run coefficient, $\hat{\theta}_{FE} = \hat{\beta}_{FE}/\left(1 - \hat{\lambda}_{FE}\right)$, is given by

$$\underset{N,T\to\infty}{Plim}\left(\hat{\theta}_{FE}\right) = \frac{\theta}{1-\rho\Psi_2},$$

where

$$\Psi_2 = \frac{(1+\lambda)\,\omega_{33}}{(1+\rho)\left[\left(\frac{\sigma^2}{\tau^2}\right)(1-\lambda\rho)^2 + \left(\beta^2+\omega_{33}\right)\right]}.$$

Thus note that

$$Plim\left(\hat{\theta}_{FE}\right) > \theta, \quad \text{if} \quad \rho > 0.$$

In the case where x_{it} is trended or if $\rho \to 1$ from below we have

$$\underset{\rho\to 1}{Plim}\left(\hat{\lambda}_{FE}\right) = 1, \text{ and } \underset{\rho\to 1}{Plim}\left(\hat{\beta}_{FE}\right) = 0,$$

irrespective of the true value of λ. See Pesaran and Smith (1995) for further details.

Example 65 *The FE 'static' private savings regressions reported in Tables 28.1 and 28.2 within Example 63 are subject to a substantial degree of residual serial correlation, which can lead to inconsistent estimates even under slope homogeneity since the wealth variable, W, is in fact constructed from accumulation of past savings. The presence of residual serial correlation could be due to a host of factors: omitted variables, neglected slope heterogeneity in the case of serially correlated regressors, and of course neglected dynamics. The diagnostic statistics provided in the second part of Table 28.3, within Example 64, show statistically significant evidence of residual serial correlation in the case of eight of the twenty countries.[6] It is clear that, even when the slope coefficients are allowed to be estimated freely across countries, residual serial correlation still continues to be a problem, at least in the case of some, if not all, the countries.[7] The usual time series technique for dealing with dynamic misspecification is to estimate error correction models based on ARDL models. ARDL models have the advantage that they are robust to integration and cointegration properties of the regressors, and for sufficiently high lag-orders could be immune to the endogeneity problem, at least as far as the long-run properties of the model are concerned. In the present application, observations for each individual country are available for too short a period to estimate even a first-order ARDL model including all the 10 regressors for each country separately.[8] Pooling in the form of FE estimation can compensate for lack of time series observations but, as shown in previous example, this can have its own set of problems. To check the robustness of the 'static' FE estimates presented in Table 28.2 to dynamic misspecification, Haque, Pesaran, and Sharma (2000) estimated the following first-order dynamic panel data model*

6 The diagnostic statistics are computed using the Lagrange multiplier procedure described in Section 5.8, and are valid irrespective of whether the regressions contain lagged dependent variables, implicitly or explicitly.

7 Under slope homogeneity restrictions, residual serial correlation is a problem for *all* the countries in the panel.

8 A first-order *ARDL* model in the private savings rate for each country that contains *all* ten regressors would involve estimating twenty-two unknown parameters with only twenty-two time series observations available per country!

$$y_{it} = \mu_i + \lambda y_{i,t-1} + \boldsymbol{\beta}_0' \mathbf{x}_{it} + \boldsymbol{\beta}_{01}' (\mathbf{x}_{it} \bar{W}_i) + \boldsymbol{\beta}_1' \mathbf{x}_{i,t-1} + u_{it}. \tag{28.55}$$

The country-specific long-run coefficients are given by

$$\boldsymbol{\theta}_i = (\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 + \boldsymbol{\beta}_{01} \bar{W}_i)/(1 - \lambda). \tag{28.56}$$

The FE estimates computed using all the 21 countries over the period 1972–1993 are given in Table 28.4.[9] *Clearly, there are significant dynamics, particularly in the relationship between changes in the government surplus and expenditure variables (SUR, GCUR, and GI) and the private savings rate. There is also important evidence of cross-sectional variations in the coefficients of wealth, income and demographic variables (W, YRUS and DEP). However, unlike the static estimates in Table 28.2, the coefficients of GDP growth and the real interest rate are no longer statistically significant. Overall, this equation presents a substantial improvement over the static FE estimates. In fact, the estimated standard error of this dynamic regression is 62 percent lower than the standard error of the FE estimates favoured by Masson, Bayoumi, and Samiei (1998), and reproduced in the first column of Table 28.1. Using the formula (28.56) the following estimates of the long-run coefficients are obtained*

SUR	−0.432	
	(−3.11)	
GCUR	−0.398	
	(−4.65)	
GI	−0.202	
	(−0.91)	
GR	−0.004	
	(−0.03)	
RINT	0.154	
	(1.64)	
W	0.224	$-0.00057\bar{W}_i$
	(4.58)	(−3.77)
INF	0.248	
	(3.10)	
PCTT	0.136	
	(4.11)	
YRUS	1.384	$-0.0047\bar{W}_i$
	(2.58)	(−2.92)
DEP	0.708	$-0.0027\bar{W}_i$
	(2.19)	(−2.64)

According to these estimates the long-run coefficients of the SUR and GCUR variables are still statistically significant, although the coefficient of the SUR variable is now estimated to be much lower than the estimate based on the static regressions. The long-run coefficients of the GI, GR and RINT variables are no longer statistically significant. It appears that, in contrast to government consumption expenditures, the effect of changes in government investment expenditures on private savings is temporary and tends to zero in the long run. The inflation and the terms of trade variables (INF

[9] For relatively simple dynamic models where T (= 22) is reasonably large and of the same order of magnitude as N (= 21), the application of the IV type estimators, discussed in Chapter 27, to a first differenced version of (28.55) does not seem necessary and can lead to considerable loss of efficiency.

Table 28.4 Fixed-effects estimates of dynamic private savings equations with cross-sectionally varying slopes (21 OECD countries, 1972–1993)

Regressors	Coefficients	Regressors	Coefficients
$PSAV_{-1}$	0.670 (20.80)	W	0.074 (4.41)
SUR	−0.771 (−16.28)	$W \times \overline{W}_i$	−0.00019 (−3.62)
SUR_{-1}	0.628 (11.54)	INE	0.082 (3.11)
GCUR	−0.544 (7.78)	PCTT	0.045 (4.54)
$GCUR_{-1}$	0.412 (6.16)	YRUS	0.456 (2.49)
GI	−0.666 (−5.54)	$YRUS \times \overline{W}_i$	−0.00157 (−2.81)
GL_1	0.600 (4.80)	DEP	0.233 (2.12)
GR	−0.0014 (−0.03)	$DEP \times \overline{W}_i$	−0.00089 (−2.52)
RINT	0.051 (1.60)		
$\bar{R}^2$	0.908		
$\bar{\sigma}$	1.451		
LL	−807.61		
AIC	−845.61		
SBC	924.18		

*The figures in brackets are t-ratios.

and PCTT) have the expected signs and are also statistically significant. The long-run coefficients of the remaining variables vary with country-specific average wealth-GDP ratio and when averaged across countries yield the values of 0.043 [0.026], −0.118 [0.219] *and* −0.148 [0.125] *for W, YRUS, and DEP variables respectively. The cross-sectional standard errors of these estimates are given in square brackets. The average estimate of the coefficient of the relative income variable has the wrong sign, but it is not statistically significant. The average estimates of the other two coefficients have the expected signs, but are not statistically significant either. It seems that the effects of many of the regressors considered in the MBS study are not robust to dynamic misspecifications. However, it would be interesting to examine the consequences of jointly allowing for unrestricted short-run slope heterogeneity and dynamics.*

28.8 Mean group estimator of dynamic heterogeneous panels

Consider a dynamic model of the form

$$y_{it} = \lambda_i y_{i,t-1} + \mathbf{x}'_{it}\boldsymbol{\beta}_i + u_{it},\ i = 1, 2, \ldots, N;\ \ t = 1, 2, \ldots, T, \tag{28.57}$$

where $\mathbf{x}_{it}$ is a $k \times 1$ vector of exogenous variables, and the error term u_{it} is assumed to be independently, identically distributed over t with mean zero and variance σ_i^2, and is independent across i. Let $\psi_i = (\lambda_i, \beta_i')'$. Further assume that ψ_i is independently distributed across i with

$$E(\psi_i) = \psi = (\lambda, \beta')', \tag{28.58}$$

$$E[(\psi_i - \psi)(\psi_i - \psi)'] = \Delta. \tag{28.59}$$

Rewriting $\psi_i = \psi + \eta_i$, (28.58) and (28.59) can be equivalently written as

$$E(\eta_i) = \mathbf{0}, \quad E\left(\eta_i \eta_j'\right) = \begin{cases} \Delta & \text{if} \quad i = j, \\ \mathbf{0} & \text{if} \quad i \neq j. \end{cases} \tag{28.60}$$

Although we may maintain the Assumption (28.7) that $E\left(\eta_i \mathbf{x}_{it}'\right) = \mathbf{0}$, we can no longer assume that $E\left(\eta_i y_{i,t-1}\right) = \mathbf{0}$. Through continuous substitutions, we have

$$y_{i,t-1} = \sum_{j=0}^{\infty} (\lambda + \eta_{i1})^j \mathbf{x}_{i,t-j-1}'(\beta + \eta_{i2}) + \sum_{j=0}^{\infty} (\lambda + \eta_{i1})^j u_{i,t-j-1}, \tag{28.61}$$

where $\eta_i = \left(\eta_{i1}, \eta_{i2}'\right)'$. It follows that $E(\eta_i y_{i,t-1}) \neq \mathbf{0}$.

The violation of the independence between the regressors and the individual effects, η_i, implies that the pooled least squares regression of y_{it} on $y_{i,t-1}$, and $\mathbf{x}_{it}$ will yield inconsistent estimates of ψ, even for sufficiently large T and N. Pesaran and Smith (1995) have noted that, as $T \to \infty$, the least squares regression of y_{it} on $y_{i,t-1}$ and $\mathbf{x}_{it}$ yields a consistent estimator of ψ_i, $\hat{\psi}_i$. Hence, the authors suggest a MG estimator of ψ by taking the average of $\hat{\psi}_i$ across i,

$$\hat{\psi}_{MG} = \frac{1}{N} \sum_{i=1}^{N} \hat{\psi}_i, \tag{28.62}$$

where

$$\hat{\psi}_i = \left(\mathbf{W}_{i.}'\mathbf{W}_{i.}\right)^{-1} \mathbf{W}_{i.}'\mathbf{y}_{i.},$$

$\mathbf{W}_{i.} = (\mathbf{y}_{i.,-1}, \mathbf{X}_{i.})$ with $\mathbf{y}_{i.,-1} = (y_{i0}, y_{i1}, \ldots, y_{iT-1})'$. The variance of $\hat{\psi}_{MG}$ is consistently estimated by

$$\widehat{Var}\left(\hat{\psi}_{MG}\right) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \left(\hat{\psi}_i - \hat{\psi}_{MG}\right)\left(\hat{\psi}_i - \hat{\psi}_{MG}\right)'.$$

Note that, for finite T, $\hat{\psi}_i$ for ψ_i is biased, with a bias of order $1/T$ (Hurwicz (1950), Kiviet and Phillips (1993)). Hsiao, Pesaran, and Tahmiscioglu (1999) have shown that the MG estimator is asymptotically normal for large N, and large T, so long as $\sqrt{N}/T \to 0$ as both N and $T \to \infty$.

28.8.1 Small sample bias

The MG estimator in the case of dynamic panels is biased when T is small, due to the presence of the lagged dependent variable in the model which biases the *OLS* estimator of the short-run coefficients λ_i and $\boldsymbol{\beta}_i$. Pesaran, Smith, and Im (1996) investigate the small sample properties of various estimators of the long-run coefficients for a dynamic heterogeneous panel data model. They find that when T is small the MG estimator can be seriously biased, particularly when N is large relative to T. In particular, for finite T, as $N \to \infty$ (under the usual panel assumption of independence across groups), the MG estimator still converges to a normal distribution, but with a mean which is not the same as the true value of the parameter under consideration, if the underlying equations contain lagged dependent variables or weakly exogenous regressors. To see this, first note that, for a finite T,

$$E\left(\hat{\boldsymbol{\psi}}_{MG}\right) = \boldsymbol{\psi} + \frac{1}{N}\sum_{i=1}^{N} E\left[\left(\mathbf{W}_{i.}'\mathbf{W}_{i.}\right)^{-1}\mathbf{W}_{i.}'\mathbf{u}_{i.}\right]. \tag{28.63}$$

It is easy to see that, due to the presence of lagged dependent variables, $N \to \infty$ is not sufficient for eliminating the second term. One needs large enough T for the bias to disappear. In practice, when the model contains lagged dependent variables, we have

$$E\left[\left(\mathbf{W}_{i.}'\mathbf{W}_{i.}\right)^{-1}\mathbf{W}_{i.}'\mathbf{u}_{i.}\right] = \frac{\mathbf{K}_{iT}}{T} + O\left(T^{-\frac{3}{2}}\right),$$

where $\mathbf{K}_{iT}$ is bounded in T and a function of the unknown underlying parameters. Hence

$$E\left(\hat{\boldsymbol{\psi}}_{MG}\right) = \boldsymbol{\psi} + \frac{1}{T}\sum_{i=1}^{N}\frac{\mathbf{K}_{iT}}{N} + O\left(T^{-\frac{3}{2}}\right).$$

Pesaran and Zhao (1999) propose a number of bias reduction techniques for the MG estimator of the long-run coefficients in dynamic models. Estimation of such coefficients poses additional difficulties due to the nonlinearity of long-run coefficients in terms of the underlying short-run parameters is an additional source of bias for the MG estimation of dynamic models. In a set of Monte Carlo experiments, Hsiao, Pesaran, and Tahmiscioglu (1999) showed that the MG estimator is unlikely to be a good estimator when either N or T is small.

28.9 Bayesian approach

Under the assumption that y_{i0} are fixed and known and η_i and u_{it} are independently normally distributed, we can implement the Bayes estimator of $\boldsymbol{\psi}_i$ conditional on σ_i^2 and $\boldsymbol{\Delta}$, namely

$$\hat{\boldsymbol{\psi}}_B = \left\{\sum_{i=1}^{N}\left[\sigma_i^2(\mathbf{W}_i'\mathbf{W}_i)^{-1} + \boldsymbol{\Delta}\right]^{-1}\right\}^{-1}\sum_{i=1}^{N}\left[\sigma_i^2(\mathbf{W}_i'\mathbf{W}_i)^{-1} + \boldsymbol{\Delta}\right]\hat{\boldsymbol{\psi}}_i, \tag{28.64}$$

where $\mathbf{W}_i = (\mathbf{y}_{i,-1}, \mathbf{X}_i)$ with $\mathbf{y}_{i,-1} = (y_{i0}, y_{i1}, \ldots, y_{iT-1})'$. This Bayes estimator is a weighted average of the least squares estimator of individual units with the weights being inversely proportional to individual variances. When $T \to \infty, N \to \infty$, and $N/T^{3/2} \to 0$, the Bayes estimator is asymptotically equivalent to the MG estimator (28.62) (Hsiao, Pesaran, and Tahmiscioglu (1999)).

In practice, the variance components, σ_i^2 and Δ are rarely known. The Monte Carlo studies conducted by Hsiao, Pesaran, and Tahmiscioglu (1999) show that, following the approach of Lindley and Smith (1972) in assuming that the prior-distributions of σ_i^2 and Δ are independent and are distributed as

$$P(\Delta^{-1}, \sigma_1^2, \ldots, \sigma_n^2) = W(\Delta^{-1} | (rR)^{-1}, r) \prod_{i=1}^{N} \sigma_i^{-2}, \tag{28.65}$$

yields a Bayes estimator almost as good as the Bayes estimator with known Δ and σ_i^2, where $W(.)$ represents the Wishart distribution with scale matrix, rR, and degrees of freedom r.

The Hsiao, Pesaran, and Tahmiscioglu (1999) Bayes estimator is derived under the assumption that the initial observation y_{i0} are fixed constants. As discussed in Anderson and Hsiao (1981, 1982), this assumption is clearly unjustifiable for a panel with finite T. However, contrary to the sampling approach where the correct modelling of initial observations is quite important, the Hsiao, Pesaran, and Tahmiscioglu (1999) Bayesian approach appears to perform fairly well in the estimation of the mean coefficients for dynamic random coefficient models as demonstrated in their Monte Carlo studies.

28.10 Pooled mean group estimator

Consider the *ARDL* model (28.44). Pesaran, Shin, and Smith (1999) has proposed an estimation method for *ARDL* models, under the assumption that the long-run coefficients on $\mathbf{X}_i$, defined by $\boldsymbol{\theta}_i = -\boldsymbol{\beta}_i/\phi_i$, are the same across the groups, namely

$$\boldsymbol{\theta}_i = \boldsymbol{\theta}, \quad i = 1, 2, \ldots, N.$$

This estimator, known as the pooled mean group estimator, provides a useful intermediate alternative between estimating separate regressions, which allows all coefficients and error variances to differ across the groups, and standard FE estimators that assume the slope coefficients are the same across i. Under the above assumptions, the error correction model can be written more compactly as

$$\Delta \mathbf{y}_i = \phi_i \boldsymbol{\xi}_i(\boldsymbol{\theta}) + \mathbf{W}_i \boldsymbol{\kappa}_i + \boldsymbol{\varepsilon}_i, \tag{28.66}$$

where

$$\mathbf{W}_i = (\Delta \mathbf{y}_{i,-1}, \Delta \mathbf{y}_{i,-2}, \ldots, \Delta \mathbf{y}_{i,-p+1}, \Delta \mathbf{X}_i, \Delta \mathbf{X}_{i,-1}, \ldots, \Delta \mathbf{X}_{i,-q+1}),$$
$$\boldsymbol{\xi}_i(\boldsymbol{\theta}) = \mathbf{y}_{i,-1} - \mathbf{X}_i \boldsymbol{\theta},$$

is the error correction component, and

$$\kappa_i = (\lambda_{i1}^*, \lambda_{i2}^*, \ldots, \lambda_{i,p-1}^*; \delta_{i0}^{*\prime}, \delta_{i1}^{*\prime}, \ldots, \delta_{i,q-1}^{*\prime})'.$$

There are three issues to be noted in estimating (28.66). First, the regression equations for each group are nonlinear in ϕ_i and $\boldsymbol{\theta}$. A further complication arises from the cross-equation parameter restrictions existing by virtue of the long-run homogeneity assumption. Finally, note that the error variances differ across groups. The log-likelihood function is

$$\ell_T(\varphi) = -\frac{T}{2}\sum_{i=1}^{N} \ln 2\pi\sigma_i^2 - \frac{1}{2}\sum_{i=1}^{N} \sigma_i^{-2} Q_i, \tag{28.67}$$

where

$$Q_i = [\Delta \mathbf{y}_i - \phi_i \xi_i(\boldsymbol{\theta})]' \mathbf{H}_i [\Delta \mathbf{y}_i - \phi_i \xi_i(\boldsymbol{\theta})],$$
$$\mathbf{H}_i = \mathbf{I}_T - \mathbf{W}_i(\mathbf{W}_i'\mathbf{W}_i)^{-1}\mathbf{W}_i',$$

$\mathbf{I}_T$ is an identity matrix of order T, $\boldsymbol{\varphi} = (\boldsymbol{\theta}', \boldsymbol{\phi}', \boldsymbol{\sigma}')'$, $\boldsymbol{\phi} = (\varphi_1, \varphi_2, \ldots, \varphi_N)'$, and $\boldsymbol{\sigma} = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)'$. In the case where the $\mathbf{x}_{it}^{\prime s}$ are $I(0)$, the pooled observation matrix on the regressors

$$\frac{1}{NT}\sum_{i=1}^{N} \frac{\phi_i^2}{\sigma_i^2} \mathbf{X}_i'\mathbf{H}_i\mathbf{X}_i,$$

converges in probability to a fixed positive definite matrix. In the case where the $\mathbf{x}_{it}$'s are $I(1)$, the matrix

$$\frac{1}{NT^2}\sum_{i=1}^{N} \frac{\phi_i^2}{\sigma_i^2} \mathbf{X}_i'\mathbf{H}_i\mathbf{X}_i,$$

converges to a random positive definite matrix with probability 1. These conditions should hold for all feasible values of ϕ_i and σ_i^2 as $T \to \infty$ either for a fixed N, or for $N \to \infty$ and $T \to \infty$, jointly. See Pesaran, Shin, and Smith (1999) for details.

The *ML* estimates of the long-run coefficients, $\boldsymbol{\theta}$, and the group-specific error-correction coefficients, ϕ_i, can be computed by maximizing (28.67) with respect to φ. These *ML* estimators are termed pooled mean group (PMG) estimators in order to highlight the pooling effect of the homogeneity restrictions on the estimates of the long-run coefficients, and the fact that averages across groups are used to obtain group-wide mean estimates of the error-correction coefficients and the other short-run parameters of the model.

Pesaran, Shin, and Smith (1999) propose two different likelihood-based algorithms for the computation of the PMG estimators which are computationally less demanding than estimating the pooled regression. The first is a 'back-substitution' algorithm that only makes use of the first derivatives of the log-likelihood function

$$\hat{\theta} = -\left(\sum_{i=1}^{N} \frac{\hat{\phi}_i^2}{\hat{\sigma}_i^2} \mathbf{X}_i' \mathbf{H}_i \mathbf{X}_i\right)^{-1} \left[\sum_{i=1}^{N} \frac{\hat{\phi}_i}{\hat{\sigma}_i^2} \mathbf{X}_i' \mathbf{H}_i \left(\Delta \mathbf{y}_i - \hat{\phi}_i \mathbf{y}_{i,-1}\right)\right], \tag{28.68}$$

$$\hat{\phi}_i = \left(\hat{\xi}_i' \mathbf{H}_i \hat{\xi}_i\right)^{-1} \hat{\xi}_i' \mathbf{H}_i \Delta \mathbf{y}_i, \tag{28.69}$$

$$\hat{\sigma}_i^2 = T^{-1} (\Delta \mathbf{y}_i - \hat{\phi}_i \hat{\xi}_i)' \mathbf{H}_i (\Delta \mathbf{y}_i - \hat{\phi}_i \hat{\xi}_i), \tag{28.70}$$

where $\hat{\xi}_i = \mathbf{y}_{i,-1} - \mathbf{X}_i \hat{\theta}$. Starting with an initial estimate of θ, say $\hat{\theta}^{(0)}$, estimates of ϕ_i and σ_i^2 can be computed using (28.69) and (28.70), which can then be substituted in (28.68) to obtain a new estimate of θ, say $\hat{\theta}^{(1)}$, and so on until convergence is achieved. Alternatively, the PMG estimators can be computed using (a variation of) the Newton-Raphson algorithm which makes use of both the first and the second derivatives. An overview of alternative numerical optimization techniques is provided in Section A.16 of Appendix A.

Note that, for small T, the PMG estimator (as well as the group-specific estimator) will be subject to the familiar downward bias on the coefficient of the lagged dependent variable. Because the bias is in the same direction for each group, averaging or pooling does not reduce this bias. Bias corrections are available in the literature (e.g., Kiviet and Phillips (1993)), but these apply to the short-run coefficients. Because the long-run coefficient is a nonlinear function of the short-run coefficients, procedures that remove the bias in the short-run coefficients can leave the long-run coefficient biased. Pesaran and Zhao (1999) discuss how the bias in the long-run coefficients can be reduced.

Example 66 *Continuing from Example 65, Haque, Pesaran, and Sharma (2000) then allowed for both unrestricted short-run slope heterogeneity and dynamics. To this end, they estimate individual country regressions containing first-order lagged values of the savings rates, $PSAV_{i,t-1}$. The MG and pooled mean group (PMG) estimates of the long-run coefficients based on these dynamic individual country regressions are given in Table 28.5. For ease of comparison, the MG estimator based on a static version of these regressions, as well as the corresponding FE estimates, are reported. Unlike the FE estimates, the consequences of allowing for dynamics on the MG estimates are rather limited. Once again only the coefficients of the SUR and the GCUR variables are statistically significant, although the dynamic MG estimates suggest the coefficient of the PCTT variable to be also marginally significant. Finally, the last column of Table 28.5 provides the pooled mean group estimates of the long-run coefficients, where the short-run dynamics are allowed to differ freely across countries but equality restrictions are imposed on one or more of the long-run coefficients; the rationale being that due to differences in factors such as adjustment costs or the institutional set-up across countries slope homogeneity is more likely to be valid in the long run. The PMG estimates in Table 28.5 impose the slope homogeneity restrictions only on the long-run coefficients of the SUR variable. As expected, the PMG estimates are generally more precisely estimated and confirm that, amongst the various determinants of private savings considered by MBS, only the effects of the SUR and the GCUR variables seem to be reasonably robust to the presence of slope heterogeneity and yield plausible estimates for the offsetting effects of government budget surpluses and government consumption expenditures on private savings across OECD countries.*

Table 28.5 Private saving equations: fixed-effects, mean group and pooled MG estimates (20 OECD countries, 1972–1993)

	FE Estimates		Mean Group Estimates		Pooled MGE
Regressors	Static	Dynamic	Static	Dynamic	Dynamic
SUR	−0.518	−0.968	−0.671	−0.911	−0.870
	(−8.50)	(−7.76)	(−8.07)	(−5.48)	(−19.81)
GCUR	−0.461	−0.665	−0.401	−0.394	−0.474
	(−10.76)	(−8.17)	(−5.95)	(−4.38)	(−6.88)
GI	−0.555	−0.789	−0.335	−0.109	−0.401
	(−5.28)	(−4.14)	(−1.01)	(−0.22)	(−1.14)
GR	−0.059	0.091	0.046	0.057	0.029
	(−1.09)	(−0.93)	(0.88)	(0.92)	(0.48)
RINT	0.205	0.127	0.104	0.183	0.139
	(4.11)	(1.41)	(1.28)	(1.61)	(1.66)
W	0.020	0.028	0.001	0.002	−0.004
	(4.51)	(3.49)	(0.061)	(0.115)	(−0.21)
INF	0.161	0.069	0.089	0.137	0.103
	(3.91)	(0.93)	(1.02)	(1.18)	(1.11)
PCTT	0.044	0.094	0.048	0.103	0.077
	(2.83)	(3.31)	(1.34)	(2.21)	(2.37)
YP	−0.087	−0.076	−0.069	−0.056	−0.031
	(−2.54)	(−1.23)	(−0.77)	(−0.60)	(−0.35)
DEP	−0.161	−0.241	0.080	0.058	0.050
	(−5.13)	(−4.22)	(0.63)	(0.45)	(0.39)

*The dependent variable is PSAV_{it}. The estimates refer to the long-run coefficients. Dynamic fixed-effects (FE) estimates are based on a first-order autoregressive panel data model containing the lagged dependent variables, $\text{PSAV}_{i,t-1}$. The dynamic Mean Group (MG) estimates are based on country-specific regressions also containing. $\text{PSAV}_{i,t-1}$. The Pooled MG estimates impose the restrictions that the long-run coefficients of the SUR variable is the same across countries, but are otherwise comparable to the dynamic MG estimates. Due to the presence of YRUS variable in the model, country-specific parameters for the U.S. are not identifies, and the U.S. is dropped from the panel.

28.11 Testing for slope homogeneity

Given the adverse statistical consequences of neglected slope heterogeneity, it is important that the assumption of slope homogeneity is tested. To this end, consider the panel data

$$y_{it} = \alpha_i + \boldsymbol{\beta}_i' \mathbf{x}_{it} + u_{it}, \tag{28.71}$$

where α_i are bounded on a compact set, $\mathbf{x}_{it}$ is a k-dimensional vector of regressors, $\boldsymbol{\beta}_i$ is a k-dimensional vector of unknown slope coefficients, and $u_{it} \sim IID(0, \sigma_i^2)$. The null hypothesis of interest is

$$H_0 : \boldsymbol{\beta}_i = \boldsymbol{\beta}, \text{ for all } i, \ \|\boldsymbol{\beta}\| < K < \infty, \tag{28.72}$$

against the alternatives

$$H_1 : \boldsymbol{\beta}_i = \boldsymbol{\beta}, \text{ for a non-zero fraction of slopes.}$$

One assumption underlying existing tests for slope homogeneity is that, under H_1, the fraction of the slopes that are not the same does not tend to zero as $N \to \infty$.

28.11.1 Standard F-test

There are a number of procedures that can be used to test H_0, the most familiar of which is the standard F-test defined by

$$F = \left(\frac{N(T-k-1)}{k(N-1)}\right)\frac{RSSR - USSR}{USSR},$$

where $RSSR$ and $USSR$ are restricted and unrestricted residual sum of squares, respectively, obtained under the null ($\boldsymbol{\beta}_i = \boldsymbol{\beta}$) and the alternative hypotheses. This test is applicable when N is fixed as $T \to \infty$, and the error variances are homoskedastic, $\sigma_i^2 = \sigma^2$. But it is likely to perform rather poorly in cases where N is relatively large, the regressors contain lagged values of the dependent variable and/or if the error variances are cross sectionally heteroskedastic.

28.11.2 Hausman-type test by panels

For cases where $N > T$, Pesaran, Smith, and Im (1996) propose using the Hausman (1978) procedure by comparing the fixed-effects (FE) estimator of $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}}_{FE} = \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_\tau\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_\tau\mathbf{y}_i, \tag{28.73}$$

with the mean group (MG) estimator

$$\hat{\boldsymbol{\beta}}_{MG} = N^{-1}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i, \tag{28.74}$$

where $\mathbf{M}_\tau = \mathbf{I}_T - \boldsymbol{\tau}_T\left(\boldsymbol{\tau}_T'\boldsymbol{\tau}_T\right)^{-1}\boldsymbol{\tau}_T'$, $\boldsymbol{\tau}_T$ is a $\mathbf{T}\times 1$ vector of ones, $\mathbf{I}_T$ is an identity matrix of order T, and

$$\hat{\boldsymbol{\beta}}_i = \left(\mathbf{X}_i'\mathbf{M}_\tau\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{M}_\tau\mathbf{y}_i. \tag{28.75}$$

For the Hausman test to have the correct size and be consistent two conditions must be met (see also Section 26.9.1)

(a) Under H_0, $\hat{\boldsymbol{\beta}}_{FE}$ and $\hat{\boldsymbol{\beta}}_{MG}$ must both be consistent for $\boldsymbol{\beta}$, with $\hat{\boldsymbol{\beta}}_{FE}$ being asymptotically more efficient such that

$$AVar\left(\hat{\boldsymbol{\beta}}_{MG} - \hat{\boldsymbol{\beta}}_{FE}\right) = AVar\left(\hat{\boldsymbol{\beta}}_{MG}\right) - AVar\left(\hat{\boldsymbol{\beta}}_{FE}\right) > 0, \tag{28.76}$$

where $AVar(\cdot)$ stands for the asymptotic variance operator.

(b) Under H_1, $\hat{\boldsymbol{\beta}}_{MG} - \hat{\boldsymbol{\beta}}_{FE}$ should tend to a non-zero vector.

In the context of dynamic panel data models with exogenous regressors both of these conditions are met, so long as the exogenous regressors are not drawn from the same distribution. In such a case a Hausman-type test based on the difference $\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{MG}$ would be valid and is shown to have reasonable small sample properties. See Pesaran, Smith, and Im (1996) and Hsiao and Pesaran (2008).

However, as is well known, the Hausman procedure can lack power for certain parameter values as its implicit null does not necessarily coincide with the null hypothesis of interest. This problem turns out to be much more serious in the application of the Hausman procedure to the testing problem that concerns us here. For example, in the case of panel data models containing only strictly exogenous regressors, a test of slope homogeneity based on $\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{MG}$ will lack power in *all* directions, if under the alternative hypothesis, the slopes are random draws from the same distribution. To see this, suppose that under H_1 the slopes satisfy the familiar random coefficient specification

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{v}_i, \mathbf{v}_i \sim IID(\mathbf{0}, \boldsymbol{\Sigma}_v),$$

where $\boldsymbol{\Sigma}_v \neq \mathbf{0}$ is a non-negative definite matrix, and $E(\mathbf{X}_j'\mathbf{v}_i) = \mathbf{0}$ for all i and j. Then

$$\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{MG} = \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_\tau\mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \left(\mathbf{X}_i'\mathbf{M}_\tau\mathbf{X}_i\right)\mathbf{v}_i - N^{-1}\sum_{i=1}^{N}\mathbf{v}_i + \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_\tau\mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{M}_\tau\boldsymbol{\varepsilon}_i - N^{-1}\sum_{i=1}^{N}\left(\mathbf{X}_i'\mathbf{M}_\tau\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{M}_\tau\boldsymbol{\varepsilon}_i,$$

and it readily follows that, under the random coefficients alternatives and strictly exogenous regressors, we have $E(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{MG} | H_1) = \mathbf{0}$. This result holds for N and T fixed as well as when N and $T \to \infty$, and hence condition (b) of Hausman's procedure is not satisfied.

Another important case where the Hausman test does not apply arises when testing the homogeneity of slopes in pure autoregressive panel data models. To simplify the exposition, consider the following stationary AR(1) panel data model

$$y_{it} = \alpha_i(1 - \beta_i) + \beta_i y_{i,t-1} + \varepsilon_{it}, \text{ with } |\beta_i| < 1. \tag{28.77}$$

It is now easily seen that with N fixed and as $T \to \infty$, under H_0 (where $\beta_i = \beta$) we have

$$\sqrt{NT}\left(\hat{\beta}_{FE} - \beta\right) \to_d N\left(0, 1 - \beta^2\right),$$

and

$$\sqrt{NT}\left(\hat{\beta}_{MG} - \beta\right) \to_d N\left(0, 1 - \beta^2\right).$$

Hence the variance inequality part of condition (a), namely (28.76), is not satisfied, and the application of the Hausman test to autoregressive panels will not have the correct size.

28.11.3 *G*-test of Phillips and Sul

Phillips and Sul (2003) propose a different type of Hausman test where, instead of comparing two different pooled estimators of the regression coefficients (as discussed in Section 28.11.2), they propose basing the test of slope homogeneity on the difference between the individual estimates and a suitably defined pooled estimator. In the context of the panel regression model (28.71), their test statistic can be written as

$$G = \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\tau}_N \otimes \hat{\boldsymbol{\beta}}_{FE}\right)' \hat{\boldsymbol{\Sigma}}_g^{-1} \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\tau}_N \otimes \hat{\boldsymbol{\beta}}_{FE}\right),$$

where $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1', \hat{\boldsymbol{\beta}}_2', \ldots, \hat{\boldsymbol{\beta}}_N')'$ is an $Nk \times 1$ stacked vector of all the N individual least square estimates of $\boldsymbol{\beta}_i$, $\hat{\boldsymbol{\beta}}_{FE}$ is a fixed-effect estimator as before, and $\hat{\boldsymbol{\Sigma}}_g$ is a consistent estimator of $\boldsymbol{\Sigma}_g$, the asymptotic variance matrix of $\hat{\boldsymbol{\beta}} - \boldsymbol{\tau}_N \otimes \hat{\boldsymbol{\beta}}_{FE}$, under H_0. Under standard assumptions for stationary dynamic models, and assuming H_0 holds and N is fixed, then $G \to_d \chi^2(Nk)$ as $T \to \infty$, so long as $\boldsymbol{\Sigma}_g$ is a non-stochastic positive definite matrix.

As compared to the Hausman test based on $\hat{\boldsymbol{\beta}}_{MG} - \hat{\boldsymbol{\beta}}_{FE}$, the G test is likely to be more powerful; but its use will be limited to panel data models where N is small relative to T. Also, the G test will not be valid in the case of pure dynamic models, very much for the same kind of reasons noted above in relation to the Hausman test based on $\hat{\boldsymbol{\beta}}_{MG} - \hat{\boldsymbol{\beta}}_{FE}$. This is easily established in the case of the stationary first-order autoregressive panel data model considered by Phillips and Sul (2003). In the case of $AR(1)$ panel regressions with $\sigma_i^2 = \sigma^2$, it is easily verified that under H_0

$$\begin{aligned} Avar\left[\sqrt{T}\left(\hat{\beta}_i - \hat{\beta}_{FE}\right)\right] &= Avar\left[\sqrt{T}\left(\hat{\beta}_i - \beta\right) - \sqrt{T}\left(\hat{\beta}_{FE} - \beta\right)\right] \\ &= \left(1 - \beta^2\right) - \left(\frac{1-\beta^2}{N}\right), \\ Acov\left[\sqrt{T}\left(\hat{\beta}_i - \hat{\beta}_{FE}\right), \sqrt{T}\left(\hat{\beta}_j - \hat{\beta}_{FE}\right)\right] &= -\left(\frac{1-\beta^2}{N}\right). \end{aligned}$$

Therefore

$$\boldsymbol{\Sigma}_g = \left(\frac{1-\beta^2}{T}\right)\left(\mathbf{I}_N - N^{-1}\boldsymbol{\tau}_N\boldsymbol{\tau}_N'\right).$$

It is now easily seen that $rank\left(\boldsymbol{\Sigma}_g\right) = N - 1$, and $\boldsymbol{\Sigma}_g$ is non-invertible.

28.11.4 Swamy's test

Swamy (1970) proposes a test of slope homogeneity based on the dispersion of individual slope estimates from a suitable pooled estimator. Like the F-test, Swamy's test is developed for panels where N is small relative to T, but allows for cross-sectional heteroskedasticity. Swamy's statistic applied to the slope coefficients can be written as

$$\hat{S} = \sum_{i=1}^{N} \left(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_{WFE}\right)' \frac{\mathbf{X}_i' \mathbf{M}_\tau \mathbf{X}_i}{\hat{\sigma}_i^2} \left(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_{WFE}\right), \tag{28.78}$$

where $\hat{\sigma}_i^2$ is an estimator of σ_i^2 based on $\hat{\boldsymbol{\beta}}_{WFE}$, namely

$$\hat{\sigma}_i^2 = \frac{1}{T-k-1} \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_{WFE}\right)' \mathbf{M}_\tau \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_{WFE}\right),$$

and $\hat{\boldsymbol{\beta}}_{WFE}$ is the weighted pooled estimator also computed using $\hat{\sigma}_i^2$, namely

$$\hat{\boldsymbol{\beta}}_{WFE} = \left(\sum_{i=1}^{N} \frac{\mathbf{X}_i' \mathbf{M}_\tau \mathbf{X}_i}{\hat{\sigma}_i^2}\right)^{-1} \sum_{i=1}^{N} \frac{\mathbf{X}_i' \mathbf{M}_\tau \mathbf{y}_i}{\hat{\sigma}_i^2}.$$

In the case where N is fixed and T tends to infinity, under H_0 the Swamy statistic, $\hat{S}$, is asymptotically chi-square-distributed with $k(N-1)$ degrees of freedom.

28.11.5 Pesaran and Yamagata Δ-test

Based on Swamy (1970)'s work, Pesaran and Yamagata (2008) propose a standardized dispersion statistic that is asymptotically normally distributed for large N and T. One version of the dispersion test, denoted by $\hat{\Delta}$, makes use of the Swamy statistic, $\hat{S}$ defined by (28.78), and another version, denoted by $\tilde{\Delta}$, is based on a modified version of the Swamy statistic where regression standard errors for the individual cross-sectional units are computed using the pooled fixed-effects, rather than the ordinary least squares estimator, as proposed by Swamy. It is shown that, in the case of models with strictly exogenous regressors, but with non-normal errors, both versions of the Δ-test tend to the standard normal distribution as $(N, T) \to_j \infty$, subject to certain restrictions on the relative expansion rates of N and T. For the $\hat{\Delta}$-test it is required that $\sqrt{N}/T \to 0$, as $(N, T) \to_j \infty$, whilst for the $\tilde{\Delta}$-test the condition is less restrictive and is given by $\sqrt{N}/T^2 \to 0$. When the errors are normally distributed, mean-variance bias adjusted versions of the Δ-tests, denoted by $\hat{\Delta}_{adj}$ and $\tilde{\Delta}_{adj}$, are proposed that are valid as $(N, T) \to_j \infty$ without any restrictions on the relative expansion rates of N and T.

More specifically, $\hat{\Delta}$ and $\tilde{\Delta}$-tests are defined by

$$\hat{\Delta} = \sqrt{N}\left(\frac{N^{-1}\hat{S} - k}{\sqrt{2k}}\right), \quad \tilde{\Delta} = \sqrt{N}\left(\frac{N^{-1}\tilde{S} - k}{\sqrt{2k}}\right) \tag{28.79}$$

where

$$\tilde{S} = \sum_{i=1}^{N} \left(\hat{\beta}_i - \tilde{\beta}_{WFE}\right)' \frac{\mathbf{X}_i'\mathbf{M}_\tau\mathbf{X}_i}{\tilde{\sigma}_i^2} \left(\hat{\beta}_i - \tilde{\beta}_{WFE}\right), \tag{28.80}$$

$$\tilde{\beta}_{WFE} = \left(\sum_{i=1}^{N} \frac{\mathbf{X}_i'\mathbf{M}_\tau\mathbf{X}_i}{\tilde{\sigma}_i^2}\right)^{-1} \sum_{i=1}^{N} \frac{\mathbf{X}_i'\mathbf{M}_\tau\mathbf{y}_i}{\tilde{\sigma}_i^2},$$

and

$$\tilde{\sigma}_i^2 = \frac{1}{T-1} \left(\mathbf{y}_i - \mathbf{X}_i\hat{\beta}_{FE}\right)' \mathbf{M}_\tau \left(\mathbf{y}_i - \mathbf{X}_i\hat{\beta}_{FE}\right).$$

Although the difference between $\hat{S}$ and $\tilde{S}$ might appear slight at first, the different choices of the estimator of σ_i^2 used in construction of these statistics have important implications for the properties of the two tests as N and T tends to infinity. To see this let

$$\mathbf{Q}_{iT} = T^{-1}\left(\mathbf{X}_i'\mathbf{M}_\tau\mathbf{X}_i\right), \tag{28.81}$$

$$\mathbf{Q}_{NT} = (NT)^{-1}\left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_\tau\mathbf{X}_i\right), \tag{28.82}$$

$$\mathbf{P}_i = \mathbf{M}_\tau\mathbf{X}_i\left(\mathbf{X}_i'\mathbf{M}_\tau\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{M}_\tau, \tag{28.83}$$

$$\mathbf{M}_i = \mathbf{I}_T - \mathbf{Z}_i(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i', \tag{28.84}$$

where $\mathbf{Z}_i = (\boldsymbol{\tau}_T, \mathbf{X}_i)$, and consider the following assumptions:

Assumption H.5:

(i) $\varepsilon_{it}|\mathbf{X}_i \sim IID(0, \sigma_i^2)$, $\sigma_{\max}^2 = \max_{1\le i\le N}(\sigma_i^2) < K$, and $\sigma_{\min}^2 = \min_{1\le i\le N}(\sigma_i^2) > 0$.

(ii) ε_{it} and ε_{js} are independently distributed for $i \neq j$ and/or $t \neq s$.

(iii) $E(\varepsilon_{it}^9|\mathbf{X}_i) < K$.

Assumption H.6: (i) The $k \times k$ matrices $\mathbf{Q}_{iT}$, $i = 1, 2, \ldots, N$, defined by (28.81) are positive definite and bounded, $\max_{1\le i\le N} E\left\|\mathbf{Q}_{iT}\right\| < K$, and $\mathbf{Q}_{iT}$ tends to a non-stochastic positive definite matrix, $\mathbf{Q}_i$, $\max_{1\le i\le N} E\left\|\mathbf{Q}_i\right\| < K$, as $T \to \infty$.

(ii) The $k \times k$ pooled observation matrix $\mathbf{Q}_{NT}$ defined by (28.82) is positive definite, and $\mathbf{Q}_{NT}$ tends to a non-stochastic positive definite matrix, $\mathbf{Q} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\mathbf{Q}_i$, as $(N, T) \xrightarrow{j} \infty$.

Assumption H.7: There exists a finite T_0 such that for $T > T_0$, $E\{[\boldsymbol{v}_i'\mathbf{M}_\tau\boldsymbol{v}_i/(T-1)]^{-4-\epsilon}\} < K$ and $E\{[\boldsymbol{v}_i'\mathbf{M}_i\boldsymbol{v}_i/(T-k-1)]^{-4-\epsilon}\} < K$, for each i and for some small positive constant ϵ, where $\boldsymbol{v}_i = \boldsymbol{\varepsilon}_i/\sigma_i$.

Assumption H.8: Under H_1, the fraction of slopes that are not the same does not tend to zero as $N \to \infty$.

Under Assumptions H.5–H.7 and assuming that H_0 (the null of slope homogeneity) holds, then the dispersion statistics $\hat{S}$ and $\tilde{S}$ defined above can be written as

$$N^{-1/2}\hat{S} = N^{-1/2}\sum_{i=1}^{N} \hat{z}_{iT} + O_p\left(N^{-1/2}\right) + O_p\left(T^{-1/2}\right), \tag{28.85}$$

$$N^{-1/2}\tilde{S} = N^{-1/2}\sum_{i=1}^{N} \tilde{z}_{iT} + O_p\left(N^{-1/2}\right) + O_p\left(T^{-1/2}\right), \tag{28.86}$$

where

$$\hat{z}_{iT} = \frac{(T-k-1)\boldsymbol{v}_i'\mathbf{P}_i\boldsymbol{v}_i}{\boldsymbol{v}_i'\mathbf{M}_i\boldsymbol{v}_i}, \text{ and } \tilde{z}_{iT} = \frac{(T-1)\boldsymbol{v}_i'\mathbf{P}_i\boldsymbol{v}_i}{\boldsymbol{v}_i'\mathbf{M}_\tau\boldsymbol{v}_i}. \tag{28.87}$$

Under Assumptions H.4–H.7, $\hat{z}_{iT}$ and $\tilde{z}_{iT}$ are independently (but not necessarily identically) distributed random variables across i with finite means and variances, and for all i we have

$$E(\hat{z}_{iT}) = k + O(T^{-1}),\; Var(\hat{z}_{iT}) = 2k + O(T^{-1}), \tag{28.88}$$

$$E(\tilde{z}_{iT}) = k + O(T^{-2}),\; Var(\tilde{z}_{iT}) = 2k + O(T^{-1}), \tag{28.89}$$

$$E\left|\hat{z}_{iT}\right|^{2+\epsilon/2} < K, \text{ and } E\left|\tilde{z}_{iT}\right|^{2+\epsilon/2} < K. \tag{28.90}$$

Also under the null hypothesis that the slopes are homogenous, we have

$$\hat{\Delta} \to_d N(0,1), \text{ as } (N,T) \xrightarrow{j} \infty, \text{ so long as } \sqrt{N}/T \to 0,$$

$$\tilde{\Delta} \to_d N(0,1), \text{ as } (N,T) \xrightarrow{j} \infty, \text{ so long as } \sqrt{N}/T^2 \to 0,$$

where the standardized dispersion statistics, $\hat{\Delta}$ and $\tilde{\Delta}$ are defined above. Furthermore, if the errors, ε_{it}, are normally distributed, under H_0 we have

$$\hat{\Delta} \to_d N(0,1), \text{ as } (N,T) \xrightarrow{j} \infty, \text{ so long as } \sqrt{N}/T \to 0,$$

$$\tilde{\Delta} \to_d N(0,1), \text{ as } (N,T) \xrightarrow{j} \infty.$$

The small sample properties of the dispersion tests can be improved under the normally distributed errors by considering the following mean and variance bias adjusted versions of $\hat{\Delta}$ and $\tilde{\Delta}$

$$\tilde{\Delta}_{adj} = \sqrt{\frac{N(T+1)}{(T-k-1)}}\left(\frac{N^{-1}\tilde{S} - k}{\sqrt{2k}}\right), \tag{28.91}$$

$$\hat{\Delta}_{adj} = \sqrt{N}\left(\frac{N^{-1}\hat{S} - E(\hat{z}_{iT})}{\sqrt{Var(\hat{z}_{iT})}}\right),$$

where

$$E(\hat{z}_{iT}) = \frac{k(T-k-1)}{T-k-3},\ Var(\hat{z}_{iT}) = \frac{2k\,(T-k-1)^2\,(T-3)}{(T-k-3)^2\,(T-k-5)}. \tag{28.92}$$

The Monte Carlo results reported in Pesaran and Yamagata (2008) suggest that the $\tilde{\Delta}_{adj}$ test works well even if there are major departures from normality, and is to be recommended.

28.11.6 Extensions of the Δ-tests

The Δ-tests can be readily extended to test the homogeneity of a subset of slope coefficients. Consider the following partitioned form of (28.71)

$$\underset{T\times 1}{\mathbf{y}_i} = \alpha_i \boldsymbol{\tau}_T + \underset{T\times k_1}{\mathbf{X}_{i1}}\,\boldsymbol{\beta}_{i1} + \underset{T\times k_2}{\mathbf{X}_{i2}}\,\boldsymbol{\beta}_{i2} + \boldsymbol{\varepsilon}_i,\ i = 1, 2, \ldots, N,$$

or

$$\underset{T\times 1}{\mathbf{y}_i} = \underset{T\times(k_1+1)}{\mathbf{Z}_{i1}}\ \boldsymbol{\delta}_i + \underset{T\times k_2}{\mathbf{X}_{i2}}\,\boldsymbol{\beta}_{i2} + \boldsymbol{\varepsilon}_i,$$

where $\mathbf{Z}_{i1} = (\boldsymbol{\tau}_T, \mathbf{X}_{i1})$ and $\boldsymbol{\delta}_i = \left(\alpha_i, \boldsymbol{\beta}'_{i1}\right)'$. Suppose the slope homogeneity hypothesis of interest is given by

$$H_0:\ \boldsymbol{\beta}_{i2} = \boldsymbol{\beta}_2,\ \text{for } i = 1, 2, \ldots, N. \tag{28.93}$$

The dispersion test statistic in this case is given by

$$\tilde{S}_2 = \sum_{i=1}^{N}\left(\hat{\boldsymbol{\beta}}_{i2} - \tilde{\boldsymbol{\beta}}_{2,WFE}\right)'\frac{\mathbf{X}'_{i2}\mathbf{M}_{i1}\mathbf{X}_{i2}}{\tilde{\sigma}_i^2}\left(\hat{\boldsymbol{\beta}}_{i2} - \tilde{\boldsymbol{\beta}}_{2,WFE}\right),$$

where

$$\hat{\boldsymbol{\beta}}_{i2} = \left(\mathbf{X}'_{i2}\mathbf{M}_{i1}\mathbf{X}_{i2}\right)^{-1}\mathbf{X}'_{i2}\mathbf{M}_{i1}\mathbf{y}_i,$$

$$\tilde{\boldsymbol{\beta}}_{2,WFE} = \left(\sum_{i=1}^{N}\frac{\mathbf{X}'_{i2}\mathbf{M}_{i1}\mathbf{X}_{i2}}{\tilde{\sigma}_i^2}\right)^{-1}\sum_{i=1}^{N}\frac{\mathbf{X}'_{i2}\mathbf{M}_{i1}\mathbf{y}_i}{\tilde{\sigma}_i^2},$$

$$\mathbf{M}_{i1} = \mathbf{I}_T - \mathbf{Z}_{i1}\left(\mathbf{Z}'_{i1}\mathbf{Z}_{i1}\right)^{-1}\mathbf{Z}'_{i1},$$

$$\tilde{\sigma}_i^2 = \frac{\left(\mathbf{y}_i - \mathbf{X}_{i2}\hat{\boldsymbol{\beta}}_{2,FE}\right)'\mathbf{M}_{i1}\left(\mathbf{y}_i - \mathbf{X}_{i2}\hat{\boldsymbol{\beta}}_{2,FE}\right)}{T-k_1-1},$$

and

$$\hat{\boldsymbol{\beta}}_{2,FE} = \left(\sum_{i=1}^{N} \mathbf{X}_{i2}'\mathbf{M}_{i1}\mathbf{X}_{i2}\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_{i2}'\mathbf{M}_{i1}\mathbf{y}_i.$$

Using a similar line of reasoning as above, it is now easily seen that under H_0 defined by (28.93), and for $(N, T) \xrightarrow{j} \infty$, such that $\sqrt{N}/T^2 \to 0$, then

$$\tilde{\Delta}_2 = \sqrt{N}\left(\frac{N^{-1}\tilde{S}_2 - k_2}{\sqrt{2k_2}}\right) \to_d N(0, 1).$$

In the case of normally distributed errors, the following mean-variance bias adjusted statistic can be used

$$\tilde{\Delta}_{adj} = \sqrt{\frac{N(T - k_1 + 1)}{(T - k - 1)}}\left(\frac{N^{-1}\tilde{S}_2 - k_2}{\sqrt{2k_2}}\right).$$

The Δ-tests can also be extended to unbalanced panels. Denoting the number of time series observations on the i^{th} cross-section by T_i, the standardized dispersion statistic is given by

$$\tilde{\Delta} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\tilde{d}_i - k}{\sqrt{2k}}\right), \tag{28.94}$$

$$\tilde{d}_i = \left(\hat{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\beta}}_{WFE}\right)' \frac{\mathbf{X}_i'\mathbf{M}_{\tau_i}\mathbf{X}_i}{\tilde{\sigma}_i^2}\left(\hat{\boldsymbol{\beta}}_i - \tilde{\boldsymbol{\beta}}_{WFE}\right),$$

$\mathbf{X}_i = \left(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT_i}\right)'$, $\mathbf{M}_{\tau_i} = \mathbf{I}_{T_i} - \boldsymbol{\tau}_{T_i}\left(\boldsymbol{\tau}_{T_i}'\boldsymbol{\tau}_{T_i}\right)^{-1}\boldsymbol{\tau}_{T_i}'$ with $\boldsymbol{\tau}_{T_i}$ being a $T_i \times 1$ vector of unity,

$$\hat{\boldsymbol{\beta}}_i = \left(\mathbf{X}_i'\mathbf{M}_{\tau_i}\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{M}_{\tau_i}\mathbf{y}_i, \tag{28.95}$$

$$\tilde{\boldsymbol{\beta}}_{WFE} = \left(\sum_{i=1}^{N}\frac{\mathbf{X}_i'\mathbf{M}_{\tau_i}\mathbf{X}_i}{\tilde{\sigma}_i^2}\right)^{-1}\sum_{i=1}^{N}\frac{\mathbf{X}_i'\mathbf{M}_{\tau_i}\mathbf{y}_i}{\tilde{\sigma}_i^2}, \tag{28.96}$$

$\mathbf{y}_i = \left(y_{i1}, y_{i2}, \ldots, y_{iT_i}\right)'$,

$$\tilde{\sigma}_i^2 = \frac{\left(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_{FE}\right)'\mathbf{M}_{\tau_i}\left(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_{FE}\right)}{T_i - 1},$$

and

$$\hat{\boldsymbol{\beta}}_{FE} = \left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\tau_i}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\tau_i}\mathbf{y}_i. \tag{28.97}$$

The $\tilde{\Delta}$-test can also be applied to stationary dynamic models. Pesaran and Yamagata (2008) show that the test will be valid for dynamic panel data models so long as $N/T \rightarrow \kappa$, as $(N, T) \rightarrow_j \infty$, where $0 \leq \kappa < \infty$. This condition is more restrictive than the one obtained for panels with exogenous regressors, but is the same as the condition required for the validity of the fixed-effects estimator of the slope in $AR(1)$ models in large N and T panels.

Using Monte Carlo experiments it is shown that the $\tilde{\Delta}$-test has the correct size and satisfactory power in panels with strictly exogenous regressors for various combinations of N and T. Similar results are also obtained for dynamic panels, but only if the autoregressive coefficient is not too close to unity and so long as $T \geq N$. See Pesaran and Yamagata (2008) for further discussion.

28.11.7 Bias-corrected bootstrap tests of slope homogeneity for the *AR*(1) model

One possible way of improving on the asymptotic test developed for the *AR* models would be to follow the recent literature and use bootstrap techniques.[10] Here we make use of a bias-corrected version of the recursive bootstrap procedure.[11]

One of the main problems in the application of bootstrap techniques to dynamic models in small T samples is the fact that the *OLS* estimates of the individual coefficients, λ_i, or their *FE* (or *WFE*) counterparts are biased when T is small; a bias that persists with $N \rightarrow \infty$. To deal with this problem we focus on the $AR(1)$ case and use the bias-corrected version of $\tilde{\lambda}_{WFE}$ as proposed by Hahn and Kuersteiner (2002).[12] Denoting the bias-corrected version of $\tilde{\lambda}_{WFE}$ by $\circ$, we have

$$\mathring{\lambda}_{WFE} = \tilde{\lambda}_{WFE} + \frac{1}{T}\left(1 + \tilde{\lambda}_{WFE}\right), \tag{28.98}$$

and estimate the associated intercepts as

$$\mathring{\alpha}_{i,WFE} = \bar{y}_i - \mathring{\lambda}_{WFE}\,\bar{y}_{i,-1},$$

where $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$, and $\bar{y}_{i,-1} = T^{-1}\sum_{t=1}^{T} y_{i,t-1}$. The residuals are given by

$$\mathring{e}_{it} = y_{it} - \mathring{\alpha}_{i,WFE} - \mathring{\lambda}_{WFE}\, y_{i,t-1},$$

with the associated bias-corrected estimator of σ_i^2 given by $\mathring{\sigma}_i^2 = (T-1)^{-1}\sum_{t=1}^{T}(\mathring{e}_{it})^2$. The b^{th} bootstrap sample, $y_{it}^{(b)}$ for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$ can now be generated as

$$y_{it}^{(b)} = \mathring{\alpha}_{i,WFE} + \mathring{\lambda}_{WFE}\, y_{i,t-1}^{(b)} + \mathring{\sigma}_i \zeta_{it}^{(b)}, \text{ for } t = 1, 2, \ldots, T,$$

10 For example, see Beran (1988), Horowitz (1994), Li and Maddala (1996), and Bun (2004), although none of these authors makes any bias corrections in their bootstrapping procedures.

11 Bias-corrected estimates are also used in the literature on the derivation of the bootstrap confidence intervals to generate the bootstrap samples in dynamic $AR(p)$ models. See Kilian (1998), among others.

12 Bias corrections for the *OLS* estimates of individual λ_i are provided by Kendall (1954) and Marriott and Pope (1954), and further elaborated by Orcutt and Winokur (1969). See also Section 14.5. No bias corrections seem to be available for *FE* or *WFE* estimates of $AR(p)$ panel data models in the case of $p \geq 2$.

where $y_{i0}^{(b)} = y_{i0}$, and $\zeta_{it}^{(b)}$ are random draws with replacements from the set of pooled standardized residuals, $\mathring{e}_{it}/\mathring{\sigma}_i$, $i = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$. With $y_{it}^{(b)}$, for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$ the bootstrap statistics

$$\tilde{\Delta}^{(b)} = \sqrt{N}\left(\frac{N^{-1}\tilde{S}^{(b)} - 1}{\sqrt{2}}\right), \; b = 1, 2, \ldots, B,$$

where $\tilde{S}^{(b)}$ is the modified Swamy statistic, defined by (28.80), computed using the b^{th} bootstrapped sample. The statistics $\tilde{\Delta}^{(b)}$ for $b = 1, 2, \ldots, B$, can now be used to obtain the bootstrap p-values

$$p_B = \frac{1}{B}\sum_{b=1}^{B} I\left(\tilde{\Delta}^{(b)} - \tilde{\Delta}\right),$$

where B is the number of bootstrap sample, $I(A)$ takes the value of unity if $A > 0$ or zero otherwise, and $\tilde{\Delta}$ is the standardized dispersion statistic applied to the actual observations. If $p_B < 0.05$, say, the null hypothesis of slope homogeneity is rejected at the 5 per cent significance level.

28.11.8 Application: testing slope homogeneity in earnings dynamics

In this section we examine the slope homogeneity of the dynamic earnings equations with the panel study of income dynamics (*PSID*) data set used in Meghir and Pistaferri (2004). Briefly, these authors select male heads aged 25 to 55 with at least nine years of usable earnings data. The selection process leads to a sample of 2, 069 individuals and 31, 631 individual-year observations. To obtain a panel data set with a larger T, only individuals with at least 15 time series observations are included in the panel. This leaves us with 1, 031 individuals and 19, 992 individual-year observations. Following Meghir and Pistaferri (2004), the individuals are categorized into three education groups: High School Dropouts (HSD, those with less than 12 grades of schooling), High School Graduates (HSG, those with at least a high school diploma, but no college degree), and College Graduates (CLG, those with a college degree or more). In what follows, the earning equations for the different educational backgrounds, HSD, HSG, and CLG, are denoted by the superscripts $e = 1, 2$, and 3, and for the pooled sample by 0. The numbers of individuals in the three categories are $N^{(1)} = 249$, $N^{(2)} = 531$, and $N^{(3)} = 251$. The panel is unbalanced with $t = 1, \ldots T_i^{(e)}$ and $i = 1, \ldots, N^{(e)}$, and an average time period of around 18 years.

In the research on earnings dynamics, it is standard to adopt a two-step procedure where in the first stage the log of real earnings is regressed on a number of control variables such as age, race and year dummies. The dynamics are then modelled based on the residuals from this first stage regression. The use of the control variables and the grouping of the individuals by educational backgrounds is aimed at eliminating (minimizing) the effects of individual heterogeneities at the second stage.

It is, therefore, of interest to examine the extent to which the two-step strategy has been successful in dealing with the heterogeneity problem. With this in mind we follow closely the two-step procedure adopted by Meghir and Pistaferri (2004) and first run regressions of log real earnings, $w_{it}^{(e)}$, on the control variables: a square of "age" ($AGE_{it}^{(e)2}$), race ($WHITE_i^{(e)}$), year dummies ($YEAR(t)$), region of residence ($NE_{it}^{(e)}$, $CE_{it}^{(e)}$, $STH_{it}^{(e)}$), and residence in a standard metropolitan statistical area ($SMSA_{it}^{(e)}$), for each education group $e = 0, 1, 2, 3$, separately.[13] The residuals from these regressions, which we denote by $y_{it}^{(e)}$, are then used in the second stage to estimate dynamics of the earnings process.

Specifically,

$$y_{it}^{(e)} = \alpha_i^{(e)} + \lambda^{(e)} y_{it-1}^{(e)} + \sigma_i^{(e)} \varepsilon_{it}^{(e)}, \ e = 0, 1, 2, 3,$$

where within each education group $\lambda^{(e)}$ is assumed to be homogeneous across the different individuals. Our interest is to test the hypothesis that $\lambda^{(e)} = \lambda_i^{(e)}$ for all i in e.

The test results are given in the first panel of Table 28.6. The $\tilde{\Delta}$ statistics and the associated bootstrapped p values by education groups all lead to strong rejections of the homogeneity hypothesis. Judging by the size of the $\tilde{\Delta}$ statistics, the rejection is stronger for the pooled sample as compared with the sub-samples, confirming the importance of education as a discriminatory factor in the characterizations of heterogeneity of earnings dynamics across individuals. The test results also indicate the possibility of other statistically significant sources of heterogeneity within each of the education groups, and casts some doubt on the two-step estimation procedure adopted in the literature for dealing with heterogeneity, a point recently emphasized by Browning, Ejrnæs, and Alvarez (2010).

In Table 28.6 we also provide a number of different FE estimates of $\lambda^{(e)}$, $e = 0, 1, 2, 3$, on the assumption of within group slope homogeneity. Given the relatively small number of time series observations available (on average 18), the bias corrections to the FE estimates are quite large. The cross-section error variance heterogeneity also plays an important role in this application, as can be seen from a comparison of FE and WFE estimates with the latter being larger. Focusing on the bias-corrected WFE estimates, we also observe that the persistence of earnings dynamics rises systematically from 0.52 in the case of the school drop outs to 0.72 for the college graduates. This seems sensible, and partly reflects the more reliable job prospects that are usually open to individuals with a higher level of education.

The homogeneity test results suggest that further efforts are needed also to take account of within group heterogeneity. One possibility would be to adopt a Bayesian approach, assuming that $\lambda_i^{(e)}, i = 1, 2, \ldots, N^{(e)}$ are draws from a common probability distribution and focus attention on the whole posterior density function of the persistent coefficients, rather than the average estimates that tend to divert attention from the heterogeneity problem. Another possibility would be to follow Browning, Ejrnæs, and Alvarez (2010) and consider particular parametric functions, relating $\lambda_i^{(e)}$ to individual characteristics as a way of capturing within group heterogeneity. Finally, one could consider a finer categorization of the individuals in the panel; say by further splitting of the education groups or by introducing new categories such as occupational classifications. The slope homogeneity tests provide an indication of the statistical importance of the heterogeneity problem, but are silent as how best to deal with the problem.

[13] Log real earnings are computed as $w_{it}^{(e)} = \ln\left(LABY_{it}^{(e)}/PCED_t\right)$, where $LABY_{it}^{(e)}$ is earnings in the current US dollar, and $PCED_t$ is the personal consumption expenditure deflator, base year 1992.

Table 28.6 Slope homogeneity tests for the $AR(1)$ model of the real earnings equations

		Pooled Sample $e = 0$	High School dropout $e = 1$	High school graduate $e = 2$	College graduate $e = 3$
N		1,031	249	531	251
Average T_i		18.39	18.36	18.22	18.79
Total observations		18,961	4,572	9,673	4,716
Tests for slope homogeneity					
$\tilde{\Delta}$ **test**	Statistic	25.59	7.20	13.65	18.32
	Normal approximation p-value	[0.0000]	[0.0000]	[0.0000]	[0.0000]
	Bias-corrected bootstrap p-value	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Autoregressive coefficient (λ)					
FE estimates ($\hat{\lambda}_{FE}$)		0.4841 (0.0065)	0.4056 (0.0147)	0.4497 (0.0095)	0.5538 (0.0106)
WFE estimates ($\tilde{\lambda}_{WFE}$)		0.5429 (0.0056)	0.4246 (0.0133)	0.5169 (0.0086)	0.6002 (0.0095)
Bias-corrected WFE ($\mathring{\lambda}_{WFE}$)		0.6504 (0.0055)	0.5188 (0.0126)	0.6192 (0.0080)	0.7214 (0.0101)

Notes: The *FE* estimator and the *WFE* estimator are defined by (28.97), and (28.96), respectively, and their associated standard errors (shown in round brackets) are based on $\widehat{Var}\left(\hat{\lambda}_{FE}\right) = \hat{\sigma}^2\left(\sum_{i=1}^{N} \mathbf{y}_{i,-1}' \mathbf{M}_{\tau_i} \mathbf{y}_{i,-1}\right)^{-1}$, where

$$\hat{\sigma}^2 = \left(\mathrm{T} - N - 1\right)^{-1} \sum_{i=1}^{N} \left(\mathbf{y}_i - \hat{\lambda}_{FE}\mathbf{y}_{i,-1}\right)' \mathbf{M}_{\tau_i} \left(\mathbf{y}_i - \hat{\lambda}_{FE}\mathbf{y}_{i,-1}\right),$$

$$\mathrm{T} = \sum_{i=1}^{N} \mathrm{T}_i, \text{ and } \widehat{Var}\left(\tilde{\lambda}_{WFE}\right) = \left(\sum_{i=1}^{N} \tilde{\sigma}_i^{-2} \mathbf{y}_{i,-1}' \mathbf{M}_{\tau_i} \mathbf{y}_{i,-1}\right)^{-1}.$$

Bias corrected estimates are based on $\mathring{\lambda}_{WFE} = \tilde{\lambda}_{WFE} + (\mathrm{T}/N)\left(1 + \tilde{\lambda}_{WFE}\right)$ and $\widehat{Var}\left(\mathring{\lambda}_{WFE}\right) = \mathrm{T}^{-1}\left(1 - \mathring{\lambda}_{WFE}^2\right)$. Bias-corrected bootstrapped tests also use $\mathring{\lambda}_{WFE}$ and the associated estimates to generate bootstrap samples (see Section 28.11.7 for further details).

28.12 Further reading

Further details on estimation and inference on large heterogeneous panels can be found in Hsiao (1975), Pesaran and Smith (1995), and Hsiao and Pesaran (2008).

28.13 Exercises

1. Suppose that

$$y_{it} = \boldsymbol{\beta}_{it}' \mathbf{x}_{it} + u_{it},\ i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T,$$