Chap.3 Covariance Structure and Robust Covariance Estimation

When assumption A1.1 is abandoned, what is a covariance structure in a panel data context? Two different types of correlation must be considered: correlation among cross-sections and serial correlation, as in traditional time series analysis. At the same time, two different types of heteroscedasticity also must be considered: cross-sectional heteroscedasticity and period heteroscedasticity. Moreover, fixed and random effects are two different alternative ways of considering covariance structures. They can not be combined.

3.1 Heteroscedasticity and Cross-Sectional Correlation in fixed effects models

1) General structure of disturbance Covariance Matrix

First, recall (1.55), the specification organized as a set of individual-specific equations:

$$Y = i_{nT}\alpha + X\beta + \left(I_n \otimes i_T\right)u + \left(i_n \otimes I_T\right)\gamma + \varepsilon$$

where the general form of the disturbance covariance matrix is given by

$$V = E\left(\varepsilon\varepsilon'\middle|X\right) = \begin{pmatrix} E\left(\varepsilon_1\varepsilon_1'\middle|X\right) & E\left(\varepsilon_1\varepsilon_2'\middle|X\right)\cdots E\left(\varepsilon_1\varepsilon_n'\middle|X\right) \\ E\left(\varepsilon_2\varepsilon_1'\middle|X\right) & E\left(\varepsilon_2\varepsilon_2'\middle|X\right)\cdots E\left(\varepsilon_2\varepsilon_n'\middle|X\right) \\ \cdots & \cdots \qquad\qquad \cdots \\ E\left(\varepsilon_n\varepsilon_1'\middle|X\right) & E\left(\varepsilon_n\varepsilon_2'\middle|X\right)\cdots E\left(\varepsilon_n\varepsilon_n'\middle|X\right) \end{pmatrix} \tag{3.1}$$

If instead we treat the specification as a set of period-specific equations, the stacked (by period) representation is given by

$$Y = i_{nT}\alpha + X\beta + \left(i_n \otimes I_T\right)u + \left(I_n \otimes i_T\right)\gamma + \varepsilon \tag{3.2}$$

and its disturbance matrix is given by

$$V = E\left(\varepsilon\varepsilon'\middle|X\right) = \begin{pmatrix} E\left(\varepsilon_1\varepsilon_1'\middle|X\right) & E\left(\varepsilon_1\varepsilon_2'\middle|X\right)\cdots E\left(\varepsilon_1\varepsilon_T'\middle|X\right) \\ E\left(\varepsilon_2\varepsilon_1'\middle|X\right) & E\left(\varepsilon_2\varepsilon_2'\middle|X\right)\cdots E\left(\varepsilon_2\varepsilon_T'\middle|X\right) \\ \cdots & \cdots \qquad\qquad \cdots \\ E\left(\varepsilon_T\varepsilon_1'\middle|X\right) & E\left(\varepsilon_T\varepsilon_2'\middle|X\right)\cdots E\left(\varepsilon_T\varepsilon_T'\middle|X\right) \end{pmatrix} \tag{3.3}$$

2) Cross-sectional heteroscedasticity

Cross-sectional heteroscedasticity allows for a different disturbance variance for each cross-section, constant over time with zero covariance, thus

$$E\left(\varepsilon_{it}^2\middle|X_i, u_i, \gamma\right) = \sigma_i^2 \quad \text{for all } t,\ i = 1, \cdots, n \tag{3.4}$$

$$E\left(\varepsilon_i\varepsilon_i'\middle|X_i, u_i, \gamma\right) = \sigma_i^2 I_T \tag{3.5}$$

and

$$V = diag\left(\sigma_1^2, \ldots, \sigma_n^2\right) \otimes I_T \tag{3.6}$$

FGLS for this specification is straightforward. First we perform preliminary estimation of the within transformed model obtain individual-specific residual:

$$\varepsilon_{it}^* = y_{it}^* - X_{it}^{*\prime} \hat{\beta}^{(w)} \tag{3.7}$$

Then we use these residuals to form estimates of the cross-sectional variances:

$$\hat{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_{it}^{*2} \qquad (i = 1, \cdots, n) \tag{3.8}$$

So that we have a weighted least squares procedure to form the FGLS estimates:

$$\hat{\hat{\beta}} = \left(X'\hat{V}^{-1}X\right)^{-1} X'\hat{V}^{-1}Y = \left(\sum_{i=1}^{n} \frac{1}{\hat{\sigma}_i^2} X_i' X_i\right)\left(\sum_{i=1}^{n} \frac{1}{\hat{\sigma}_i^2} X_i' Y_i\right) \tag{3.9}$$

3) Period heteroscedasticity

Here we assume that the variances are different from one period to another but constant over cross-sections for a given period. Again zero covariances are assumed between different cross sections. Therefore:

$$E\left(\varepsilon_{it}^2 \middle| X_t, u, \gamma_t\right) = \sigma_t^2 \text{ for all } i,\ t = 1, \cdots, T \tag{3.10}$$

$$E\left(\varepsilon_t \varepsilon_t' \middle| X_t, u, \gamma_t\right) = \sigma_t^2 I_n \tag{3.11}$$

Denote $\Lambda = diag\left(\sigma_1^2, \ldots, \sigma_T^2\right)$, we have

$$V = diag\left(\Lambda, \Lambda, \ldots, \Lambda\right) = I_n \otimes \Lambda \tag{3.12}$$

We perform preliminary estimation to obtain period-specific residual ε_{it}^*, then we use these residuals to form estimates of the period variances:

$$\hat{\sigma}_t^2 = \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_{it}^{*2} \qquad (t = 1, \cdots, T) \tag{3.13}$$

and form the FGLS estimates:

$$\hat{\hat{\beta}} = \left(X'\hat{V}^{-1}X\right)^{-1} X'\hat{V}^{-1}Y = \left(\sum_{i=1}^{n} X_i' \hat{\Lambda}^{-1} X_i\right)\left(\sum_{i=1}^{n} X_i' \hat{\Lambda}^{-1} Y_i\right) \tag{3.14}$$

4) Contemporaneous Correlation (Cross-Sectional Correlation)

In this case,

$$E\left(\varepsilon_{it} \varepsilon_{jt} \middle| X_t, u, \gamma_t\right) = \sigma_{ij} \tag{3.15}$$

$$E\left(\varepsilon_{is} \varepsilon_{jt} \middle| X_t, X_s, \gamma_t, \gamma_s, u\right) = 0 \tag{3.16}$$

for all i, j, s and t with $s \neq t$. Note that the contemporaneous covariances do not vary over t. Using the period-specific disturbance vectors, we may rewrite this assumption as:

$$E\left(\varepsilon_t\varepsilon_t'\middle|X_t,u,\gamma_t\right)=\Sigma_n \quad \text{for all} \quad t \tag{3.17}$$

where

$$\Sigma_n=\begin{pmatrix}\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n}\\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n}\\ \vdots & \vdots & \cdots & \vdots\\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn}\end{pmatrix} \tag{3.18}$$

and

$$V=\left(\sigma_{ij}I_T\right)=\Sigma_n\otimes I_T \tag{3.19}$$

Since it involves covariances across cross-sections as in a seemingly unrelated regressions type framework discussed in Chap.2, we term it a cross-section SUR specification, the FGLS estimation for this specification is similar to that for SUR model we discussed in Chap.2.

5) Period SUR (Period Heteroscedasticity and Serial Correlation)

In this case,

$$E\left(\varepsilon_{is}\varepsilon_{it}\middle|X_i,u_i,\gamma\right)=\sigma_{st} \tag{3.20}$$

$$E\left(\varepsilon_{is}\varepsilon_{jt}\middle|X_i,u_i,X_j,u_j,\gamma\right)=0 \tag{3.21}$$

for all i, j, s and t with $i\neq j$. Note that the heteroscedasticity and serial correlation do not vary cross-sections. Using the individual-specific disturbance vectors, we may rewrite this assumption as:

$$E\left(\varepsilon_i\varepsilon_i'\middle|X_i,u_i,\gamma\right)=\Sigma_T \quad \text{for all} \quad i \tag{3.22}$$

where

$$\Sigma_T=\begin{pmatrix}\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1T}\\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2T}\\ \vdots & \vdots & \cdots & \vdots\\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_{TT}\end{pmatrix} \tag{3.23}$$

and

$$V=I_n\otimes\Sigma_T \tag{3.24}$$

We term this a period SUR specification since it involves covariances across periods within a given cross-section, as in a seemingly unrelated regressions framework with period specific equations. In estimating this specification, we employ residuals obtained form first stage estimates to form an estimate of Σ_T. In the second stage, we perform FGLS,

$$\hat{\hat{\beta}}=\left(\sum_{i=1}^{n}X_i'\hat{\Sigma}_T^{-1}X_i\right)\left(\sum_{i=1}^{n}X_i'\hat{\Sigma}_T^{-1}Y_i\right) \tag{3.25}$$

3.2 Cross-sectional heteroscedasticity and correlation tests in fixed model

3.2.1 Testing for cross-sectional heteroscedasticity

For testing the cross-sectional heteroscedasticity assumption in (3.5), the full set of the test strategies that we have used before is available.

$H_0 : \sigma_1^2 = \cdots = \sigma_n^2$

i) LR Test

We now assume $\varepsilon_t \left| X \sim N\left(0, \Sigma\right)\right.$, where $\varepsilon_t = \left(\varepsilon_{t1}, \cdots, \varepsilon_{tn}\right)'$, $\Sigma = diag\left(\sigma_1^2, \cdots, \sigma_n^2\right)$. Taking logs and summing over the T periods gives the log-likelihood for the sample

$$\ln L = -\frac{nT}{2}\ln 2\pi - \frac{T}{2}\ln\left|\Sigma\right| - \frac{1}{2}\sum_{t=1}^{T}\varepsilon_t'\Sigma^{-1}\varepsilon_t \tag{3.26}$$

in which, $\varepsilon_{it} = y_{it} - X_{it}'\beta \qquad i = 1, \cdots, n$.

We can also carry out a likelihood ratio test discussed in cross-sectional heteroscedasticity (section 7.5 last term). Using $V = diag\left(\sigma_1^2, \cdots, \sigma_n^2\right) \otimes I_T$, the log-likelihood function is

$$\ln L = -\frac{nT}{2}\ln 2\pi - \frac{T}{2}\sum_{i=1}^{n}\ln\sigma_i^2 - \frac{1}{2}\sum_{i=1}^{n}\frac{\varepsilon_i'\varepsilon_i}{\sigma_i^2} \tag{3.27}$$

and its concentrated form is

$$\ln L_{1c} = -\frac{nT}{2}\ln 2\pi - \frac{T}{2}\sum_{i=1}^{n}\ln\sigma_i^2 - \frac{nT}{2}$$

Under H_0, its concentrated form is

$$\ln L_{10} = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma^2 - \frac{nT}{2}$$

So

$$LR = -2\left(\ln\hat{L}_{oc} - \ln\hat{L}_{1c}\right) = T\left(n\ln\hat{\sigma}^2 - \sum_{i=1}^{n}\ln\hat{\sigma}_i^2\right) \overset{a}{\sim} \chi^2\left(n-1\right) \tag{3.28}$$

where $\hat{\sigma}_i^2 = \frac{\hat{\varepsilon}_i'\hat{\varepsilon}_i}{T}$ and $\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{nT}$, with all residuals computed using ML estimators.

ii) LM Test

$$LM = g'\left[-E\left(H\right)\right]^{-1} g \tag{3.29}$$

$$g=\begin{bmatrix}\dfrac{\partial \ln L}{\partial \beta}\\ \dfrac{\partial \ln L}{\partial \sigma_i^2}\end{bmatrix}=\begin{bmatrix}\sum_{i=1}^{n}\dfrac{1}{\sigma_i^2}X_i'\varepsilon_i\\ -\dfrac{T}{2\sigma_i^2}+\dfrac{\varepsilon_i'\varepsilon_i}{2\sigma_i^4}, i=1,\cdots,n\end{bmatrix} \tag{3.30}$$

$$H=\left[\frac{\partial^2 \ln L}{\partial\theta\partial\theta'}\right] \quad \theta=\left(\beta,\sigma^2\right)'$$

$$=\begin{pmatrix}-\sum_{i=1}^{n}\dfrac{1}{\sigma_i^2}X_i'X_i & -\sum_{i=1}^{n}\dfrac{1}{\sigma_i^4}X_i'\varepsilon_i, i=1,\cdots,n\\ -\sum_{i=1}^{n}\dfrac{1}{\sigma_i^4}X_i'\varepsilon_i, i=1,\cdots,n & diag\left(\dfrac{T}{2\sigma_i^4}-\dfrac{\varepsilon_i'\varepsilon_i}{\sigma_i^6}, i=1,\cdots,n\right)\end{pmatrix} \tag{3.31}$$

$$-E\left(H\right)=\begin{pmatrix}\sum_{i=1}^{n}\dfrac{1}{\sigma_i^2}X_i'X_i & 0\\ 0 & diag\left(\dfrac{T}{2\sigma_i^4}, i=1,\cdots,n\right)\end{pmatrix} \tag{3.32}$$

Under the null hypothesis $\sigma_i^2=\sigma^2\left(i=1,\cdots,n\right)$, the first derivative of the log-likelihood function with respect to this common σ^2 is

$$\frac{\partial \ln L_R}{\partial \sigma^2}=-\frac{nT}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}\varepsilon_i'\varepsilon_i \tag{3.33}$$

Equating (3.33) to zero, we have

$$\sigma^2=\frac{1}{nT}\sum_{i=1}^{n}\varepsilon_i'\varepsilon_i == \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 \tag{3.34}$$

Under the null hypothesis of equal variances, regardless of what the common restricted estimator of σ_i^2 is, the first-order condition for equating $\partial \ln L/\partial\beta$ to zero will be the OLS normal equations. So we can use the LS residuals at the restricted solution to obtain $\hat{\sigma}^2=\dfrac{1}{nT}e'e$ and $\hat{\sigma}_i^2=\dfrac{1}{T}e_i'e_i$. With these results in hand and using the estimate of the $-E\left(H\right)$, the LM statistic reduces to

$$LM=\begin{bmatrix}-\dfrac{T}{2\hat{\sigma}^2}+\dfrac{e_1'e_1}{2\hat{\sigma}^4}\\ \vdots\\ -\dfrac{T}{2\hat{\sigma}^2}+\dfrac{e_n'e_n}{2\hat{\sigma}^4}\end{bmatrix}'\begin{pmatrix}\dfrac{2\hat{\sigma}^4}{T} & & 0\\ & \ddots & \\ 0 & & \dfrac{2\hat{\sigma}^4}{T}\end{pmatrix}\begin{bmatrix}-\dfrac{T}{2\hat{\sigma}^2}+\dfrac{e_1'e_1}{2\hat{\sigma}^4}\\ \vdots\\ -\dfrac{T}{2\hat{\sigma}^2}+\dfrac{e_n'e_n}{2\hat{\sigma}^4}\end{bmatrix}$$

$$= \begin{bmatrix} -\frac{T}{2\hat{\sigma}^2} + \frac{T\hat{\sigma}_1^2}{2\hat{\sigma}^4} \\ \vdots \\ -\frac{T}{2\hat{\sigma}^2} + \frac{T\hat{\sigma}_n^2}{2\hat{\sigma}^4} \end{bmatrix}' \begin{pmatrix} \frac{2\hat{\sigma}^4}{T} & & 0 \\ & \ddots & \\ 0 & & \frac{2\hat{\sigma}^4}{T} \end{pmatrix} \begin{bmatrix} -\frac{T}{2\hat{\sigma}^2} + \frac{T\hat{\sigma}_1^2}{2\hat{\sigma}^4} \\ \vdots \\ -\frac{T}{2\hat{\sigma}^2} + \frac{T\hat{\sigma}_n^2}{2\hat{\sigma}^4} \end{bmatrix}$$

$$= \sum_{i=1}^{n} \left[\frac{T}{2\hat{\sigma}^2} \left(\frac{\hat{\sigma}_i^2}{\hat{\sigma}^2} - 1 \right) \right]^2 \left(\frac{2\hat{\sigma}^2}{T} \right) = \frac{T}{2} \sum_{i=1}^{n} \left[\frac{\hat{\sigma}_i^2}{\hat{\sigma}^2} - 1 \right]^2 \sim \chi^2 (n-1) \tag{3.35}$$

Iii) Wald Test and Modified Wald Statistic

With the unrestricted estimates, we may use the Wald statistic. If we assume normality, then we have $A\text{var}\left(\hat{\sigma}_i^2\right) = \frac{2\sigma_i^4}{T}$ and the variances are asymptotically uncorrelated.

Therefore , using $\hat{\sigma}_i^2$ to estimate σ_i^2 produces the sample statistic

$$W = \sum_{i=1}^{n} \left(\hat{\sigma}_i^2 - \sigma^2 \right)^2 \left(\frac{2\hat{\sigma}_i^4}{T} \right)^{-1} \tag{3.36}$$

$$W = \frac{T}{2} \sum_{i=1}^{n} \left(\frac{\sigma^2}{\hat{\sigma}_i^2} - 1 \right)^2 \tag{3.37}$$

where the estimator of the common variance would be the pooled estimator from the first least squares.

If the assumption of normally distributed disturbances is inappropriate, then neither the LM nor the LR test is usable. Moreover the Wald statistic defined aboved is also incorrect. It remains possible to construct a usable Wald statistic as below.

Under the null hypothesis, the Wald statistic is

$$W = \sum_{i=1}^{n} \frac{\left(\hat{\sigma}_i^2 - \sigma^2 \right)^2}{Var\left(\hat{\sigma}_i^2 \right)} \tag{3.38}$$

If the null hypothesis is correct,

$$W \overset{a}{\sim} \chi^2 (n-1). \tag{3.39}$$

By hypothesis,

$$p\lim \hat{\sigma}^2 = \sigma^2 \tag{3.40}$$

where $\hat{\sigma}^2$ is the disturbance variance estimator from the pooled regression. Now we must reconsider $\text{var}\left(\hat{\sigma}_i^2\right)$, since

$$\hat{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^{T} e_{it}^2 \tag{3.41}$$

It is a mean of T observations, as such, we can use

$$V_i = \frac{1}{T}\frac{1}{T-1}\sum_{t=1}^{T}\left(e_{it}^2 - \hat{\sigma}_i^2\right)^2 \tag{3.42}$$

to estimate $\operatorname{var}\left(\hat{\sigma}_i^2\right)$. The modified Wald statistic is then

$$W = \sum_{i=1}^{n}\frac{\left(\hat{\sigma}_i^2 - \sigma^2\right)^2}{V_i} \overset{a}{\sim} \chi^2\left(n-1\right) \tag{3.43}$$

3.2.2 Testing for contemporaneous correlation

For testing the hypothesis that the off-diagonal elements of Σ are zero, there are three approaches. The likelihood ratio test is based on the statistic

$$LR = T\left(\ln\left|\hat{\Sigma}_{hetero.}\right| - \ln\left|\hat{\Sigma}_{general}\right|\right) = T\left(\sum_{i=1}^{n}\ln\hat{\sigma}_i^2 - \ln\left|\hat{\Sigma}\right|\right) \overset{a}{\sim} \chi^2_{\frac{n}{2}(n-1)} \tag{3.44}$$

where $\hat{\sigma}_i^2$ are the estimates of σ_i^2 obtained from MLE of the cross-sectional heterosecdastic model and $\hat{\Sigma}$ is the MLE in the unrestricted model. We can obtain the MLE from the iterating FGLS.

The LM statistic is

$$LM = T\sum_{i=2}^{n}\sum_{j=1}^{i-1} r_{ij}^2 \tag{3.45}$$

where r_{ij} is the ij-th residual correlation coefficient. Here the appropriate basis for computing the correlations is the residuals from the iterating FGLS estimator in the cross-sectional heteroscedastic model.

3.3 Fixed and random-effects models with an AR(1) disturbance

3.3.1 Balanced panels and AR(1) Process

1) Fixed-effects models

① Autocorrelation

The preceding discussion deal with heteroscedasticity and cross-sectional correlation. Now we relax the assumption of nonautocorrelation and assume that

$$E\left(\varepsilon_{it}\varepsilon'_{js} \middle| X_i, X_j, u_i, u_j\right) = 0 \quad \text{if } i \neq j \tag{3.46}$$

In addition, we allow for autocorrelation within the cross-sectional units, That is

$$\varepsilon_{it} = \rho_i \varepsilon_{it-1} + v_{it} \qquad |\rho_i| < 1 \tag{3.47}$$

where v_{it} is a white noise with variance σ^2_{vi}.

$$\operatorname{var}\left(\varepsilon_{it} \middle| X_i, u_i\right) = \sigma_i^2 = \frac{\sigma^2_{vi}}{1-\rho_i^2} \tag{3.48}$$

In such a case,

$$E\left(\varepsilon_i\varepsilon'_i \middle| X_i, u_i\right) = \sigma_i^2 \Omega_i \tag{3.49}$$

$$E\left(\varepsilon_i\varepsilon'_j \middle| X_i, X_j, u_i, u_j\right) = 0 \tag{3.50}$$

$$V = diag\left(\sigma_1^2\Omega_1, \cdots, \sigma_n^2\Omega_n\right) \tag{3.51}$$

For FGLS estimation of the model, suppose that $\hat{\rho}_i$ is a consistent estimator of ρ_i, then we can transformed the data using the Prais-Winsten transformation:

$$\begin{aligned} Y_{it}^* &= \sqrt{1-\hat{\rho}_i^2} \qquad t = 1 \\ &= Y_{it} - \hat{\rho}_i Y_{it-1} \quad t \geq 2 \end{aligned} \tag{3.52}$$

and similarly for each explanatory variable. The transformation has now removed the autocorrelation. As such, the cross-section heteroscedastic model applies to the transformed data, as described earlier, we have

$$\hat{\sigma}^2_{vi} = \frac{\hat{\varepsilon}_i^{*\prime}\hat{\varepsilon}_i^*}{T} \tag{3.53}$$

$$\hat{\sigma}_i^2 = \frac{\hat{\sigma}^2_{vi}}{1-\hat{\rho}_i^2}$$

where $\hat{\varepsilon}_i^*$ is the residual vector of unit i in this regression. Now the remaining question is how to obtain the estimates $\hat{\rho}_i$. The model is first estimated by the standard covariance method (within transformation), from the residuals e_i^*, a conventional estimator of ρ_i is given by

$$\hat{\rho}_i = \sum_{t=2}^{T} e_{it}^* e_{it-1}^* \Big/ \sum_{t=1}^{T} e_{it}^{*2} \tag{3.54}$$

If the disturbances have a common stochastic process with the same ρ_i, then several estimators of the common ρ are available. One of them is

$$\hat{\rho} = \sum_{i=1}^{n}\sum_{t=2}^{T} e_{it}^* e_{it-1}^* \Big/ \sum_{i=1}^{n}\sum_{t=2}^{T} e_{it}^{*2} \tag{3.55}$$

② **Autocorrelation and cross-sectional correlation**

According to (3.26), if we wish to allow for cross-sectional correlation across units, then the variance matrix in (3.1) would be $E\left(\varepsilon\varepsilon' \middle| X\right) = \left(\sigma_{ij}\Omega_{ij}\right)$.

We may further assume that (for convenience sake, the conditions are omitted) $Cov\left(v_{it}, v_{jt}\right) = \sigma_{v_{ij}}$, therefore we have

$$\begin{aligned} E\left(\varepsilon_{it}\varepsilon_{jt}\right) &= E\left[\left(\rho_i\varepsilon_{it-1} + v_{it}\right)\left(\rho_j\varepsilon_{jt-1} + v_{jt}\right)\right] \\ &= \rho_i\rho_j E\left(\varepsilon_{it-1}\varepsilon_{jt-1}\right) + E\left(v_{it}v_{jt}\right) \end{aligned} \tag{3.56}$$

We can get the diagonal elements of matrix $\sigma_{ij}\Omega_{ij}$, that is $\dfrac{\sigma_{v_{ij}}}{1-\rho_i\rho_j}$. At the same time, we have

$$\begin{aligned} E\left(\varepsilon_{it-1}\varepsilon_{jt}\right) &= E\left[\left(\rho_i\varepsilon_{it-2} + v_{it-1}\right)\left(\rho_j\varepsilon_{jt-1} + v_{jt}\right)\right] \\ &= \rho_i\rho_j E\left(\varepsilon_{it-2}\varepsilon_{jt-1}\right) + \rho_j E\left(v_{it-1}v_{jt-1}\right) \end{aligned} \tag{3.57}$$

And We can get the off-diagonal elements of matrix $\sigma_{ij}\Omega_{ij}$, that is $\rho_j \dfrac{\sigma_{v_{ij}}}{1-\rho_i\rho_j}$. On the analogy of this, we have

$$\sigma_{ij}\Omega_{ij} = \frac{\sigma_{v_{ij}}}{1-\rho_i\rho_j}\begin{pmatrix} 1 & \rho_j & \rho_j^2 & \cdots & \rho_j^{T-1} \\ \rho_i & 1 & \rho_j & \cdots & \rho_j^{T-2} \\ \vdots & & \cdots & & \\ \rho_i^{T-1} & \rho_i^{T-2} & \rho_i^{T-3} & \cdots & 1 \end{pmatrix} \tag{3.58}$$

The Parks method is FGLS for panel data models where the disturbances show heteroscedasticity、autocorrelation and cross-sectional correlation .Initial estimates of ρ_i

are required, as before. The Prais-Winsten transformation renders all the block in V diagonal. Therefore, the model of cross-sectional correlation in (3.24) applies to the transformed data. Estimates of $\sigma_{\varepsilon_{ij}}$ can be obtained from the least squares residual covariances obtained from the transformed data:

$$\hat{\sigma}_{\varepsilon_{ij}} = \frac{\hat{\sigma}_{v_{ij}}}{1-\hat{\rho}_i\hat{\rho}_j} \tag{3.59}$$

where , $\hat{\sigma}_{v_{ij}} = e'_{*i}e_{*j}/T$. e_{*i} and e_{*j} are respectively the least squares residual vectors of unit i and j .

The Parks method consists of two sequential FGLS transformations, first eliminating serial correlation of the disturbances then eliminating contemporaneous correlation of the disturbances. This is done by initially estimating within model by OLS. The residuals from this estimation are used to estimate the individual serial correlation of the disturbances, which are then used to transform the model into one with serially independent disturbances. Residuals from this estimation are then used to estimate the contemporaneous correlation of the disturbances, and the data is once again transformed to allow for OLS estimation with now spherical disturbances.

2) Random-effects models

The variance components model we discussed in Chap.1 assume that the only correlation over time is due to the presence in the panel of the same individual over several periods. This equicorrelation coefficient is given by $corr(\eta_{it},\eta_{is}) = \sigma_u^2 / (\sigma_u^2+\sigma_\varepsilon^2)$ for $t \neq s$. Note that it is the same no matter how far t is from s. This may be a restrictive assumption for economic relationships. Lillard andWillis (1978) generalized the error component model to the serially correlated case, by assuming that the remainder disturbances (the ε_{it}) in model (1.34) follow an $\mathrm{AR}(1)$ process:

$$Y_{it} = \delta + X'_{it}\beta + u_i + \varepsilon_{it} \quad ,\eta_{it} = u_i + \varepsilon_{it} \tag{3.60}$$

$$\varepsilon_{it} = \rho\varepsilon_{i,t-1} + v_{it} \tag{3.60'}$$

where $|\rho|<1$, $v_{it} \sim IID(0,\sigma_v^2)$; The $u_i \sim IID(0,\sigma_u^2)$, u_i are independent of the ε_{it} and $\varepsilon_{i0} \sim IID(0,\sigma_v^2/(1-\rho^2))$. Baltagi and Li (1991a) derived the corresponding Fuller and Battese (1974) transformation for this model. First, one applies the Prais–Winsten (PW) transformation matrix:

$$C=\begin{bmatrix}(1-\rho^2)^{1/2} & 0 & 0 & \cdots & 0 & 0 & 0\\ -\rho & 1 & 0 & \cdots & 0 & 0 & 0\\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & -\rho & 1 & 0\\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1\end{bmatrix}$$

to transform the remainder $\mathrm{AR}(1)$ disturbances into serially uncorrelated classical errors. For panel data, this has to be applied for n individuals. The transformed regression disturbances are in vector form

$$\eta^*=(I_n\otimes C)\eta=(I_n\otimes Ci_T)u+(I_n\otimes C)\varepsilon \qquad (3.61)$$

Using the fact that $Ci_T=(1-\rho)i_T^\alpha$, where $i_T^{\alpha'}=(\alpha,i'_{T-1})$ and $\alpha=\sqrt{(1+\rho)/(1-\rho)}$, one can rewrite (3.61) as

$$\eta^*=(1-\rho)(I_n\otimes i_T^\alpha)u+(I_n\otimes C)\varepsilon \qquad (3.62)$$

Since $E(C\varepsilon_i\varepsilon_i'C')=\sigma_v^2 I_T$, $E\left[(I_n\otimes C)\varepsilon\varepsilon'(I_n\otimes C')\right]=\sigma_v^2 I_{nT}$;The variance–covariance matrix of the transformed disturbances is

$$\Omega^*=E\left(\eta^*\eta^{*\prime}\right)=\sigma_u^2(1-\rho)^2\left[I_n\otimes i_T^\alpha i_T^{\alpha'}\right]+\sigma_v^2 I_{nT}$$

Alternatively, this can be rewritten as

$$\Omega^*=d^2\sigma_u^2(1-\rho)^2\left[I_n\otimes i_T^\alpha i_T^{\alpha'}/d^2\right]+\sigma_v^2 I_{nT} \qquad (3.63)$$

where $d^2=i_T^{\alpha'}i_T^\alpha=\alpha^2+(T-1)$. This replaces $J_T^\alpha=i_T^\alpha i_T^{\alpha'}$ by $d^2\bar{J}_T^\alpha$, its idempotent counterpart, where $\bar{J}_T^\alpha=i_T^\alpha i_T^{\alpha'}/d^2$. Extending the Wansbeek and Kapteyn trick, we replace I_T by $E_T^\alpha+\bar{J}_T^\alpha$, where $E_T^\alpha=I_T-\bar{J}_T^\alpha$. Collecting terms with the same matrices, one obtains the spectral decomposition of Ω^*,

$$\Omega^*=\sigma_\alpha^2(I_n\otimes\bar{J}_T^\alpha)+\sigma_v^2(I_n\otimes E_T^\alpha) \qquad (3.64)$$

where $\sigma_\alpha^2=d^2\sigma_u^2(1-\rho)^2+\sigma_v^2$. Therefore

$$\sigma_v\Omega^{*-1/2}=(\sigma_v/\sigma_\alpha)(I_n\otimes\bar{J}_T^\alpha)+(I_n\otimes E_T^\alpha)=I_{nT}-\theta_\alpha(I_n\otimes\bar{J}_T^\alpha) \qquad (3.65)$$

where $\theta_\alpha=1-(\sigma_v/\sigma_\alpha)$.

Premultiplying the PW transformed observations $Y^*=(I_n\otimes C)Y$ by $\sigma_v\Omega^{*-1/2}$ one gets $Y^{**}=\sigma_v\Omega^{*-1/2}Y^*$. The typical elements of $Y^{**}=\sigma_v\Omega^{*-1/2}Y^*$ are given by

$$\left(Y_{i1}^{*}-\theta_{\alpha}\alpha b_{i},Y_{i2}^{*}-\theta_{\alpha}b_{i},\cdots,Y_{iT}^{*}-\theta_{\alpha}b_{i}\right)' \qquad (3.66)$$

where $b_i=\left[\left(\alpha Y_{i1}^{*}+\sum_{2}^{T}Y_{it}^{*}\right)\Big/d^2\right]$ for $i=1,\cdots,n$. The first observation gets special attention in the $\mathrm{AR}(1)$ error component model. First, the PW transformation gives it a special weight $\sqrt{1-\rho^2}$ in Y^{*}. Second, the Fuller and Battese transformation gives it a special weight $\alpha=\sqrt{(1+\rho)/(1-\rho)}$ in computing the weighted average b_i and the pseudo-difference in (3.66). Note that (i) if $\rho=0$, then $\alpha=1, d^2=T, \sigma_{\alpha}^2=T\sigma_u^2+\sigma_{\varepsilon}^2$ and $\theta_{\alpha}=\theta$. Therefore, the typical element of Y_{it}^{**} reverts to the familiar $\left(Y_{it}-\theta\bar{Y}_i\right)$ transformation for the one-way error component model with no serial correlation. (ii) If $\sigma_u^2=0$ then $\sigma_{\alpha}^2=\sigma_v^2$ and $\theta_{\alpha}=0$. Therefore, the typical element of Y_{it}^{**} reverts to the PW transformation Y_{it}^{*}.

The above estimators of the variance components arise naturally from the spectral decomposition of Ω^{*}. In fact, $\left(I_n\otimes E_T^{\alpha}\right)\eta^{*}\sim\left(0,\sigma_v^2\left[I_n\otimes E_T^{\alpha}\right]\right)$ and $\left(I_n\otimes \bar{J}_T^{\alpha}\right)\eta^{*}\sim\left(0,\sigma_{\alpha}^2\left[I_n\otimes \bar{J}_T^{\alpha}\right]\right)$ and

$$\hat{\sigma}_v^2=\eta^{*\prime}\left(I_n\otimes E_T^{\alpha}\right)\eta^{*}\Big/n(T-1) \text{ and } \hat{\sigma}_{\alpha}^2=\eta^{*\prime}\left(I_n\otimes \bar{J}_T^{\alpha}\right)\eta^{*}\Big/n \qquad (3.67)$$

provide the above estimators of σ_v^2 and σ_{α}^2, respectively. Baltagi and Li (1991a) suggest estimating ρ from Within residuals $\tilde{\varepsilon}_{it}$ as $\tilde{\rho}=\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{i,t-1}\Big/\sum_{i=1}^{n}\sum_{t=2}^{T}\tilde{\varepsilon}_{i,t-1}^2$. Then, $\hat{\sigma}_v^2$ and $\hat{\sigma}_{\alpha}^2$ are estimated from (3.67) by substituting OLS residuals $\hat{\eta}^{*}$ from the PW transformed equation using $\tilde{\rho}$. Using Monte Carlo experiments, Baltagi and Li (1997) found that $\tilde{\rho}$ performs poorly for small T and recommended an alternative estimator of ρ which is based on the autocovariance function $Q_s=E\left(\eta_{it}\eta_{i,t-s}\right)$. For the AR(1) model given in (3.60), it is easy to show that $\rho+1=\left(Q_0-Q_2\right)/\left(Q_0-Q_1\right)$. Hence, a consistent estimator of ρ (for large n) is given by

$$\hat{\rho}=\frac{\tilde{Q}_0-\tilde{Q}_2}{\tilde{Q}_0-\tilde{Q}_1}-1=\frac{\tilde{Q}_1-\tilde{Q}_2}{\tilde{Q}_0-\tilde{Q}_1}$$

where $\tilde{Q}_s = \sum_{i=1}^{n}\sum_{t=s+1}^{T}\hat{\eta}_{it}\hat{\eta}_{i,t-s} \big/ n\left(T-s\right)$ and $\hat{\eta}_{it}$ denotes the OLS residuals on (3.60). $\hat{\sigma}_v^2$ and $\hat{\sigma}_\alpha^2$ are estimated from (3.67) by substituting OLS residuals $\hat{\eta}^*$ from the PW transformed equation using $\hat{\rho}$ rather than $\tilde{\rho}$.

Therefore, the estimation of an AR(1) serially correlated error component model is considerably simplified by (i) applying the PW transformation in the first step, as is usually done in the time-series literature, and (ii) subtracting a pseudo-average from these transformed data as in (3.66) in the second step.

3) The Durbin–Watson Statistic for Balanced Panel Data

For the fixed effects model described in (1.22) with ε_{it} following an AR(1) process, Bhargava, Franzini and Narendranathan (1982), hereafter BFN, suggested testing for $H_0 : \rho = 0$ against the alternative that $\left|\rho\right| < 1$, using the Durbin–Watson statistic only based on the Within residuals (the e_{it}) rather than OLS residuals:

$$d_p = \sum_{i=1}^{n}\sum_{t=2}^{T}\left(e_{it} - e_{i,t-1}\right)^2 \Big/ \sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2 \tag{3.68}$$

BFN showed that for arbitrary regressors, d_p is a locally most powerful invariant test in the neighborhood of $\rho = 0$. They argued that exact critical values are both impractical and unnecessary for panel data since they involve the computation of the nonzero eigenvalues of a large $nT \times nT$ matrix. Instead, BFN show how one can easily compute upper and lower bounds of d_p, and they tabulate the 5% levels for n = 50, 100, 150, 250, 500, 1000, T = 6, 10 and k = 1, 3, 5, 7, 9, 11, 13, 15. BFN remark that d_p would rarely be inconclusive since the bounds will be very tight even for moderate values of N. Also, for very large n, BFN argue that it is not necessary to compute these bounds, but simply test whether d_p is less than two when testing against positive serial correlation.

3.3.2 Unbalanced panels and AR(1) Process

Consider a linear panel-data model described as follows.

$$Y_{it} = \alpha + X_{it}'\beta + u_i + \varepsilon_{it} \quad i = 1, \cdots, n; \;\; t = 1, \cdots, T_i \tag{3.69}$$

$$\varepsilon_{it} = \rho\varepsilon_{it-1} + v_{it}$$

The data can be unbalanced and unequally spaced. Specifically, the dataset contains observations on individual i at times t_{ij} for $j=1,\cdots,n_i$ .The difference $t_{ij}-t_{ij-1}$ plays an integral role, for instance if you have quarterly data, the "time" difference between the third and fourth quarter must be 1 month not 3.

1) Estimating ρ

The estimate of ρ is always obtained after removing the group means. Let $\tilde{Y}_{it_{ij}}=Y_{it_{ij}}-\overline{Y}_i$, $\tilde{X}_{it_{ij}}=X_{it_{ij}}-\overline{X}_{i.}$, and $\tilde{\varepsilon}_{it_{ij}}=\tilde{\varepsilon}_{it_{ij}}-\overline{\varepsilon}_i$

where $\overline{Y}_i=\frac{1}{n_i}\sum_{t=1}^{n_i}Y_{it_{ij}}$, $\overline{X}_{i.}=\frac{1}{n_i}\sum_{t=1}^{n_i}X_{it_{ij}}$, $\overline{\varepsilon}_i=\frac{1}{n_i}\sum_{t=1}^{n_i}\varepsilon_{it_{ij}}$,

We can get the estimates of ρ by running a regression on

$$\tilde{Y}_{it}=\tilde{X}_{it}'\beta+\tilde{\varepsilon}_{it} \tag{3.70}$$

After estimating ρ , Baltagi and Wu(1999) derive a transformation of the data remove the AR(1) component. The transformed $Y_{it_{ij}}$ can be written as

$$Y_{it_{ij}}^*=\begin{cases}\left(1-\rho^2\right)^{1/2}Y_{it_{ij}} & if \quad t_{ij}=1\\ \left(1-\rho^2\right)^{1/2}\left[Y_{i,t_{ij}}\left\{\dfrac{1}{1-\rho^{2\left(t_{ij}-t_{i,j-1}\right)}}\right\}^{1/2}-Y_{i,t_{i,j-1}}\left\{\dfrac{\rho^{2\left(t_{ij}-t_{i,j-1}\right)}}{1-\rho^{2\left(t_{ij}-t_{i,j-1}\right)}}\right\}^{1/2}\right] & if \quad t_{ij}>1\end{cases} \tag{3.71}$$

Using the analogous transform on the independent variables generates transformed data without the AR(1) component. Performing simple OLS on the transformed data leaves behind the residuals v^*.

2) The within estimator of the fixed-effects model

To obtain the within estimator, we must transform the data come from the AR(1) transform. For the within transform to remove the fixed effects, the first observation of each panel must be dropped. Specifically, let

$$\begin{aligned}\breve{Y}_{it_{ij}}&=Y_{it_{ij}}^*-\overline{Y}_i^*+\overline{\overline{Y}}^* \qquad \forall\ j>1\\ \breve{X}_{it_{ij}}&=X_{it_{ij}}^*-\overline{X}_{i.}^*+\overline{\overline{X}}_{.}^* \qquad \forall\ j>1\\ \breve{\varepsilon}_{it_{ij}}&=\eta_{it_{ij}}^*-\overline{\varepsilon}_i^*+\overline{\overline{\varepsilon}}^* \qquad \forall\ j>1\end{aligned} \tag{3.72}$$

where $\overline{Y}_i^*=\dfrac{\sum_{j=2}^{n_i}Y_{it_{ij}}^*}{n_i-1}$, $\overline{\overline{Y}}^*=\dfrac{\sum_{i=1}^{n}\sum_{j=2}^{n_i}Y_{it_{ij}}^*}{\sum_{i=1}^{n}\left(n_i-1\right)}$

$\overline{X}_{i.}^*=\dfrac{\sum_{j=2}^{n_i}X_{it_{ij}}^*}{n_i-1}$, $\overline{\overline{X}}_{.}^*=\dfrac{\sum_{i=1}^{n}\sum_{j=2}^{n_i}X_{it_{ij}}^*}{\sum_{i=1}^{n}\left(n_i-1\right)}$

$$\overline{\varepsilon}_i^* = \frac{\sum_{j=2}^{n_i} \varepsilon_{it_{ij}}^*}{n_i - 1}, \quad \overline{\overline{\varepsilon}}^* = \frac{\sum_{i=1}^{n} \sum_{j=2}^{n_i} \varepsilon_{it_{ij}}^*}{\sum_{i=1}^{n} (n_i - 1)}$$

The within estimator of the fixed-effects model is then obtained by running OLS on

$$\breve{Y}_{it_{ij}} = \alpha + \breve{X}'_{it_{ij}} \beta + \breve{\varepsilon}_{it_{ij}} \tag{3.73}$$

3)The Baltagi-Wu GLS estimator of the random-effects model

The residuals v^* can be used to estimate the variance components. Translating the matrix formulas given in Baltagi and Wu(1999) into summations yields the following variance-components estimators:

$$\hat{\sigma}_\omega^2 = \sum_{i=1}^{n} \frac{\left(v_i^{*\prime} g_i\right)^2}{\left(g_i' g_i\right)}$$

$$\hat{\sigma}_\varepsilon^2 = \frac{\left[\sum_{i=1}^{n} \left(v_i^{*\prime} v_i^*\right) - \sum_{i=1}^{n} \left\{\frac{\left(v_i^{*\prime} g_i\right)^2}{\left(g_i' g_i\right)}\right\}\right]}{\sum_{i=1}^{n} (n_i - 1)}$$

$$\hat{\sigma}_u^2 = \frac{\left[\sum_{i=1}^{n} \left\{\frac{\left(v_i^{*\prime} g_i\right)^2}{\left(g_i' g_i\right)}\right\} - n\hat{\sigma}_\varepsilon^2\right]}{\sum_{i=1}^{n} \left(g_i' g_i\right)}$$

where

$$g_i = \left[1, \frac{\left\{1 - \rho^{\left(t_{i,2} - t_{i,1}\right)}\right\}}{\left\{1 - \rho^{2\left(t_{i,2} - t_{i,1}\right)}\right\}^{\frac{1}{2}}}, \cdots, \frac{\left\{1 - \rho^{\left(t_{i,n_i} - t_{i,n_i-1}\right)}\right\}}{\left\{1 - \rho^{2\left(t_{i,n_i} - t_{i,n_i-1}\right)}\right\}^{\frac{1}{2}}}\right]'$$

and v_i^* is the $n_i \times 1$ vector of residuals from v^* that correspond to person i.

Then

$$\hat{\theta}_i = 1 - \left(\frac{\hat{\sigma}_u}{\hat{w}_i}\right) \tag{3.74}$$

where

$$\hat{w}_i^2 = g_i' g_i \hat{\sigma}_u^2 + \hat{\sigma}_\varepsilon^2$$

With these estimated in hand, we can transform the data via

$$z_{it_{ij}}^{**} = z_{it_{ij}}^{*} - \hat{\theta}_i g_{ij} \frac{\sum_{s=1}^{n_i} g_{is} z_{it_{is}}^{*}}{\sum_{s=1}^{n_i} g_{is}^2}$$

for $z \in \{Y, X\}$.

4) The test statistics

The Baltagi-Wu LBI is the sum of terms

$$d_* = d_1 + d_2 + d_3 + d_4 \tag{3.75}$$

where

$$d_1 = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n_i} \left\{ \tilde{z}_{it_{i,j-1}} - \tilde{z}_{it_{i,j}} I\left(t_{ij} - t_{i,j-1} = 1\right) \right\}^2}{\sum_{i=1}^{n} \sum_{j=1}^{n_i} \tilde{z}_{it_{ij}}^2}$$

$$d_2 = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n_i-1} \tilde{z}_{it_{i,j-1}}^2 \left\{ 1 - I\left(t_{ij} - t_{i,j-1} = 1\right) \right\}^2}{\sum_{i=1}^{n} \sum_{j=1}^{n_i} \tilde{z}_{it_{ij}}^2}$$

$$d_3 = \frac{\sum_{i=1}^{n} \tilde{z}_{it_{i1}}^2}{\sum_{i=1}^{n} \sum_{j=1}^{n_i} \tilde{z}_{it_{ij}}^2}$$

$$d_4 = \frac{\sum_{i=1}^{n} \tilde{z}_{it_{in_i}}^2}{\sum_{i=1}^{n} \sum_{j=1}^{n_i} \tilde{z}_{it_{ij}}^2}$$

$I(\ )$ is the indicator function that takes the value of 1 if the condition is true and 0 otherwise. The $\tilde{z}_{it_{ij}}$ are residuals from the within estimator.

Baltagi and Wu(1999) also show that d_1 is the Bhargava et al. Durbin-Watson statistic modified to handle eases of unbalanced panels and unequally spaced data.

3.4 Heteroscedasticity in the RE Model

In this section, we relax the assumption of homoscedasticity of the disturbances and introduce heteroscedasticity through the u_i as first suggested by Mazodier and Trognon (1978). Next, we suggest an alternative heteroscedastic variance components specification, where only the ε_{it} are heteroscedastic.

3.4.1 The u_i are heteroscedastic

For model (1.34), we assume:

$u_i \sim \left(0, \sigma_{ui}^2\right)$ for $i = 1, \cdots, n$, $\varepsilon_{it} \sim iid\left(0, \sigma_{\varepsilon}^2\right)$ for $i = 1, \cdots, n$, $t = 1, \cdots, T$

In vector form $u \sim (0, \Sigma_u)$,where $\Sigma_u = diag(\sigma_{ui}^2)$ is a diagonal of $n \times n$, and $\varepsilon \sim (0, \sigma_\varepsilon^2 I_{nT})$ .Using $\eta = (I_n \otimes i_T) u + \varepsilon$,the resulting variance matrix of the disturbances is given by

$$V = E(\eta\eta') = diag(\sigma_{ui}^2) \otimes J_T + diag(\sigma_\varepsilon^2) \otimes I_T \tag{3.76}$$

where $diag(\sigma_\varepsilon^2)$ is also of dimension $n \times n$, $J_T = i_T i_T'$. (3.76) can be rewritten as

$$V = diag(\tau_i^2) \otimes J_T / T + diag(\sigma_\varepsilon^2) \otimes \left(I_T - \frac{1}{T} J_T \right) \tag{3.77}$$

with $\tau_i^2 = T\sigma_{ui}^2 + \sigma_\varepsilon^2$. Thus, we have

$$\sigma_\varepsilon V^{-\frac{1}{2}} = diag(\sigma_\varepsilon / \tau_i) \otimes \frac{1}{T} J_T + I_n \otimes \left(I_T - \frac{1}{T} J_T \right) \tag{3.78}$$

Hence, $Y^* = \sigma_\varepsilon V^{-\frac{1}{2}} Y$ has a typical element $y_{it}^* = y_{it} - \theta_i \bar{y}_i$,where $\theta_i = 1 - (\sigma_\varepsilon / \tau_i)$ for $i = 1, \cdots, n$.

3.4.2 The ε_{it} are heteroscedasticity

For model (1.34), we assume:

$u_i \sim iid(0, \sigma_u^2)$, $\varepsilon_{it} \sim (0, \sigma_{\varepsilon i}^2)$, using $\eta = (I_n \otimes i_T) u + \varepsilon$ we get

$$V = E(\eta\eta') = diag(\sigma_u^2) \otimes J_T + diag(\sigma_{\varepsilon i}^2) \otimes I_T \tag{3.79}$$

which can be rewritten as

$$V = diag(T\sigma_u^2 + \sigma_{\varepsilon i}^2) \otimes \frac{1}{T} J_T + diag(\sigma_{\varepsilon i}^2) \otimes \left(I_T - \frac{1}{T} J \right) \tag{3.80}$$

and

$$V^{-\frac{1}{2}} = diag(1/\tau_i) \otimes \frac{1}{T} J_T + diag(1/\sigma_{\varepsilon i}) \otimes \left(I_T - \frac{1}{T} J_T \right) \tag{3.81}$$

and $Y^* = V^{-\frac{1}{2}} Y$ has a typical element

$$y_{it}^* = (\bar{y}_i / \tau_i) + (y_{it} - \bar{y}_i) / \sigma_{\varepsilon i} \tag{3.82}$$

upon rearranging terms, we get

$$y_{it}^* = \frac{1}{\sigma_{\varepsilon i}} (y_{it} - \theta_i \bar{y}_i) \tag{3.83}$$

where $\theta_i = 1 - \sigma_{\varepsilon i} / \tau_i$.

How to implement the FGLS estimation? When $\sigma_{\varepsilon i}^2$ and σ_{ui}^2 are unknown, by replacing the unknown true values with their estimates, a feasible GLS estimator can be implemented. Unfortunately, with a single realization of u_i , there is no way one can get a consistent estimator for σ_{ui}^2 even when $T \to \infty$. The conventional formula

$$\hat{\sigma}_{ui}^2 = \bar{\hat{\eta}}_i^2 - \frac{1}{T}\hat{\sigma}_{\varepsilon i}^2, \qquad i = 1, \cdots, n \tag{3.84}$$

where $\hat{\eta}_{it}$ is the initial estimate of η_{it} (say, the least-square or LSDV estimated residual of (1.34), the $\hat{\sigma}_{ui}^2$ converges to u_i^2 , not σ_{ui}^2 . However, $\sigma_{\varepsilon i}^2$ can be consistently estimated by

$$\hat{\sigma}_{\varepsilon i}^2 = \frac{1}{T-1}\sum_{t=1}^{T}\left(\hat{\eta}_{it} - \bar{\hat{\eta}}_i\right)^2 \tag{3.85}$$

as $T \to \infty$. In the event that $\sigma_{ui}^2 = \sigma_u^2$ for all i, we can estimate σ_u^2 by taking the average of (3.84) across i as their estimates.

It should be noted that when T is finite, there is no way we can get consistent estimates of $\sigma_{\varepsilon i}^2$ and σ_{ui}^2 even when $n \to \infty$. Phillips(2003) argues that this model suffers from the incidental parameters problem and the variance estimates of u_i (the σ_{ui}^2) cannot be estimated consistently. So there is no guarantee that FGLS and true GLS will have the same asymptotic distributions. However, if $\sigma_{ui}^2 = \sigma_u^2$ for all i, then we can get consistent estimates of $\sigma_{\varepsilon i}^2$ and σ_u^2 when both n and T tend to infinity. Substituting $\hat{\sigma}_{\varepsilon i}^2$ and $\hat{\sigma}_u^2$ for $\sigma_{\varepsilon i}^2$ and σ_u^2 in (3.79), we obtain its estimation $\hat{V}$.

Alternatively, one may assume that the conditional variance of u_i conditional on X_i has the same functional form across individuals, $Var\left(u_i \middle| X_i\right) = \sigma^2\left(X_i\right)$, to allow for the consistent estimation of heteroscedastic variance, σ_{ui}^2. The FGLS estimator of β is asymptotically equivalent to the GLS estimator when both n and T approach to infinity. LI and Stengos(1994) considered the regression model given by (1.34) and (1.35) with $u_i \sim iid\left(0, \sigma_u^2\right)$ and $E\left(\varepsilon_{it} \middle| X_{it}\right) = 0$ with $\mathrm{var}\left(\varepsilon_{it} \middle| X_{it}\right) = r\left(X_{it}\right) = r_{it}$. Therefore

$E\left(\eta_{it}^2 \middle| X_{it}\right) = \sigma_u^2 + r_{it}$, and the proposed estimator of σ_u^2 is given by

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{n}\sum_{t \neq s}^{T}\hat{\eta}_{it}\hat{\eta}_{is}}{nT\left(T-1\right)}$$

where $\hat{\eta}_{it}$ denotes the OLS residual. Also

$$\hat{r}_{it} = \frac{\sum_{j=1}^{n}\sum_{s=1}^{T}\hat{\eta}_{is}k_{it,js}}{\sum_{j=1}^{n}\sum_{s=1}^{T}k_{it,js}} - \hat{\sigma}_u^2$$

where the kernel function is given by $k_{it,js} = K\left(\frac{X'_{it} - X'_{js}}{h}\right)$ and h is the smoothing parameter. These estimators of the variance components are used to construct a feasible adaptive GLS estimator of β which they denote by GLSAD.

3.5 Robust Estimators

3.5.1 Robust Coefficient Covariances Estimators

In this section, we describe the basic features of the various robust estimators, for clarity focusing on the simple cases where we compute robust covariance for models estimated by standard OLS without cross-section or period effects. The extensions to models estimated using instrumental variables, fixed or random effects, and GLS weighted least squares are straight forward.

1) White Robust Covariance

The generic pooled model as follows:

$$Y_{it} = X_{it}'\beta + \varepsilon_{it}\,,\quad i = 1,\cdots,n\,;\quad t = 1,\cdots,T \tag{3..86}$$

where β now includes the constant. We assume that the data are stacked by individual:

$$Y_i = X_i\beta + \varepsilon_i \qquad (i = 1,\cdots,n) \tag{3.87}$$

The ordinary least squares estimator of β is

$$\begin{aligned}
\hat{\beta} &= \left(X'X\right)^{-1} X'Y \\
&= \left[\sum_{i=1}^{n} X_i'X_i\right]^{-1} \sum_{i=1}^{n} X_i'Y_i \\
&= \left[\sum_{i=1}^{n} X_i'X_i\right]^{-1} \sum_{i=1}^{n} X_i'\left(X_i\beta + \varepsilon_i\right) \\
&= \beta + \left[\sum_{i=1}^{n} X_i'X_i\right]^{-1} \sum_{i=1}^{n} X_i'\varepsilon_i
\end{aligned}$$

$$\operatorname{var}\left(\hat{\beta}\right) = \left(X'X\right)^{-1}\left(XVX\right)\left(X'X\right)^{-1} \tag{3.88}$$

The true asymptotic covariance matrixes

$$\begin{aligned}
A\operatorname{var}\left(\hat{\beta}\right) &= \frac{1}{nT}\, p\lim\left(\frac{X'X}{nT}\right)^{-1} p\lim\left(\frac{X'VX}{nT}\right) p\lim\left(\frac{X'X}{nT}\right)^{-1} \\
&= \frac{1}{nT}\, p\lim\left(\frac{1}{nT}\sum_{i=1}^{n} X_i'X_i\right)^{-1} p\lim\left(\frac{1}{nT}\sum_{i=1}^{n} X_i'\Omega_i X_i\right) p\lim\left(\frac{1}{nT}\sum_{i=1}^{n} X_i'X_i\right)^{-1}
\end{aligned} \tag{3.89}$$

As before, the center matrix must be estimated by an appropriate method, so as to obtain consistent covariance estimator .

i) White Period Method

Suppose there are arbitrary serial correlation and time-varying variances in the disturbances. The White period robust coefficient covariance estimator is:

$$\frac{n^*}{n^* - k^*}\left(\sum_{i=1}^{n} X_i'X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i'\hat{\varepsilon}_i\hat{\varepsilon}_i'X_i\right)\left(\sum_{i=1}^{n} X_i'X_i\right)^{-1} \tag{3.90}$$

The White period robust coefficient variance estimator obtains from the regression that allows the unconditional variance matrix $E\left(\varepsilon_i \varepsilon_i'\right) = \Sigma_T$ to be unrestricted, and the conditional variance matrix $E\left(\varepsilon_i \varepsilon_i' \middle| X_i\right)$ may depend on X_i in general fashion.For instance. we can estimate this matrix with

$$p\lim\left(\frac{1}{nT}\sum_{i=1}^{n} X_i'\Omega_i X_i\right) = p\lim\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}X_i'\hat{\varepsilon}_i\hat{\varepsilon}_i'X_i$$

$$=\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{\varepsilon}_{it}\hat{\varepsilon}_{is}X_{it}X_{is}'\right)$$

The results is a combination of the White and Newey-West estimator. In (3.90),the leading term is a degree of freedom adjustment depending on the total number of observations in the stacked data, n^* is the total number of stacked observations, and k^*, the total number of estimated parameters.

ii) White cross-section method

$$\frac{n^*}{n^*-k^*}\left(\sum_{t=1}^{T}X_t'X_t\right)^{-1}\left(\sum_{t=1}^{T}X_t'\hat{\varepsilon}_t\hat{\varepsilon}_t'X_t\right)\left(\sum_{t=1}^{T}X_t'X_t\right)^{-1} \qquad (3.91)$$

This estimator is robust to contemporaneous correlation as well as different disturbance variances in each cross-section. Specifically, the unconditional contemporaneous variance matrix $E\left(\varepsilon_t\varepsilon_t'\right) = \Sigma_n$ is unrestricted, and the conditional variance matrix $E\left(\varepsilon_t\varepsilon_t' \middle| X_t\right)$ can depend on X_t in arbitrary unknown fashion.

iii) White diagonal method

$$\frac{n^*}{n^*-k^*}\left(\sum_{i,t}X_{it}X_{it}'\right)^{-1}\left(\sum_{i,t}\hat{\varepsilon}_{it}^2X_{it}X_{it}'\right)\left(\sum_{i,t}X_{it}X_{it}'\right)^{-1} \qquad (3.92)$$

This method allows the unconditional variance matrix $E\left(\varepsilon\varepsilon'\right) = V$ to be an unrestricted diagonal matrix, and the conditional variance $E\left(\varepsilon_{it}^2 \middle| X_{it}\right)$ to depend on X_{it} in general fashion.

iv) The cluster estimator of the VCE (for Stata only)

Stata has implemented an estimator of the VCE that is robust to the correlation of disturbances within groups and to not identically distributed disturbances. It is commonly refered to as the cluster-robust-VCE estimator, because these groups are known as clusters. Within-cluster correlation allows the matrix V in (3.88) to be block-diagonal,i.e,

$V = diag(\Omega_1, \cdots, \Omega_n)$, it allows the disturbances within each clusters to be with each other but requires that the disturbances from difference clusters be uncorrelated.

The cluster robust VCE estimator is:

$$\frac{n^* - 1}{n^* - k^*} \frac{M}{M - 1} \left(X'X \right)^{-1} \left(\sum_{i=1}^{M} \tilde{\omega}_i \tilde{\omega}_i' \right) \left(X'X \right)^{-1} \qquad (3.93)$$

Where M is the number of clusters, $\tilde{\omega}_i = \sum_{t=1}^{T_i} \hat{\varepsilon}_{it} X_{it}$, T_i is the number of observations in the i th cluster, $\hat{\varepsilon}_{it}$ is the t th residual from the i th cluster,and X_{it} is the column vector of regressors from the t th observation in the i th cluster.

3.5.2. Ordinary Least Squares with Panel-Corrected Standard Errors.

1) Assess Parks Method

The Parks method is FGLS for panel data models where the disturbances show panel heteroscedasticity, contemporaneous correlation, and individual specific serial correlation. The correlation for contemporaneous correlation of the disturbances automatically corrects for any panel heteroscedasticity, so we need only consider the corrections for contemporaneous and serial correlation of the disturbances here.

The Parks method consists of two sequential FGLS transformations, first eliminating serial correlation of the disturbances then eliminating contemporaneous correlation of the disturbances. This is done by initially estimating model (3.86) by OLS. We now consider the consequences of the two corrections separately.

The Parks correction for contemporaneously correlated disturbances cannot be used unless T is at least as big as n (Beck et al. 1993). But even when T is greater than n, so that FGLS can be used, estimation of standard errors is problematic unless T is considerably larger than n. Each element of the matrix of contemporaneous covariances of the covariances of the disturbances is estimated using, on average, $2T/n$ observations. Many cross-national panel studies have ratios of T to n very close to 1, so covariances are being estimated with only slightly more than two observations per estimate!

On the other hand, the FGLS correction for individual-specific serially correlated disturbances, used by Parks, is likely to cause serious underestimates of variability. The essence of the problem is that each ρ_i is estimated using an autoregression based on only T observations. It is well known that such estimates are biased downward (Hurwicz 1950). As a consequence, the Parks estimates, which correct based on these inaccurate autoregressions, may be inferior to OLS estimates. The underestimates of the ρ_i, when combined with trending data, can cause the Parks estimates of standard errors to misestimate variability substantially.

2) Panel-corrected standard errors (PCSE)

If the disturbances in model (3.86) are not spherical ,then OLS estimates of β will be consistent but inefficient; the degree of inefficiency depends on the data and the exact form of the disturbance process. The OLS standard errors will also be inaccurate, but they can be corrected so that they provide accurate estimated of the variability of the OLS estimates of β. This correction takes into account the contemporaneous correlation of the disturbances (and perforce heteroscedasticity). Any serial correlation of the disturbances must be eliminated before the panel-corrected standard errors are calculated.

The correct formula for the sampling variability of the OLS estimated is given by the square roots of the diagonal terms of (3.88). If the disturbances obey the spherical assumption, this simplifies to the usual OLS formula, where the OLS standard errors are the square roots of the diagonal terms of $\hat{\sigma}^2\left(X'X\right)^{-1}$, and $\hat{\sigma}^2$ is the usual OLS estimator . If the disturbances obey the panel structure, then this formula provides incorrect standard errors. However,expression (3.79), can still be used, in combination with that panel structure of the disturbances, to provide accurate, panel-corrected standard errors (PCSEs).

For example, the Cross-section SUR (PCSE) method replaces the outer product of the cross-section residuals in equation (3.86) with an estimate of the cross-section residual (contemporaneous) covariance matrix Σ_n:

$$\frac{n^*}{n^*-k^*}\left(\sum_{t=1}^{T} X_t'X_t\right)^{-1}\left(\sum_{t=1}^{T} X_t'\Sigma_n X_t\right)\left(\sum_{t=1}^{T} X_t'X_t\right)^{-1} \tag{3.94}$$

Analogously, the Period SUR (PCSE) replaces the outer product of the period residuals in equation (3.77) with an estimate of the period covariance Σ_T:

$$\frac{n^*}{n^*-k^*}\left(\sum_{i=1}^{n} X_i'X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i'\Sigma_T X_i\right)\left(\sum_{i=1}^{n} X_i'X_i\right)^{-1} \tag{3.95}$$

The two diagonal forms of these estimators, Cross-section weights (PCSE), and Period weights (PCSE), use only the diagonal elements of the relevant $\hat{\Sigma}_n$ and $\hat{\Sigma}_T$. These covariance estimators are robust to heteroskedasticity across-sections or periods, respectively, but not to general correlation of residuals.

We can compute either nondegree of freedom corrected version or $\frac{n^*}{n^*-k^*}$ degree of freedom corrected versions of all the robust coefficient covariance estimators.