A distribution-free joint test and one-directional robust tests for random individual effects and heteroscedasticity allowing for unbalanced panels

Bernard Lejeune

University of Liège, ERUDITE and CORE

Boulevard du Rectorat, 7, B33

4000 Liège Belgium

E-mail: B.Lejeune@ulg.ac.be

Econometric Society European Meeting
Santiago de Compostela, 1999

Plan

1- Introduction

2- The proposed joint test

3- A BMCP based on robust one-directional tests

4- An empirical illustration

1. Introduction

- Background: when analyzing microeconomic panel data, the error term ε_{it} of the regression model

$$Y_{it} = X_{it}\beta + \varepsilon_{it}$$

may be expected to be:

  - serially correlated (non-zero covariance) in the time-series dimension, presumably due to individual effects,
  - heteroscedastic in the cross-section dimension.

Practitioners usually focus on detecting and modelling expected serial correlation through an error components model but most of the time completely ignore the possibility of heteroscedasticity. This may result in:

  - inefficient (but nevertheless consistent) estimation,
  - misleading inferences (standard errors, tests of individual effects, diagnostic tests,...).

- Questions of interest:

  - How to detect, from preliminary OLS estimation of the model, the possible simultaneous presence of serial correlation and heteroscedasticity?
  - How to discriminate between them?

- In this paper, we propose:

  - a joint test for detecting, from OLS residuals, the presence of serial correlation — presumably due to individual effects — and/or heteroscedasticity,
  - a BMCP based on robust one-directional statistics for identifying the source(s) of departure from the joint null when it is rejected.

Throughout, we allow for unbalanced panels ("ignorable selection rule") and make no distributional assumptions.

2. The proposed joint test

2.1. Formulation of the test

- We consider the following linear regression model:

$$Y_{it} = X_{it}\beta + \varepsilon_{it} \quad i = 1, 2, ..., n \; ; \; t = 1, 2, ..., T_i$$

  Observations are assumed independent across individuals.

- We are interested in testing the joint null of no serial correlation and no heteroscedasticity. Using the possible heteroscedastic one-way error component model

$$Y_{it} = X_{it}\beta + \varepsilon_{it} \, , \; \varepsilon_{it} = \mu_i + \nu_{it} \, , \quad i = 1, 2, ..., n \; ; \; t = 1, 2, ..., T_i$$

  as an auxiliary nested alternative and stacking the T_i observations of each individual i, such a test may be expressed as testing the null

$$\mathrm{H}_0 : \begin{cases} E(Y_i|X_i, Z_i) = X_i\beta^o \\ V(Y_i|X_i, Z_i) = \sigma_\nu^{2o} I_{T_i} \end{cases} , \; i = 1, ..., n$$

  against the alternative

$$\mathrm{H}_1 : \begin{cases} E(Y_i|X_i, Z_i) = X_i\beta^o \\ V(Y_i|X_i, Z_i) = \sigma_\nu^{2o} \, \mathrm{diag}\left(\phi(Z_i\gamma^o)\right) + \sigma_\mu^{2o} J_{T_i} \end{cases} , \; i = 1, ..., n$$

  where, following Breush-Pagan (1979), $\phi(.)$ is an arbitrary function satisfying $\phi(.) > 0$, $\phi(0) = 1$ and $\phi'(0) \neq 0$ (ex.: $\exp(.)$, $(1 + .)^\delta$, ...).

  Remarks:

  - Testing H_0 against H_1 means testing $\sigma_\mu^{2o} = 0$ and $\gamma^o = 0$.
  - H_0 maintains the hypothesis that the conditional mean is correctly specified with respect to (X_i, Z_i).
  - H_1 does not explicitly allow for heteroscedasticity in μ_i. The joint test will however not be insensitive to such a pattern.
  - Regarding ν_{it}, H_1 allows for a quite broad class of heteroscedastic models.

2.2. The test statistic

- A m-testing approach: H_0 implies that

$$E\left[\mathrm{vec}(u_i^o u_i^{o\prime} - \sigma_\nu^{2o} I_{T_i})|X_i, Z_i\right] = 0\,, \quad u_i = Y_i - X_i\beta\,, \quad i = 1, ..., n$$

so that for any "regular" matrix function $F_i(X_i, Z_i)$, we have

$$E\left[F_i(X_i, Z_i)'\,\mathrm{vec}(u_i^o u_i^{o\prime} - \sigma_\nu^{2o} I_{T_i})\right] = 0\,, \quad i = 1, ..., n$$

$\rightarrow$ H_0 may be tested by checking through a m-test the nullity of the misspecification indicator

$$\tilde{\Phi}_n = \frac{1}{n}\sum_{i=1}^{n} F_i(X_i, Z_i)'\,\mathrm{vec}(\tilde{u}_i \tilde{u}_i' - \tilde{\sigma}_\nu^2 I_{T_i})$$

$F_i(X_i, Z_i)$ determines the direction(s) in which the test has power.

- A natural choice for $F_i(X_i, Z_i)$ arises from the gaussian joint LM test of $\sigma_\mu^{2o} = 0$ and $\gamma^o = 0$ in model H_1. This test basically amounts to checking the nullity of the $(1 + k_\gamma) \times 1$ misspecification indicator

$$\tilde{\Phi}_n^{I_r H} = \frac{1}{n}\sum_{i=1}^{n} W_i^{I_r H\prime}\tilde{v}_i = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} \tilde{u}_i'(J_{T_i} - I_{T_i})\tilde{u}_i \\ \frac{1}{n}\sum_{i=1}^{n} Z_i'\left(\tilde{u}_i^{\hat{}2} - \tilde{\sigma}_\nu^2 e_{T_i}\right) \end{bmatrix}$$

where

$$\begin{aligned} \tilde{v}_i &= \mathrm{vec}(\tilde{u}_i \tilde{u}_i' - \tilde{\sigma}_\nu^2 I_{T_i}) \\ W_i^{I_r H} &= \left[\ W_i^{I_r} \quad W_i^H\ \right] = \left[\ \mathrm{vec}\, J_{T_i} \quad \mathrm{diag}\,(\mathrm{vec}\, I_{T_i})\,(Z_i \otimes e_{T_i})\ \right] \end{aligned}$$

Remarks:

- The precise form of $\phi(.)$ does not matter.
- The first component of $\tilde{\Phi}_n^{I_r H}$ checks for zero covariances and the second for constant variances.

- A m-test of the nullity of $\tilde{\Phi}_n^{I_r H}$ is obtained from the asymptotic chi-square statistic

$$\mathcal{PLM}_n = n\tilde{\Phi}_n^{I_r H\prime}\tilde{V}_n^{-1}\tilde{\Phi}_n^{I_r H} \xrightarrow{d} \chi^2(1+k_\gamma)$$

where $\tilde{V}_n$ is, under H_0, a consistent estimator of

$$V_n^o = \frac{1}{n}\sum_{i=1}^{n} E\left[\left(W_i^{I_r H} - R_i P_n\right)' v_i^o v_i^{o\prime}\left(W_i^{I_r H} - R_i P_n\right)\right]$$

and

$$R_i = \operatorname{vec} I_{T_i}\,, \quad P_n = \left(\sum_{i=1}^{n} R_i' R_i\right)^{-1} \sum_{i=1}^{n} E\left[R_i' W_i^{I_r H}\right]$$

Essentially, the validity of $\mathcal{PLM}_n$ does not rely on other assumptions than H_0 itself. It is robust to non-normality (i.e., distribution-free) while, by construction, asymptotically equivalent to the gaussian joint LM test under normality. It is however not very simple.

- $\mathcal{PLM}_n$ may be considerably simplified if we are willing to assume that, under H_0, there is no dynamic in the three first conditional moments of the errors u_{it}^o, i.e., if for all i and t

$$\begin{aligned} E(u_{it}^o|X_i, Z_i, u_{it-1}^o, ..., u_{i1}^o) &= E(u_{it}^o|X_i, Z_i) = 0 \\ E(u_{it}^{o2}|X_i, Z_i, u_{it-1}^o, ..., u_{i1}^o) &= E(u_{it}^{o2}|X_i, Z_i) = \sigma_\nu^{2o} \\ E(u_{it}^{o3}|X_i, Z_i, u_{it-1}^o, ..., u_{i1}^o) &= E(u_{it}^{o3}|X_i, Z_i) \end{aligned}$$

Remark: these conditions are automatically satisfied if the u_{it}^o are conditionally independently — but not necessarily identically — distributed across t, and thus also under conditional joint normality.

→ Under this additional auxiliary assumption, $\mathcal{PLM}_n$ turns out to be the sum of two asymptotically independent pseudo-LM statistics:

$$\mathcal{PLM}_n^{I_r H} = \mathcal{PLM}_n^{I_r} + \mathcal{PLM}_n^{H} \xrightarrow{d} \chi^2(1+k_\gamma)$$

where

$$\mathcal{PLM}_n^{I_r} = \frac{\left(\sum_{i=1}^{n} \tilde{u}_i'(J_{T_i} - I_{T_i})\tilde{u}_i\right)^2}{2\tilde{\sigma}_\nu^4 \sum_{i=1}^{n}(T_i^2 - T_i)} \xrightarrow{d} \chi^2(1)$$

$$\mathcal{PLM}_n^{H} = \left(\sum_{i=1}^{n} Z_i' \left(\hat{\tilde{u}_i^2} - \tilde{\sigma}_\nu^2 e_{T_i}\right)\right)' \left(\sum_{i=1}^{n} \underline{Z}_i' \left(\mathrm{diag}(\hat{\tilde{u}_i^4}) - \tilde{\sigma}_\nu^4 I_{T_i}\right) \underline{Z}_i\right)^{-1} \left(\sum_{i=1}^{n} Z_i' \left(\hat{\tilde{u}_i^2} - \tilde{\sigma}_\nu^2 e_{T_i}\right)\right) \xrightarrow{d} \chi^2(k_\gamma)$$

Remarks :

- $\mathcal{PLM}_n^{I_r}$ is the incomplete panel version of the Breush-Pagan (1980) standard LM test for random individual effects derived in Baltagi-Li (1990). The balanced version of this standard LM test was shown to be robust to non-normality by Honda (1985) under the assumptions of non-stochastic regressors and i.i.d. errors.
- $\mathcal{PLM}_n^{H}$ is a variant of a regression-based statistic proposed by Wooldridge (1990). It contains as special cases well-known tests for heteroscedasticity. So, assuming $E(u_{it}^{o4}|X_i, Z_i) = \delta^o$, $\mathcal{PLM}_n^{H}$ collapses to the Koenker's (1981) test statistic. Likewise, assuming $E(u_{it}^{o4}|X_i, Z_i) = 3\sigma_\nu^{4o}$, we obtain the standard Breush-Pagan's (1979) test statistic.
- For practical purpose, $\mathcal{PLM}_n^{H}$ may be computed as N minus the residual sum of squares of the artificial OLS regression

$$1 = \left[\left(\tilde{u}_{it}^2 - \tilde{\sigma}_\nu^2\right) \underline{Z}_{it}\right] b + \text{residuals}, \quad i = 1, 2, ..., n \; ; \; t = 1, 2, ..., T_i$$

This is the Wooldridge's (1990) test statistic.

3. A BMCP based on robust one-directional tests

- Question : how to identify the source(s) of departure from H_0 when it is rejected ?

- First answer : given the additive structure (asymptotic independence) of $\mathcal{PLM}_n^{I_r H}$, it is tempting to identify the direction(s) in which misspecification detected by the joint statistic may lie by looking at the one-directional statistics $\mathcal{PLM}_n^{I_r}$ and $\mathcal{PLM}_n^{H}$.

  $\rightarrow$ Use a Bonferroni approach : identify the source(s) of departure from H_0 detected by the joint test at asymptotic size α as given by the one-directional test statistic(s) rejected at asymptotic size $\alpha/2$.

- Problem : the one-directional statistics $\mathcal{PLM}_n^{I_r}$ and $\mathcal{PLM}_n^{H}$ are not reliable since each one may be “contaminated” by a departure from the joint null in the other direction.

- Suggested solution : use a Bonferroni approach based on robust versions of the one-directional statistics.

3.1. Robust one-directional statistics

3.1.1. A robust test for random individual effects

- We are interested in testing the null of no serial correlation while allowing for arbitrary pattern of heteroscedasticity under the null. Keeping the same framework as above, such a test may be expressed as testing the null

$$\mathrm{H}_0^{Ir}:\begin{cases} E(Y_i|X_i,Z_i)=X_i\beta^o & i=1,...,n \\ Cov(Y_{it},Y_{it'}|X_i,Z_i)=0 & t,t'=1,...,T_i \;\; (t\neq t') \end{cases}$$

against the alternative

$$\mathrm{H}_1^{Ir}:\begin{cases} E(Y_i|X_i,Z_i)=X_i\beta^o & i=1,...,n \\ Cov(Y_{it},Y_{it'}|X_i,Z_i)=\sigma_\mu^{2o} & t,t'=1,...,T_i \;\; (t\neq t') \end{cases}$$

H_0^{Ir} and H_1^{Ir} are alike to H_0 and H_1 except that they leave unspecified the variances.

- A m-testing approach: H_0^{Ir} implies that

$$E\left[\mathrm{vech}_{\mathrm{wd}}(u_i^o u_i^{o\prime})|X_i,Z_i\right]=0\,,\quad i=1,...,n$$

so that for any "regular" matrix function $F_i(X_i,Z_i)$, we have

$$E\left[F_i(X_i,Z_i)'\mathrm{vech}_{\mathrm{wd}}(u_i^o u_i^{o\prime})\right]=0\,,\quad i=1,...,n$$

$\rightarrow$ H_0^{Ir} may be tested by checking through a m-test the nullity of the misspecification indicator

$$\tilde{\Phi}_n=\frac{1}{n}\sum_{i=1}^{n}F_i(X_i,Z_i)'\mathrm{vech}_{\mathrm{wd}}(\tilde{u}_i\tilde{u}_i')$$

$F_i(X_i,Z_i)$ determines the direction(s) in which the test has power.

- $\mathcal{PLM}_n^{Ir}$ amounts to checking the nullity of a special case of $\tilde{\Phi}_n$, namely of the misspecification indicator

$$\tilde{\Phi}_n^{Ir}=\frac{1}{n}\sum_{i=1}^{n}\tilde{u}_i'(J_{T_i}-I_{T_i})\tilde{u}_i=\frac{1}{n}\sum_{i=1}^{n}2e'_{(T_i^2-T_i)/2}\mathrm{vech}_{\mathrm{wd}}(\tilde{u}_i\tilde{u}_i')$$

- A m-test of the nullity of $\tilde{\Phi}_n^{I_r}$ is obtained from the asymptotic chi-square statistic

$$\mathcal{RPLM}_n^{Ir} = n\tilde{\Phi}_n^{I_r\prime}\tilde{V}_n^{-1}\tilde{\Phi}_n^{I_r} \xrightarrow{d} \chi^2(1)$$

where $\tilde{V}_n$ is, under H_0^{Ir}, a consistent estimator of

$$V_n^o = \frac{1}{n}\sum_{i=1}^{n} E\left[\left(2e'_{(T_i^2-T_i)/2}\mathrm{vech}_{\mathrm{wd}}(u_i^o u_i^{o\prime})\right)^2\right]$$

so that an appropriate statistic turns out to be

$$\mathcal{RPLM}_n^{Ir} = \frac{\left(\sum_{i=1}^{n}\tilde{u}_i'(J_{T_i}-I_{T_i})\tilde{u}_i\right)^2}{\sum_{i=1}^{n}\left(\tilde{u}_i'(J_{T_i}-I_{T_i})\tilde{u}_i\right)^2} \xrightarrow{d} \chi^2(1)$$

Remarks:

  - Essentially, the validity of $\mathcal{RPLM}_n^{I_r}$ does not rely on other assumptions than H_0^{Ir} itself. It is distribution-free and robust to heteroscedasticity while, by construction, asymptotically equivalent to $\mathcal{PLM}_n^{I_r}$ under H_0 and the auxiliary assumption of no dynamic in (or conditional independence of) the errors u_{it}^o.
  - A (numerically equal) variant of $\mathcal{RPLM}_n^{I_r}$ was already derived by Li-Stengos (94) in a balanced panel data framework using the unnecessary auxiliary assumption of conditional independence of the errors u_{it}^o across t.

3.1.2. A robust test for heteroscedasticity

- We are interested in testing the null of no heteroscedasticity while allowing for arbitrary pattern of serial correlation under the null. Keeping again the same framework as above, such a test may be expressed as testing the null

$$\mathrm{H}_0^H : \begin{cases} E(Y_i|X_i, Z_i) = X_i\beta^o \\ V(Y_{it}|X_i, Z_i) = \sigma_\nu^{2o} \end{cases}, \quad \begin{matrix} i = 1, ..., n \\ t = 1, ..., T_i \end{matrix}$$

against the alternative

$$\mathrm{H}_1^H : \begin{cases} E(Y_i|X_i, Z_i) = X_i\beta^o \\ V(Y_{it}|X_i, Z_i) = \sigma_\nu^{2o}\, \mathrm{diag}\,(\phi(Z_i\gamma^o)) \end{cases}, \quad \begin{matrix} i = 1, ..., n \\ t = 1, ..., T_i \end{matrix}$$

H_0^H and H_1^H are alike to H_0 and H_1 except that they leave unspecified the covariances.

- A m-testing approach: H_0^H implies that

$$E\left[(u_i^{o\,\hat{}\,2} - \sigma_\nu^{2o} e_{T_i})|X_i, Z_i\right] = 0\,, \quad i = 1, ..., n$$

so that for any "regular" matrix function $F_i(X_i, Z_i)$, we have

$$E\left[F_i(X_i, Z_i)'(u_i^{o\,\hat{}\,2} - \sigma_\nu^{2o} e_{T_i})\right] = 0\,, \quad i = 1, ..., n$$

$\rightarrow$ H_0^H may be tested by checking through a m-test the nullity of the misspecification indicator

$$\tilde{\Phi}_n = \frac{1}{n}\sum_{i=1}^{n} F_i(X_i, Z_i)'(\tilde{u}_i^{\,\hat{}\,2} - \tilde{\sigma}_\nu^2 e_{T_i})$$

$F_i(X_i, Z_i)$ determines the direction(s) in which the test has power.

- $\mathcal{PLM}_n^H$ amounts to checking the nullity of a special case of $\tilde{\Phi}_n$, namely of the misspecification indicator

$$\tilde{\Phi}_n^H = \frac{1}{n}\sum_{i=1}^{n} Z_i'\left(\tilde{u}_i^{\,\hat{}\,2} - \tilde{\sigma}_\nu^2 e_{T_i}\right)$$

- A m-test of the nullity of $\tilde{\Phi}_n^H$ is obtained from the asymptotic chi-square statistic

$$\mathcal{RPLM}_n^H = n\tilde{\Phi}_n^{H\prime}\tilde{V}_n^{-1}\tilde{\Phi}_n^H \xrightarrow{d} \chi^2(1)$$

where $\tilde{V}_n$ is, under H_0^H, a consistent estimator of

$$V_n^o = \frac{1}{n}\sum_{i=1}^{n} E\left[(Z_i - R_i P_n)' \,\bar{v}_i^o \bar{v}_i^{o\prime}\,(Z_i - R_i P_n)\right]$$

and

$$\begin{aligned}
\bar{v}_i^o &= u_i^{o\hat{}2} - \sigma_\nu^{2o} e_{T_i} \\
R_i &= e_{T_i}\,, \;\; P_n = \left(\sum_{i=1}^{n} R_i' R_i\right)^{-1} \sum_{i=1}^{n} E\left[R_i' Z_i\right]
\end{aligned}$$

so that an appropriate statistic turns out to be

$$\begin{aligned}
\mathcal{PLM}_n^H \;=\; & \left(\sum_{i=1}^{n} Z_i' \left(\tilde{u}_i^{\hat{}2} - \tilde{\sigma}_\nu^2 e_{T_i}\right)\right)' \left(\sum_{i=1}^{n} \underline{Z}_i' \left(\tilde{u}_i^{\hat{}2} - \tilde{\sigma}_\nu^2 e_{T_i}\right)\left(\tilde{u}_i^{\hat{}2} - \tilde{\sigma}_\nu^2 e_{T_i}\right)' \underline{Z}_i\right)^{-1} \\
& \left(\sum_{i=1}^{n} Z_i' \left(\tilde{u}_i^{\hat{}2} - \tilde{\sigma}_\nu^2 e_{T_i}\right)\right) \xrightarrow{d} \chi^2(k_\gamma)
\end{aligned}$$

Remarks :

- Essentially, the validity of $\mathcal{RPLM}_n^H$ does not rely on other assumptions than H_0^H itself. It is distribution-free and robust to serial correlation while, by construction, asymptotically equivalent to $\mathcal{PLM}_n^H$ under H_0 and the auxiliary assumption of no dynamic in (or conditional independence of) the errors u_{it}^o.
- For practical purpose, $\mathcal{RPLM}_n^H$ may be computed as n minus the residual sum of squares of the artificial OLS regression

$$1 = \left[\left(\tilde{u}_i^{\hat{}2} - \tilde{\sigma}_\nu^2 e_{T_i}\right)' \underline{Z}_i\right] b + \text{residuals}, \quad i = 1, 2, ..., n$$

3.2. The proposed testing procedure

- To sum up, for detecting and identifying the source(s) of departure from the joint null H_0 of no serial correlation (individual effects) and no heteroscedasticity using OLS residuals, we propose the following testing procedure:

  (1) Use $\mathcal{PLM}_n^{I_rH}$ to test at asymptotic size α the joint null H_0.

  (2) If the joint null is rejected, identify the source(s) of departure from H_0 as given by the robust one-directional test statistics $\mathcal{RPLM}_n^{Ir}$ and $\mathcal{RPLM}_n^{H}$ rejected at asymptotic size $\alpha/2$.

- Remarks :
  - It is perfectly possible that the test statistic $\mathcal{PLM}_n^{I_rH}$ rejects the joint null H_0 while none of the one-directional test statistics $\mathcal{RPLM}_n^{Ir}$ and $\mathcal{RPLM}_n^{H}$ rejects its own null (H_0^{Ir} and H_0^{H}).
  - A rejection of the test statistics $\mathcal{PLM}_n^{I_rH}$, $\mathcal{RPLM}_n^{Ir}$ and $\mathcal{RPLM}_n^{H}$ may actually be due to misspecification of the conditional mean, and for $\mathcal{PLM}_n^{I_rH}$, to the presence of dynamic in (or lack of independence of) the errors under the null.
  - For practical purpose, we may view this procedure as a simple and convenient way to decide, from preliminary OLS estimation, which model and inferential method to use in further analysis of the data. Anyway, the validation of the chosen model — both in mean and variance — will require new diagnostic tests.

4. An empirical illustration

- Data from the "Marchés et Statégies d'Entreprises" Division of INSEE :
  - Inputs-output production records.
  - 824 french firms observed over the period 1979-1988.
  - 5 201 observations : only 1/3 of the firms are observed over the entire period.

- Estimation of an inter-sectorial translog production function :

$$V_{it} = \beta_{(sc\times t)} + \beta^k K_{it} + \beta^l L_{it} + \beta^{kk} K_{it}^2 + \beta^{ll} L_{it}^2 + \beta^{kl} K_{it} L_{it} + \varepsilon_{it}$$

where

$$V_{it} = \ln va_{it}, \quad K_{it} = (\ln k_{it} - \ln k^*), \quad L_{it} = (\ln l_{it} - \ln l^*),$$

va = deflated value added, k = stock of capital, l = number of workers

- Pseudo-LM tests :

Sectorial dummies, K_{it} and L_{it} are used as Z_{it} variables. This allow variances to change according to sectors, size and input ratios.

Table 1 : Pseudo-LM tests

	$\mathcal{PLM}_n^{I_r H}$	$\mathcal{PLM}_n^{I_r}$	$\mathcal{PLM}_n^{H}$	$\mathcal{RPLM}_n^{I_r}$	$\mathcal{RPLM}_n^{H}$
Stat.	11 330.5	10 981.0	349.4	160.7	58.5
D.f.	11	1	10	1	10
p-value	0.00000	0.00000	0.00000	0.00000	0.00000

$\rightarrow$ Both individual effects and heteroscedasticity like patterns are present in the data. This suggests considering a heteroscedastic error components model.
Note also the large differences between the robust and non-robust statistics.