A full heteroscedastic one-way error components model for incomplete panel: Pseudo maximum likelihood estimation and specification testing

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
Istanbul, 25-29 August 1996

Plan

1- Motivation: microeconomic panel data and heteroscedasticity

2- The proposed heteroscedastic one-way error components model

3- Second order pseudo maximum likelihood estimation and properties

4- Specification testing:

- Joint pseudo Lagrange multiplier testing and a BMCP
- Information matrix testing

5- An empirical illustration

1. Motivation: Microeconomic panel data and heteroscedasticity

- Heteroscedasticity = endemic when working with microeconomic cross-section data

  – Standard regression model:

$$\begin{aligned} &Y_i = X_i\beta + \varepsilon_i, \\ &\varepsilon_i \sim \text{i.d.}(0,\sigma^2) \end{aligned} \qquad i = 1, ..., m$$

  – Various sources of heteroscedasticity:

    - Differences in size (mechanical source)
    - Heteroscedasticity not directly related to size (heterogeneity)

- Heteroscedasticity in cross-section $\Rightarrow$ heteroscedasticity in panel data

  – Standard one-way error components model:

$$\begin{aligned} &Y_{ij} = X_{ij}\beta + \varepsilon_{ij}, \quad \varepsilon_{ij} = \mu_i + \nu_{ij}, \\ &\nu_{ij} \sim \text{i.d.}(0,\sigma^2_\nu), \quad \mu_i \sim \text{i.d.}(0,\sigma^2_\mu), \end{aligned} \qquad \begin{aligned} &i = 1, ..., m; \\ &j = 1, 2, ..., t \end{aligned}$$

  – Same heteroscedasticity sources but at two different levels: within and/or between

- Heteroscedastic specifications available in the literature:

  a- Mazodier-Trognon (1978), Baltagi-Griffin (1988):

  $$\nu_{ij} \sim \text{i.d.}(0, \sigma^2_{\nu}) \text{ and } \mu_i \sim \text{i.d.}(0, \sigma^2_{\mu_i})$$

  b- Rao and al. (1981), Magnus (1982), Baltagi (1988) - Wansbeek (1988):

  $$\nu_{ij} \sim \text{i.d.}(0, \sigma^2_{\nu_i}) \text{ and } \mu_i \sim \text{i.d.}(0, \sigma^2_{\mu})$$

  c- Li-Stengos (1994):

  $$\nu_{ij} \sim \text{i.d.}(0, \sigma^2_{\nu_{ij}}) \text{ and } \mu_i \sim \text{i.d.}(0, \sigma^2_{\mu})$$
  $$\text{where } \sigma^2_{\nu_{ij}} = \text{a non-parametric function f}\,(Z_{ij})$$

  d- Verbon (1980): (for LM testing purpose)

  $$\nu_{ij} \sim \text{i.d.}(0, \sigma^2_{\nu_i}) \text{ and } \mu_i \sim \text{i.d.}(0, \sigma^2_{\mu_i})$$
  $$\text{where } \sigma^2_{\nu_i} = \sigma^2_{\nu}\ \text{f}\,(Z_i\theta) \text{ and } \sigma^2_{\mu_i} = \sigma^2_{\mu}\ \text{f}\,(Z_i\theta)$$

- Limitations of these specifications:

  – Only *one* heteroscedastic error or imposed identical patterns

  – Grouped heteroscedasticity = incidental parameter problems when m is large but t is small (usual in microeconomic panel data)

  – Does not allow for unbalanced (incomplete) panel

2. The proposed heteroscedastic one-way error components model

- Specification:

$$\begin{array}{lll} Y_{ij} = X_{ij}\beta + \varepsilon_{ij}, & \varepsilon_{ij} = \mu_i + \nu_{ij}, & \\ \nu_{ij} \sim \text{i.d.}(0, \sigma^2_{\nu_{ij}}), & \sigma^2_{\nu_{ij}} = \mathrm{h}_\nu(Z_{ij}\theta_1), & \begin{array}{l} i = 1, ..., m; \\ j = 1, ..., n_i \end{array} \\ \mu_i \sim \text{i.d.}(0, \sigma^2_{\mu_i}), & \sigma^2_{\mu_i} = \mathrm{h}_\mu(W_i\theta_2), & \end{array}$$

  - μ_i and ν_{ij} mutually independent, $\{X_{ij}, Z_{ij}, W_i\}$ strictly exogenous and independently distributed (i.d.) across i
  - $\mathrm{h}_\nu(.)$ and $\mathrm{h}_\mu(.)$ are (strictly) positive twice continuously differentiable functions. Attractive choice (Harvey (1976)): $\mathrm{h}_\nu(.) = \mathrm{h}_\mu(.) = \exp(.)$
  - The total number of observations is $N = \sum_{i=1}^{m} n_i$
  - The missing data generating mechanism is assumed "ignorable"
  - Encompass the previously proposed parametric specifications
  - Provides a way to account for large differences in size and/or varying heterogeneity

- Stacking the n_i observations of each individual i, the model may be written:

$$E\left(Y_i/X_i, Z_i, W_i\right) = X_i\beta \qquad\qquad i = 1, ..., m$$
$$V\left(Y_i/X_i, Z_i, W_i\right) = \operatorname{diag}\left(\mathrm{h}_\nu(Z_i\theta_1)\right) + J_{n_i}\,\mathrm{h}_\mu\left(W_i\theta_2\right)$$

- Stacking again the above vectors and matrices, we obtain the general form:

$$E\left(Y/X, Z, W\right) = X\beta$$
$$V\left(Y/X, Z, W\right) = \operatorname{diag}\left(\mathrm{h}_\nu(Z\theta_1)\right) + D\operatorname{diag}\left(\mathrm{h}_\mu(W\theta_2)\right)D'$$

3. Second order pseudo maximum likelihood estimation and properties

- At first sight, the most natural estimator of the model is:

$$\hat{\beta}_{\text{FGLS}} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}Y$$

where $\hat{\Omega}$ is a consistent $(m \to \infty,\ n_i$ bounded) estimator of Ω

Properties of the FGLS estimator (= gaussian quasi-generalized pseudo maximum likelihood):

- Consistent and efficient if $\hat{\Omega}$ is consistent
- Consistent but inefficient if $\hat{\Omega}$ is inconsistent (but positive definite and $O_p(1)$). In this case, $V\,(\hat{\beta}_{\text{FGLS}})$ must be computed using a heteroscedasticity-consistent covariance matrix estimator

- In the present context, the gaussian pseudo maximum likelihood of order 2 (GPML2) is attractive:

  – From a computational point of view:

    - Given the general form of the variance functions, $\hat{\Omega}$ can not be obtained in a simple way, i.e., in avoiding non-linear optimization. GPML2 also requires non-linear optimization but simultaneously provides mean and variance parameters

  – From a statistical point of view:

    - For the mean parameters, GPML2 has the same properties than FGLS (including robustness to variance misspecification). But it has additional by-product properties for the variance parameters. Among them, under normality, it provides an efficient estimator of the variance parameters

• Gaussian pseudo maximum likelihood of order 2 (GPML2):

$$\begin{aligned}\underset{\varphi\in\Theta}{\text{Max}}\; L(Y/X,Z,W;\varphi) &= \sum_{i=1}^{m}\ln l_i(Y_i/X_i,Z_i,W_i;\varphi)\\ &= -\frac{N}{2}\ln(2\pi)-\frac{1}{2}\sum_{i=1}^{m}\ln|\Omega_i|-\frac{1}{2}\sum_{i=1}^{m}u_i'\Omega_i^{-1}u_i\end{aligned}$$

where $u_i = Y_i - X_i\beta, \qquad \varphi' = (\beta', \theta_1', \theta_2')$

$\Omega_i = \text{diag}\,(\text{h}_\nu(Z_i\theta_1)) + J_{n_i}\, h_\mu\,(W_i\theta_2)$

• Computation of the GPML2 estimator:

– Needed ingredients for “efficient” computation:

- A numerical algorithm: scoring method (rapid, stable) requires (“vec” versus “trace” matrix expressions):

  - Analytical gradient
  - (Analytical hessian: H)
  - Analytical conditionally expected hessian: $E_o\,(H)$

- A sensible set of starting values (inconsistent):

  - $Y_{ij} = D_{ij}\alpha + X_{ij}^*\beta^* + \upsilon_{ij}$
  - $\text{h}_\nu^{-1}(\hat{u}_{ij}^2) = Z_{ij}\theta_1 + \upsilon_{\nu_{ij}}$, $\hat{u}_{ij}^2 = Y_{ij} - D_{ij}\hat{\alpha} + X_{ij}^*\hat{\beta}^*$
  - $\text{h}_\mu^{-1}((\hat{\alpha}_i - \overline{\alpha})^2) = W_i\theta_2 + \upsilon_{\mu_i}$

• Consistency of GPML2:

- RPML2 consistency theorem:

Assume that usual regularity conditions are satisfied and let $\hat{\varphi}_m = \left(\hat{\beta}'_m, \hat{\theta}'_m\right)'$ *be the pseudo maximum likelihood estimator of order 2 obtained from*

$$\operatorname*{Max}_{\varphi \in \Theta} L(Y, X, \varphi) = \sum_{i=1}^{m} \ln l_i \left(Y_i, m_i(X_i, \beta), \Omega_i(X_i, \theta)\right)$$

Then ***sufficient*** *and* ***necessary*** *conditions for* $\hat{\varphi}_m$ *to be consistent for* $\varphi^o = \left(\beta^{o\prime}, \theta^{o\prime}\right)'$ *when both the conditional mean and the conditional variance are correctly specified, i.e.,* $E\left(Y_i/X_i\right) = m_i(X_i, \beta^o)$ *and* $V\left(Y_i/X_i,\right) = \Omega_i(X_i, \theta^o)$*, and to be consistent for* $\varphi^*_m = \left(\beta^{o\prime}, \theta^{*\prime}_m\right)'$ *when the conditional mean is correctly specified but the conditional variance is misspecified, i.e.,* $E\left(Y_i/X_i\right) = m_i(X_i, \beta^o)$ *but* $\nexists\, \theta : \forall\, X_i, V\left(Y_i/X_i\right) = \Omega_i(X_i, \theta)$ *(*θ^*_m *is a pseudo-true value), are:*
(1) *the mean* (β) *and variance* (θ) *parameters vary independently*
(2) $\forall i$, $l_i(Y_i, m_i, \Omega_i)$ *belongs to the restricted quadratic exponential family* $l_i(Y_i, m_i, \Omega_i) = \exp\left(A_i(m_i, \Omega_i) + B_i(Y_i) + C_i(m_i, \Omega_i)'Y_i + Y_i' D_i(\Omega_i) Y_i\right)$, *where* $Y_i \in \mathcal{R}^{n_i}$, $m_i = E\left(Y_i\right)$, $\Omega_i = V\left(Y_i\right)$, $A_i(m_i, \Omega_i)$ *and* $B_i(Y_i)$ *are scalar,* $C_i(m_i, \Omega_i)$ *is a* $(n_i \times 1)$ *vector and* $D_i(\Omega_i)$ *is a* $(n_i \times n_i)$ *matrix*

- Since the normal distribution belongs to the restricted quadratic exponential family, we have for GPML2:

– Under correctly specified conditional mean and variance:

$$\hat{\varphi}_m \xrightarrow{a.s.} \varphi^o, \text{ as } m \to \infty \ (n_i \text{ bounded}), \quad \varphi' = (\beta', \theta'_1, \theta'_2)$$

– Under correctly specified conditional mean but misspecified variance:

$$\hat{\beta}_m \xrightarrow{a.s.} \beta^o \text{ and } \hat{\theta}_m - \theta^*_m \xrightarrow{a.s.} 0, \text{ as } m \to \infty \ (n_i \text{ bounded})$$

where $\theta' = (\theta'_1, \theta'_2)$ and θ^*_m is a pseudo-true value (KLIC interpretation)

• Asymptotic normality of GPML2:

- Standard PML asymptotic normality theorem (e.g. White (1994)):

Assume that usual regularity conditions are satisfied and let $\hat{\varphi}_m$ *be a PML consistent estimator of the pseudo-true value* φ_m^* *obtained from*

$$\operatorname*{Max}_{\varphi \in \Theta} L(Y, X, \varphi) = \sum_{i=1}^{m} L_i(Y_i, X_i, \varphi)$$

then

$$\sqrt{m}\,(\hat{\varphi}_m - \varphi_m^*) \approx N(0, J_m^{*^{-1}} I_m^* J_m^{*^{-1}})$$

where $J_m^* = \frac{1}{m}\sum_{i=1}^m E\left(H_i^*\right)$, $I_m^* = \frac{1}{m}\ V\left(\sum_{i=1}^m s_i^*\right)$, $H_i = \frac{\partial^2 L_i}{\partial\varphi\partial\varphi'}$ *and* $s_i = \frac{\partial L_i}{\partial \varphi}$

- The form of the asymptotic covariance of GPML2 depends on the extent of the misspecification:

1- If the model is correctly specified in its entirety (conditional distribution and conditional moments):

$$\sqrt{m}\,(\hat{\varphi}_m - \varphi^o) \approx N(0, C_m^o), \qquad \varphi' = (\beta', \theta_1', \theta_2')$$

with

$$C_m^o = \begin{bmatrix} C_{m_{\beta\beta}}^o & 0 \\ 0 & C_{m_{\theta\theta}}^o \end{bmatrix} = \begin{bmatrix} I_{m_{\beta\beta}}^{o^{-1}} & 0 \\ 0 & I_{m_{\theta\theta}}^{o^{-1}} \end{bmatrix}, \qquad \theta' = (\theta_1', \theta_2')$$

where $I_m^o = \frac{1}{m}\sum_{i=1}^m E\left(s_i^o s_i^{o\prime}\right) = -J_m^o = -\frac{1}{m}\sum_{i=1}^m E\left(H_i^o\right)$

Consistent covariance matrix estimator ($E_o\left(.\right)$ = conditional expect.):

$$\hat{C}_{m_{\beta\beta}}^o = \left(-\frac{1}{m}\sum_{i=1}^m E_o\left(\hat{H}_{i_{\beta\beta}}\right)\right)^{-1} = \left(\frac{1}{m}\sum_{i=1}^m X_i'\hat{\Omega}_i^{-1}X_i\right)^{-1}$$

$$\hat{C}_{m_{\theta\theta}}^o = \left(-\frac{1}{m}\sum_{i=1}^m E_o\left(\hat{H}_{i_{\theta\theta}}\right)\right)^{-1} = \left(\frac{1}{2m}\sum_{i=1}^m \left(\frac{\partial \operatorname{vec}\hat{\Omega}_i}{\partial\theta'}\right)' \left(\hat{\Omega}_i^{-1}\otimes\hat{\Omega}_i^{-1}\right)\frac{\partial \operatorname{vec}\hat{\Omega}_i}{\partial\theta'}\right)^{-1}$$

2- If the model is only correctly specified for the conditional mean and the conditional variance:

$$\sqrt{m}\,(\hat{\varphi}_m - \varphi^o) \approx N(0, C^o_m), \qquad \varphi' = (\beta', \theta'_1, \theta'_2)$$

with

$$C^o_m = \begin{bmatrix} C^o_{m\beta\beta} & C^o_{m\beta\theta} \\ C^{o\prime}_{m\beta\theta} & C^o_{m\theta\theta} \end{bmatrix} = \begin{bmatrix} I^{o^{-1}}_{m\beta\beta} & J^{o^{-1}}_{m\beta\beta} I^o_{m\beta\theta} J^{o^{-1}}_{m\theta\theta} \\ J^{o^{-1}}_{m\theta\theta} I^o_{m\theta\beta} J^{o^{-1}}_{m\beta\beta} & J^{o^{-1}}_{m\theta\theta} I^o_{m\theta\theta} J^{o^{-1}}_{m\theta\theta} \end{bmatrix}, \ \theta' = (\theta'_1, \theta'_2)$$

where
$$\begin{aligned}
I^o_{m\beta\beta} &= \tfrac{1}{m}\textstyle\sum_{i=1}^m E\left(s^o_{i_\beta} s^{o\prime}_{i_\beta}\right) = -J^o_{m\beta\beta} = -\tfrac{1}{m}\textstyle\sum_{i=1}^m E\left(H^o_{i_{\beta\beta}}\right) \\
J^o_{m\theta\theta} &= \tfrac{1}{m}\textstyle\sum_{i=1}^m E\left(H^o_{i_{\theta\theta}}\right), \quad I^o_{m\beta\theta} = I^{o\prime}_{m\theta\beta} = \tfrac{1}{m}\textstyle\sum_{i=1}^m E\left(s^o_{i_\beta} s^{o\prime}_{i_\theta}\right) \\
I^o_{m\theta\theta} &= \tfrac{1}{m}\textstyle\sum_{i=1}^m E\left(s^o_{i_\theta} s^{o\prime}_{i_\theta}\right)
\end{aligned}$$

Consistent covariance matrix estimator ($E_o\,(.)$ = conditional expect.):

$$\begin{aligned}
\hat{C}^o_{m\beta\beta} &= \left(-\frac{1}{m}\sum_{i=1}^m E_o\left(\hat{H}_{i_{\beta\beta}}\right)\right)^{-1} = \left(\frac{1}{m}\sum_{i=1}^m X'_i \hat{\Omega}_i^{-1} X_i\right)^{-1} \\
\hat{C}^o_{m\beta\theta} &= \left(\frac{1}{m}\sum_{i=1}^m E_o\left(\hat{H}_{i_{\beta\beta}}\right)\right)^{-1} \left(\frac{1}{m}\sum_{i=1}^m \hat{s}_{i_\beta}\hat{s}'_{i_\theta}\right) \left(\frac{1}{m}\sum_{i=1}^m E_o\left(\hat{H}_{i_{\theta\theta}}\right)\right)^{-1} \\
\hat{C}^o_{m\theta\theta} &= \left(\frac{1}{m}\sum_{i=1}^m E_o\left(\hat{H}_{i_{\theta\theta}}\right)\right)^{-1} \left(\frac{1}{m}\sum_{i=1}^m \hat{s}_{i_\theta}\hat{s}'_{i_\theta}\right) \left(\frac{1}{m}\sum_{i=1}^m E_o\left(\hat{H}_{i_{\theta\theta}}\right)\right)^{-1}
\end{aligned}$$

where

$$\begin{aligned}
&E_o\left(\hat{H}_{i_{\beta\beta}}\right) = -X'_i \hat{\Omega}_i^{-1} X_i, \quad \hat{s}_{i_\beta} = X'_i \hat{\Omega}_i^{-1} \hat{u}_i \\
&E_o\left(\hat{H}_{i_{\theta\theta}}\right) = -\frac{1}{2}\left(\frac{\partial \operatorname{vec}\hat{\Omega}_i}{\partial\theta'}\right)' \left(\hat{\Omega}_i^{-1} \otimes \hat{\Omega}_i^{-1}\right) \frac{\partial \operatorname{vec}\hat{\Omega}_i}{\partial\theta'} \\
&\hat{s}_{i_\theta} = \frac{1}{2}\left(\frac{\partial \operatorname{vec}\hat{\Omega}_i}{\partial\theta'}\right)' \left(\hat{\Omega}_i^{-1} \otimes \hat{\Omega}_i^{-1}\right) \operatorname{vec}\left(\hat{u}_i\hat{u}'_i - \hat{\Omega}_i\right)
\end{aligned}$$

3- If the model is correctly specified for the conditional mean but misspecified for the conditional variance:

$$\sqrt{m}\,(\hat{\varphi}_m - \varphi^*_m) \approx N(0, C^*_m), \qquad \varphi^*_m = \left(\beta^{o\prime}, \theta^{*\prime}_{1_m}, \theta^{*\prime}_{2_m}\right)'$$

with

$$C^*_m = \begin{bmatrix} C^*_{m\beta\beta} & C^*_{m\beta\theta} \\ C^{*\prime}_{m\beta\theta} & C^*_{m\theta\theta} \end{bmatrix} = \begin{bmatrix} J^{*^{-1}}_{m\beta\beta} I^{*^{-1}}_{m\beta\beta} J^{*^{-1}}_{m\beta\beta} & J^{*^{-1}}_{m\beta\beta} I^{*}_{m\beta\theta} J^{*^{-1}}_{m\theta\theta} \\ J^{*^{-1}}_{m\theta\theta} I^{*}_{m\theta\beta} J^{*^{-1}}_{m\beta\beta} & J^{*^{-1}}_{m\theta\theta} I^{*}_{m\theta\theta} J^{*^{-1}}_{m\theta\theta} \end{bmatrix}, \theta' = (\theta'_1, \theta'_2)$$

where
$$\begin{aligned} J^*_{m\beta\beta} &= \tfrac{1}{m}\textstyle\sum_{i=1}^m E\,(H^*_{i\beta\beta}), \qquad I^*_{m\beta\beta} = \tfrac{1}{m}\textstyle\sum_{i=1}^m E\,(s^*_{i_\beta} s^{*\prime}_{i_\beta}) \\ J^*_{m\theta\theta} &= \tfrac{1}{m}\textstyle\sum_{i=1}^m E\,(H^*_{i\theta\theta}), \qquad I^*_{m\beta\theta} = I^{*'}_{m\theta\beta} = \tfrac{1}{m}\textstyle\sum_{i=1}^m E\,(s^*_{i_\beta} s^{*\prime}_{i_\theta}) \\ I^*_{m\theta\theta} &= \tfrac{1}{m}\textstyle\sum_{i=1}^m E\,(s^*_{i_\theta} s^{*\prime}_{i_\theta}) - U^*_{m\theta\theta},\; U^*_{m\theta\theta} = \tfrac{1}{m}\textstyle\sum_{i=1}^m E\,(s^*_{i_\theta}) E\,(s^*_{i_\theta})' \end{aligned}$$

Consistent covariance matrix estimator ($E_o\,(.)$ = conditional expect.):

$$\begin{aligned} \hat{C}^*_{m\beta\beta} &= \left(\frac{1}{m}\sum_{i=1}^m E_o\,(\hat{H}_{i_{\beta\beta}})\right)^{-1} \left(\frac{1}{m}\sum_{i=1}^m \hat{s}_{i_\beta}\hat{s}'_{i_\beta}\right) \left(\frac{1}{m}\sum_{i=1}^m E_o\,(\hat{H}_{i_{\beta\beta}})\right)^{-1} \\ &= m\left(\sum_{i=1}^m X'_i\hat{\Omega}_i^{-1}X_i\right)^{-1} \left(\sum_{i=1}^m X'_i\hat{\Omega}_i^{-1}\hat{u}_i\hat{u}'_i\hat{\Omega}_i^{-1}X_i\right) \left(\sum_{i=1}^m X'_i\hat{\Omega}_i^{-1}X_i\right)^{-1} \\ \hat{C}^*_{m\beta\theta} &= \left(\frac{1}{m}\sum_{i=1}^m E_o\,(\hat{H}_{i_{\beta\beta}})\right)^{-1} \left(\frac{1}{m}\sum_{i=1}^m \hat{s}_{i_\beta}\hat{s}'_{i_\theta}\right) \left(\frac{1}{m}\sum_{i=1}^m \hat{H}_{i\theta\theta}\right)^{-1} \\ \hat{Q}^*_{m\theta\theta} &= \left(\frac{1}{m}\sum_{i=1}^m \hat{H}_{i\theta\theta}\right)^{-1} \left(\frac{1}{m}\sum_{i=1}^m \hat{s}_{i_\theta}\hat{s}'_{i_\theta}\right) \left(\frac{1}{m}\sum_{i=1}^m \hat{H}_{i\theta\theta}\right)^{-1} \end{aligned}$$

$$\hat{Q}^*_{m\theta\theta} \xrightarrow{a.s.} J^{*^{-1}}_{m\theta\theta}\left(I^*_{m\theta\theta} + U^*_{m\theta\theta}\right) J^{*^{-1}}_{m\theta\theta} \gg C^*_{m\theta\theta} \text{ (allow conservative test)}$$

where

$$\begin{aligned} \hat{H}_{i\theta\theta} &= -\frac{1}{2}\left(\frac{\partial \operatorname{vec}\hat{\Omega}_i}{\partial\theta'}\right)' \left(\hat{\Omega}_i^{-1}\otimes\hat{\Omega}_i^{-1}\right)\frac{\partial \operatorname{vec}\hat{\Omega}_i}{\partial\theta'} \\ &\quad -\frac{1}{2}\left(\left(\operatorname{vec}\left(\hat{u}_i\hat{u}'_i - \hat{\Omega}_i\right)\right)'\otimes I_{k_\theta}\right)\left(\partial \operatorname{vec}\left(\tfrac{\partial \operatorname{vec}\hat{\Omega}_i^{-1}}{\partial\theta'}\right)'\Big/\,\partial\theta'\right) \end{aligned}$$

4. Specification testing

- Purpose: to provide distribution-free pre and post estimation (mis)specification tests of the conditional variance specification

  - **Joint pseudo Lagrange multiplier test**: distribution-free test statistic which allows to get some insight of the potential relevance of the proposed model before estimating it (by only resorting to OLS residuals)

  - **Information matrix test**: distribution-free test statistic which allows to check the validity of the conditional variance specification of proposed model after having estimated it

4.1. Joint pseudo Lagrange multiplier testing and a BMCP

- We are interested in testing the null of no individual effects and homoscedasticity against the alternative of (possibly heteroscedastic) individual effects and a general form of heteroscedasticity (a set of locally equivalent alternatives) in the usual white noise disturbance

  To do that, standard one-directional tests are not suitable since they are only valid and eventually hold optimal properties when all "other assumptions" are satisfied

  Two basic solutions to "undertesting":

  - To resort to "robust statistics" (e.g. Li-Stengos (1994))

  - To perform joint testing

  The present test statistic, derived in the gaussian PML framework, is a mixture of these two solutions: it is a joint test and it is robust to distributional misspecification

- If the joint test statistic reject the null, how to identify the source(s) of departure from H_0 ?

  One possible answer: to use a Bonferroni Multiple Comparison Procedure

4.1.1. The joint pseudo LM statistic

- Reparametrization of the general model:

$$E\left(Y_i/X_i,Z_i\right) = X_i\beta \qquad\qquad i=1,...,m$$
$$V\left(Y_i/X_i,Z_i\right) = \operatorname{diag}\left(\sigma_\nu^2\,\mathrm{h}(\ddot{Z}_i\ddot{\theta}_1)\right) + \sigma_\mu^2 J_{n_i}$$

$\Rightarrow$ PLM test H_0: $\ddot{\theta}_1 = 0$ and $\sigma_\mu^2 = 0$ $\;(\Leftarrow \varepsilon_{ij} \sim \text{i.d.}(0,\sigma_\nu^2)$ under $\mathrm{H}_0)$

– $\mathrm{h}(.)$ = arbitrary fct. satisfying: $\mathrm{h}(.) > 0$, $\mathrm{h}(0) = 1$ and $\mathrm{h}'(0) \neq 0$

– No variance fct. associated with σ_μ^2 since we are testing $\sigma_\mu^2 = 0$

- General form of the PLM statistic for testing H_0: $R\varphi = 0$ (r constraints, $\varphi' = (\beta', \sigma_\nu^2, \sigma_\mu^2, \ddot{\theta}_1'))$:

$$\frac{1}{m}\left(\sum_{i=1}^m \tilde{s}_i\right)' J_m^{o^{-1}} R' \left(R\,J_m^{o^{-1}} I_m^o J_m^{o^{-1}} R'\right)^{-1} R\,J_m^{o^{-1}} \left(\sum_{i=1}^m \tilde{s}_i\right) \xrightarrow{d} \chi^2\left(r\right)$$

where $\tilde{s}_i$ is the score evaluated at the constrained GPML2 estimator
J_m^o and I_m^o may be replaced by any consistent estimator under H_0

- In the present case (key features: J_m^o is block-diagonal and we are testing only variance parameters), the PLM statistic for H_0: $\ddot{\theta}_1 = 0$ and $\sigma_\mu^2 = 0$ may be written as ($\theta' = (\sigma_\nu^2, \sigma_\mu^2, \ddot{\theta}_1')$):

$$\left(\sum_{i=1}^m \tilde{r}_i'\tilde{B}_i\right)\left(E\,(\sum_{i=1}^m B_i^{o\prime}B_i^o)\right)^{-1} R_\theta'$$
$$\left[R_\theta\left(E\,(\sum_{i=1}^m B_i^{o\prime}B_i^o)\right)^{-1}\left(E\,(\sum_{i=1}^m B_i^{o\prime}r_i^o r_i^{o\prime}B_i^o)\right)\left(E\,(\sum_{i=1}^m B_i^{o\prime}B_i^o)\right)^{-1} R_\theta'\right]^{-1}$$
$$R_\theta\left(E\,(\sum_{i=1}^m B_i^{o\prime}B_i^o)\right)^{-1}\left(\sum_{i=1}^m \tilde{B}_i'\tilde{r}_i\right) \xrightarrow{d} \chi^2\left(1 + k_{\ddot{\theta}_1}\right)$$

where $B_i = \left(\frac{\partial\,\mathrm{vec}\,\Omega_i}{\partial\theta'}\right)$ and $r_i = \mathrm{vec}\left(u_iu_i' - \sigma_\nu^2 I_{n_i}\right)$

- Simplifying the above expression, the PLM statistic turns out to be:

$$\mathcal{PLM}_{I_rH} = \mathcal{PLM}_{I_r} + \mathcal{PLM}_H \xrightarrow{d} \chi^2\left(1 + k_{\ddot{\theta}_1}\right)$$

where

$$\begin{aligned}
\mathcal{PLM}_{I_r} &= \frac{1}{2}\frac{\left(\left(\frac{1}{\tilde{\sigma}_\nu^2}\sum_{i=1}^{m}\left(\tilde{u}_i' e_{n_i}\right)^2\right) - N\right)^2}{\left(\sum_{i=1}^{m} n_i^2\right) - N} \xrightarrow{d} \chi^2(1) \\
\mathcal{PLM}_H &= \left(\tilde{u}^2 - \tilde{\sigma}_\nu^2 e_N\right)' \ddot{Z}\left[\ddot{Z}' M_{e_N}\left(\operatorname{diag}\left(\tilde{u}^4\right) - \tilde{\sigma}_\nu^4 I_N\right) M_{e_N}\ddot{Z}\right]^{-1} \\
&\quad \ddot{Z}'\left(\tilde{u}^2 - \tilde{\sigma}_\nu^2 e_N\right) \xrightarrow{d} \chi^2\left(k_{\ddot{\theta}_1}\right)
\end{aligned}$$

- Remarks:
  - $\mathcal{PLM}_{I_r}$ = the incomplete panel version of the Breush-Pagan (1980) standard LM test for error component derived in Baltagi-Li (1990). The balanced version of the standard LM test was shown to be robust to non-normality by Honda (1985)
  - A statistic asymptotically equivalent to $\mathcal{PLM}_H$, previously proposed by Wooldridge (1990), may be computed as N minus the residual sum of squares ($= NR_u^2$) of the OLS regression ($\overline{\ddot{Z}}_{ij} = [M_{e_N}\ddot{Z}]_{ij}$):

    $$1 = \left[\left(\tilde{u}_{ij}^2 - \tilde{\sigma}_\nu^2\right)\overline{\ddot{Z}_{ij}}\right] b + \text{residuals}, \quad i = 1, ..., m;\ j = 1, ..., n_i$$

  - If constant fourth-order moments are assumed under H_0, i.e., $E_o\left(u_{ij}^4\right) = \delta\ \forall\ i, j$, $\mathcal{PLM}_H$ may be simplified and we obtain the Koenker (1981) statistic:

    $$\mathcal{PLM}_H^K = \frac{1}{\tilde{\delta} - \tilde{\sigma}_\nu^4}\left(\tilde{u}^2 - \tilde{\sigma}_\nu^2 e_N\right)' \ddot{Z}\left[\ddot{Z}' M_{e_N}\ddot{Z}\right]^{-1}\ddot{Z}'\left(\tilde{u}^2 - \tilde{\sigma}_\nu^2 e_N\right)$$

  - If normality is assumed under H_0, i.e., $E_o\left(u_{ij}^4\right) = 3\ \sigma_\nu^4\ \forall\ i, j$, $\mathcal{PLM}_H$ may be further simplified and we obtain the standard Breush-Pagan (1979) statistic:

    $$\mathcal{PLM}_H^{BP} = \frac{1}{2\tilde{\sigma}_\nu^4}\left(\tilde{u}^2 - \tilde{\sigma}_\nu^2 e_N\right)' \ddot{Z}\left[\ddot{Z}' M_{e_N}\ddot{Z}\right]^{-1}\ddot{Z}'\left(\tilde{u}^2 - \tilde{\sigma}_\nu^2 e_N\right)$$

4.1.2. The Bonferroni Multiple Comparison Procedure

- Source(s) of departure from H_0 when it is rejected ?

  $\Rightarrow$ BMCP based on $\mathcal{PLM}_{I_r}$ and $\mathcal{PLM}_H$

- Basic idea of the BMCP:

  To replace a joint test, e.g., H_0: $\theta_1 = 0$ and $\theta_2 = 0$, by an induced test of same size based on a finite number of separate tests, e.g., H_0^1: $\theta_1 = 0$ and H_0^2: $\theta_2 = 0$

  induced test: accept H_0 if and only if H_0^1 and H_0^2 are accepted
  reject H_0 in others cases

  H_0^1 and H_0^2 allow to get some insight about the source(s) of departure from H_0

- Crucial ingredient:

  Determination of the sizes α_i of the separate tests such that the induced test has an overall well-defined size α_I :

  Bonferroni bounds: $\max\{\alpha_1, \alpha_2\} \leq \alpha_I \leq \alpha_1 + \alpha_2$

  If H_0^1 and H_0^2 are independent: $\alpha_I = \alpha_1 + \alpha_2 - \alpha_1\alpha_2$

  $\Rightarrow$ Here, a sensible choice: $\alpha_I/2$

- Problem:

  Under the alternative, one separate statistic may contaminate the other one

  $\Rightarrow$ must be handle with circumspection

4.2. Information matrix testing

- We are interested in testing the null that the conditional variance specification (the conditional mean is implicitly assumed correctly specified) is correct against the alternative that it is not (no precise alternative)

  To do that, we can perfom either:

  - An Hausman-type test on the variance parameters
  - Or an information matrix test on the covariance matrix of the mean parameters, i.e.,

$$E\left(s^o_{i_\beta} s^{o\prime}_{i_\beta} + H^o_{i_{\beta\beta}}\right) = 0, \quad i = 1, ..., m$$

$$\Leftrightarrow E\left(X_i' \Omega_i^{o^{-1}} u_i^o u_i^{o\prime} \Omega_i^{o^{-1}} X_i - X_i' \Omega_i^{o^{-1}} X_i\right) = 0$$

- Information matrix test (implicit null = conditional mean and conditional variance correctly specified), H_0:

$$E\left[im_i^o\right] = E\left[im_i(Y_i, X_i, Z_i, W_i, \beta^o, \theta_1^o, \theta_2^o)\right] = 0, \quad i = 1, ..., m$$

$$\Leftrightarrow E\left[S\left(X_i' \otimes X_i'\right)\left(\Omega_i^{o^{-1}} \otimes \Omega_i^{o^{-1}}\right) \operatorname{vec}\left(u_i^o u_i^{o\prime} - \Omega_i^o\right)\right] = 0$$

  where S is a $q \times k_\beta^2$ selection matrix

  $\Rightarrow$ this is a classical m-test which may be based on the indicator

$$\widehat{IM}_m = \frac{1}{m}\sum_{i=1}^m \widehat{im}_i = \frac{1}{m}\sum_{i=1}^m im_i(Y_i, X_i, Z_i, W_i, \hat\beta, \hat\theta_1, \hat\theta_2)$$

- General form of the m-test statistic for testing H_0: $E\left[im_i^o\right] = 0$ $\left(im_i^o \equiv q \times 1,\ \varphi' = (\beta', \theta_1', \theta_2')\right)$

$$\left(\sum_{i=1}^m \widehat{im}_i\right)' \left[E\left(\sum_{i=1}^m \left(im_i^o - G_m^o J_m^{o^{-1}} s_i^o\right)\left(im_i^o - G_m^o J_m^{o^{-1}} s_i^o\right)'\right)\right]^{-1} \left(\sum_{i=1}^m \widehat{im}_i\right) \xrightarrow{d} \chi^2(q)$$

where $\widehat{im}_i$ is evaluated at the GPML2 estimator, $G_m^o = \frac{1}{m}\sum_{i=1}^m E\left(\frac{\partial\, im_i^o}{\partial \varphi'}\right)$
the term between square brackets may be replaced by any consistent estimator under H_0

- In the present case (key features: $G^o_{m\beta} = 0$ and J^o_m is block-diagonal) the information matrix test statistic of the variance specification may be written as $\left(\theta' = (\theta_1', \theta_2')\right)$:

$$
\begin{aligned}
\mathcal{IM}_\beta \;=\; & \left(\sum_{i=1}^{m} \hat{r}_i' \hat{\Phi}_i^{-1} \hat{A}_i \right) S' \\
& \left[S \left(\sum_{i=1}^{m} \left(\hat{r}_i' \hat{\Phi}_i^{-1} (\hat{A}_i - \hat{B}_i \hat{P}) \right)' \left(\hat{r}_i' \hat{\Phi}_i^{-1} (\hat{A}_i - \hat{B}_i \hat{P}) \right) \right) S' \right]^{-1} \\
& S \left(\sum_{i=1}^{m} \hat{r}_i' \hat{\Phi}_i^{-1} \hat{A}_i \right)' \xrightarrow{d} \chi^2\left(q\right)
\end{aligned}
$$

where $A_i = (X_i \otimes X_i)\,,\ \Phi_i^{-1} = \left(\Omega_i^{-1} \otimes \Omega_i^{-1}\right)\,,\ r_i = \operatorname{vec}\left(u_i u_i' - \Omega_i\right)$

$B_i = \left(\frac{\partial \operatorname{vec} \Omega_i}{\partial \theta'}\right)\,,\ P = \left(\sum_{i=1}^{m} B_i' \hat{\Phi}_i^{-1} B_i\right)^{-1} \left(\sum_{i=1}^{m} B_i' \hat{\Phi}_i^{-1} A_i\right)$

all quantities are evaluated at the GPML2 estimator

- Remark:

  - A statistic $\mathcal{IM}_\beta'$ numerically equal to $\mathcal{IM}_\beta$ may be computed as m minus the residual sum of squares ($= mR_u^2$) of the OLS regression:

$$1 = \left[\hat{r}_i' \hat{\Phi}_i^{-\frac{1}{2}} \hat{\Phi}_i^{-\frac{1}{2}} (\hat{A}_i - \hat{B}_i \hat{P})\, S' \right] b + \text{residuals}, \quad i = 1, ..., m$$

where $\hat{\Phi}_i^{-\frac{1}{2}} (\hat{A}_i - \hat{B}_i \hat{P})$ $\left(\hat{\Phi}_i^{-\frac{1}{2}} = \hat{\Omega}_i^{-\frac{1}{2}} \otimes \hat{\Omega}_i^{-\frac{1}{2}}\right)$ may itself be computed as the matrix residuals ($n_i^2 \times k_\beta^2$) of the OLS multivariate regression:

$$\hat{\Phi}_i^{-\frac{1}{2}} \hat{A}_i = \left[\hat{\Phi}_i^{-\frac{1}{2}} \hat{B}_i \right] P + \text{residuals}, \quad i = 1, ..., m$$

$\mathcal{IM}_\beta'$ may be view as a Wooldridge's modified m-test (Wooldridge (1990, 1991a, 1991b), White (1994)). Thus, contrary to $\mathcal{IM}_\beta$, $\mathcal{IM}_\beta'$ remains a valid statistic (and asymptotically equivalent to $\mathcal{IM}_\beta$) if all quantities are evaluated at any n-root consistent estimator under H_0 rather than at the GPML2 estimator

5. An empirical illustration

- Data from Fecher-Perelman (1989)
  - Inputs-output production records
  - 1286 firms from 14 sectors of the Belgian manufacturing industry
  - Observations over the period 1977-1983 (almost perfectly balanced)
  - The obs. variability prominently lies in the between dimension

- Estimation of sector-specific translog production functions of the form:

$$\begin{aligned}\ln y_{ij} - \overline{\ln y_{ij}} &= \alpha_0 + \alpha_t t + \frac{1}{2}\alpha_{tt}t^2 + \sum_{k=1}^{3} \alpha_{tk} t(\ln x_{ij}^k - \overline{\ln x_{ij}^k}) \\ &+ \sum_{k=1}^{3} \beta_k (\ln x_{ij}^k - \overline{\ln x_{ij}^k}) \\ &+ \frac{1}{2}\sum_{k=1}^{3}\sum_{l=1}^{3} \beta_{kl}(\ln x_{ij}^k - \overline{\ln x_{ij}^k})(\ln x_{ij}^l - \overline{\ln x_{ij}^l}) + \varepsilon_{ij}\end{aligned}$$

  where $\beta_{kl} = \beta_{lk}$, y_{ij} denotes the firm output, t is a trend ($t = 1, 2, ..., 7$) and the 3 inputs x_{ij}^k are capital, labour and raw materials

- Purpose = to show that:

  1- Heteroscedasticity is likely to be a problem in this kind of production model

  2- The proposed model may offer a sensible way to deal with it, at least for efficiency reason, by approximately taking into account the scedastic structure of the data

- Sector-specific PLM test statistics:

Exogenous variables: $Z_{ij}^{*} = (\ln x_{ij}^{k} - \overline{\ln x_{ij}^{k}})$, $k = 1, 2, 3$

$\Rightarrow$ allow variances to change according to both size and input ratios

Table 1: Sector-specific PLM test statistics

Sector	N	m	$\mathcal{PLM}_{I_r H}$ $\chi^2(4)$	$\mathcal{PLM}_{I_r}$ $\chi^2(1)$	$\mathcal{PLM}_{H}$ $\chi^2(3)$
1	327	51	618.48	593.16	25.32
2	161	23	260.57	253.59	6.98*
3	728	105	844.57	817.50	27.07
4	532	76	952.82	926.60	26.22
5	405	58	635.99	590.11	45.88
6	391	56	414.24	409.92	4.32*
7	823	118	1085.19	1074.38	10.81*
8	461	66	718.11	690.92	27.19
9	1559	223	2581.08	2523.15	57.93
10	1091	156	2008.57	1975.51	33.06
11	420	60	781.46	753.97	27.49
12	748	107	1064.57	1047.43	17.14
13	824	118	1756.90	1733.16	23.74
14	480	69	816.33	795.63	20.70

* Not significant at 1%

- Joint PLM Statistic: all rejected with P-values < 0.00001
- Marginal statistic for heteroscedasticity: 11 out of 14 rejected with P-values almost always < 0.0001

- GPML2 estimation for sector 10 (Textile industry):

- Z_{ij} = intercept and $(\ln x_{ij}^{k} - \overline{\ln x_{ij}^{k}})$, $k = 1, 2, 3$
- W_i = intercept and $\frac{1}{n_i} \sum_{j=1}^{n_i} (\ln x_{ij}^{k} - \overline{\ln x_{ij}^{k}})$, $k = 1, 2, 3$
- Std. errors computed assuming misspecified conditional variance

Table 2: Sector 10 (Textile industry) GPML2 estimates

Variable	Parameter	Std. error	t-stat.	P-value
Intercept	-0.20938	0.01580	-13.256	0.0000
Trend	0.04926	0.00561	8.777	0.0000
Trend^2	-0.00358	0.00062	-5.756	0.0000
K x Trend	0.00003	0.00121	0.027	0.9783
L x Trend	0.00583	0.00243	2.401	0.0164
M x Trend	-0.00882	0.00235	-3.754	0.0002
K	0.02105	0.00858	2.452	0.0142
L	0.20494	0.01358	15.090	0.0000
M	0.73676	0.01548	47.602	0.0000
K^2	0.00437	0.00282	1.549	0.1214
L^2	0.05252	0.01093	4.804	0.0000
M^2	0.06138	0.00909	6.754	0.0000
L x K	-0.00295	0.00802	-0.368	0.7130
M x K	0.00203	0.00544	0.372	0.7096
L x M	-0.09520	0.01719	-5.538	0.0000
$\sigma^2_{\nu_{ij}} = \exp(.)$				
Intercept	-5.59141	0.09919	-56.371	0.0000
K	0.16487	0.08128	2.028	0.0425
L	-0.00743	0.11930	-0.062	0.9503
M	-0.44146	0.09096	-4.854	0.0000
$\sigma^2_{\mu_i} = \exp(.)$				
Intercept	-4.09073	0.10683	-38.292	0.0000
$\overline{\mathrm{K}}$	0.24047	0.09857	2.439	0.0147
$\overline{\mathrm{L}}$	-0.27645	0.13238	-2.088	0.0368
$\overline{\mathrm{M}}$	-0.49759	0.09501	-5.237	0.0000

- IM statistic $\mathcal{IM}_\beta$: $\chi^2(1) = 14.2621$, P-value $= 0.00016$
  (indicator = sum of all non-redundant elements of the IM equality)

$\Rightarrow$ The condidional variance is misspecified. However, it can be seen that a heteroscedasticity-like phenomena is present in both μ_i and ν_{ij} (conservative test on the pseudo-true value) and that this heteroscedasticity-like phenomena seems to be related to both inputs ratio and size