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:
    \[
    Y_i = X_i \beta + \varepsilon_i, \quad \varepsilon_i \sim \text{i.d.}(0, \sigma_i^2), \quad i = 1, \ldots, m
    \]

  - Various sources of heteroscedasticity:
    - Differences in size (mechanical source)
    - Heteroscedasticity not directly related to size (heterogeneity)

- Heteroscedasticity in cross-section ⇒ heteroscedasticity in panel data

  - Standard one-way error components model:
    \[
    Y_{ij} = X_{ij} \beta + \varepsilon_{ij}, \quad \varepsilon_{ij} = \mu_i + \nu_{ij}, \quad \nu_{ij} \sim \text{i.d.}(0, \sigma_{\nu}^2), \quad \mu_i \sim \text{i.d.}(0, \sigma_{\mu}^2), \quad i = 1, \ldots, m; \quad j = 1, 2, \ldots, t
    \]

  - 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):

\[ \nu_{ij} \sim \text{i.d.}(0, \sigma_{\nu}^2) \quad \text{and} \quad \mu_i \sim \text{i.d.}(0, \sigma_{\mu}^2) \]


\[ \nu_{ij} \sim \text{i.d.}(0, \sigma_{\nu_{ij}}^2) \quad \text{and} \quad \mu_i \sim \text{i.d.}(0, \sigma_{\mu_i}^2) \]

c- Li-Stengos (1994):

\[ \nu_{i,j} \sim \text{i.d.}(0, \sigma_{\nu_{ij}}^2) \quad \text{and} \quad \mu_i \sim \text{i.d.}(0, \sigma_{\mu_i}^2) \]

where \( \sigma_{\nu_{ij}}^2 = \text{a non-parametric function } f(Z_{ij}) \)

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

\[ \nu_{ij} \sim \text{i.d.}(0, \sigma_{\nu_{ij}}^2) \quad \text{and} \quad \mu_i \sim \text{i.d.}(0, \sigma_{\mu_i}^2) \]

where \( \sigma_{\nu_i}^2 = \sigma_{\nu}^2 f(Z_i \theta) \) and \( \sigma_{\mu_i}^2 = \sigma_{\mu}^2 f(Z_i \theta) \)

• 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:

\[ Y_{ij} = X_{ij}\beta + \varepsilon_{ij}, \quad \varepsilon_{ij} = \mu_i + \nu_{ij}, \]
\[ \nu_{ij} \sim \text{i.d.}(0, \sigma_{\nu_{ij}}^2), \quad \sigma_{\nu_{ij}}^2 = h_\nu(Z_{ij}\theta_1), \quad i = 1, \ldots, m; \]
\[ \mu_i \sim \text{i.d.}(0, \sigma_{\mu_i}^2), \quad \sigma_{\mu_i}^2 = h_\mu(W_i\theta_2), \]

- \( \mu_i \) and \( \nu_{ij} \) mutually independent, \( \{X_{ij}, Z_{ij}, W_i\} \) strictly exogenous and independently distributed (i.d.) across \( i \)
- \( h_\nu(.) \) and \( h_\mu(.) \) are (strictly) positive twice continuously differentiable functions. Attractive choice (Harvey (1976)): \( h_\nu(.) = h_\mu(.) = \exp(.) \)
- 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:

\[ E(Y_i/X_i, Z_i, W_i) = X_i\beta \quad \quad i = 1, \ldots, m \]
\[ V(Y_i/X_i, Z_i, W_i) = \text{diag}(h_\nu(Z_i\theta_1)) + J_{n_i} h_\mu(W_i\theta_2) \]

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

\[ E(Y/X, Z, W) = X\beta \]
\[ V(Y/X, Z, W) = \text{diag}(h_\nu(Z\theta_1)) + D \text{diag}(h_\mu(W\theta_2)) D' \]
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 \( \Omega \)

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):

\[
\operatorname{Max}_{\varphi \in \Theta} L(Y/X, Z, W; \varphi) = \sum_{i=1}^{m} \ln l_i(Y_i/X_i, Z_i, W_i; \varphi) = -\frac{N}{2} \ln (2\pi) - \frac{1}{2} \sum_{i=1}^{m} \ln |\Omega_i| - \frac{1}{2} \sum_{i=1}^{m} u_i' \Omega_i^{-1} u_i
\]

where \( u_i = Y_i - X_i \beta \), \( \varphi' = (\beta', \theta'_1, \theta'_2) \)

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

• 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):
    - \( Y_{ij} = D_{ij} \alpha + X_{ij}^s \beta^s + v_{ij} \)
    - \( h_\nu^{-1}(\hat{u}_{ij}^2) = Z_{ij} \theta_1 + v_{nij}, \hat{u}_{ij}^2 = Y_{ij} - D_{ij} \hat{\alpha} + X_{ij}^s \hat{\beta}^s \)
    - \( h_\mu^{-1}((\hat{\alpha}_i - \bar{\alpha})^2) = W_i \theta_2 + v_{\mu_i} \)
• 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

\[
\max_{\varphi \in \Theta} L(Y, X, \varphi) = \sum_{i=1}^{m} \ln l_i (Y_i, m_i(X_i, \beta), \Omega_i(X_i, \theta))
\]

Then sufficient and necessary conditions for \( \hat{\varphi}_m \) to be consistent for \( \varphi^o = (\beta^o, \theta^o)' \) when both the conditional mean and the conditional variance are correctly specified, i.e., \( E(Y_i/X_i) = m_i(X_i, \beta^o) \) and \( V(Y_i/X_i) = \Omega_i(X_i, \theta^o) \), and to be consistent for \( \varphi^*_m = (\beta^o, \theta^o^*)' \) when the conditional mean is correctly specified but the conditional variance is misspecified, i.e., \( E(Y_i/X_i) = m_i(X_i, \beta^o) \) but \( \nexists \theta : \forall X_i, V(Y_i/X_i) = \Omega_i(X_i, \theta) \) (\( \theta^o_m \) is a pseudo-true value), are:

1. the mean (\( \beta \)) and variance (\( \theta \)) 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 \( Y_i \in \mathcal{R}^{n_i}, m_i = E(Y_i), \Omega_i = V(Y_i), A_i(m_i, \Omega_i) \) 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:
    \[
    \hat{\beta}_m \xrightarrow{a.s.} \beta^o \text{ and } \hat{\theta}_m - \theta^o_m \xrightarrow{a.s.} 0, \text{ as } m \to \infty (n_i \text{ bounded})
    \]

where \( \theta' = (\theta'_1, \theta'_2) \) and \( \theta^o_m \) 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 \( \varphi^*_m \) obtained from

\[
\Max_{\varphi \in \Theta} L(Y, X, \varphi) = \sum_{i=1}^{m} L_i(Y_i, X_i, \varphi)
\]

then

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

where \( J_m = \frac{1}{m} \sum_{i=1}^{m} E(H_i^2) \), \( I_m = \frac{1}{m} V(\sum_{i=1}^{m} s_i^2) \), \( H_i = \frac{\partial L_i}{\partial \varphi} \) 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):

\[
\sqrt{m} (\hat{\varphi}_m - \varphi^o) \approx N(0, C_m^o), \quad \varphi^o = (\beta^o, \theta_1^o, \theta_2^o)
\]

with

\[
C_m^o = \begin{bmatrix}
C_{m,33}^o & 0 \\
0 & C_{m,00}^o
\end{bmatrix} = \begin{bmatrix}
I_{m,33}^{-1} & 0 \\
0 & I_{m,00}^{-1}
\end{bmatrix}, \quad \theta^o = (\theta_1^o, \theta_2^o)
\]

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

Consistent covariance matrix estimator \( (E_o(\cdot) = \text{conditional expect.}):\)

\[
\hat{C}_{m,33}^o = \left( -\frac{1}{m} \sum_{i=1}^{m} E_o(H_{i,33}) \right)^{-1} = \left( \frac{1}{m} \sum_{i=1}^{m} X_i' \hat{\Omega}_i^{-1} X_i \right)^{-1}
\]

\[
\hat{C}_{m,00}^o = \left( -\frac{1}{m} \sum_{i=1}^{m} E_o(H_{i,00}) \right)^{-1} = \left( \frac{1}{2m} \sum_{i=1}^{m} \left( \frac{\partial \text{vec } \hat{\Omega}_i}{\partial \theta} \right)' \left( \hat{\Omega}_i^{-1} \otimes \hat{\Omega}_i^{-1} \right) \frac{\partial \text{vec } \hat{\Omega}_i}{\partial \theta} \right)^{-1}
\]
2- If the model is only correctly specified for the conditional mean and the conditional variance:

\[ \sqrt{m}(\varphi_m - \varphi^o) \approx N(0, C_m) \]

with

\[ C_m = \begin{bmatrix}
  C_{m_{\beta \beta}} & C_{m_{\beta \theta}} \\
  C_{m_{\beta \theta}} & C_{m_{\theta \theta}}
\end{bmatrix} = \begin{bmatrix}
  I_{m_{\beta \beta}}^{-1} & J_{m_{\beta \theta}}^{-1} & J_{m_{\beta \theta}}^{-1} \\
  J_{m_{\theta \theta}}^{-1} & I_{m_{\theta \theta}}^{-1} & J_{m_{\theta \theta}}^{-1}
\end{bmatrix} , \quad \theta' = (\theta_1', \theta_2')
\]

where

\[ I_{m_{\beta \beta}} = \frac{1}{m} \sum_{i=1}^{m} E_o(\hat{H}_{i_{\beta \beta}}) = -\frac{1}{m} \sum_{i=1}^{m} E(H_{i_{\beta \beta}}) \]

\[ J_{m_{\beta \theta}} = \frac{1}{m} \sum_{i=1}^{m} E(\hat{s}_{i_{\beta}} \hat{s}_{i_{\theta}}') \]

Consistent covariance matrix estimator \((E_o(.) = \text{conditional expect.}):\)

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

\[ \hat{C}_{m_{\beta \theta}} = \left( \frac{1}{m} \sum_{i=1}^{m} E_o(\hat{H}_{i_{\beta \theta}}) \right)^{-1} \left( \frac{1}{m} \sum_{i=1}^{m} \hat{s}_{i_{\beta}} \hat{s}_{i_{\theta}}' \right) \left( \frac{1}{m} \sum_{i=1}^{m} E_o(\hat{H}_{i_{\theta \theta}}) \right)^{-1} \]

\[ \hat{C}_{m_{\theta \theta}} = \left( \frac{1}{m} \sum_{i=1}^{m} E_o(\hat{H}_{i_{\theta \theta}}) \right)^{-1} \left( \frac{1}{m} \sum_{i=1}^{m} \hat{s}_{i_{\theta}} \hat{s}_{i_{\theta}}' \right) \left( \frac{1}{m} \sum_{i=1}^{m} E_o(\hat{H}_{i_{\theta \theta}}) \right)^{-1} \]

where

\[ E_o(\hat{H}_{i_{\beta \beta}}) = -X_i' \hat{Q}_i^{-1} X_i , \quad \hat{s}_{i_{\beta}} = X_i' \hat{Q}_i^{-1} \hat{u}_i \]

\[ E_o(\hat{H}_{i_{\theta \theta}}) = -\frac{1}{2} \left( \frac{\partial \text{vec} \hat{\Omega}_i}{\partial \theta'} \right)' \left( \hat{\Omega}_i^{-1} \otimes \hat{\Omega}_i^{-1} \right) \frac{\partial \text{vec} \hat{\Omega}_i}{\partial \theta'} \]

\[ \hat{s}_{i_{\theta}} = \frac{1}{2} \left( \frac{\partial \text{vec} \hat{\Omega}_i}{\partial \theta'} \right)' \left( \hat{\Omega}_i^{-1} \otimes \hat{\Omega}_i^{-1} \right) \text{vec} \left( \hat{u}_i \hat{u}_i' - \hat{\Omega}_i \right) \]
3- If the model is correctly specified for the conditional mean but mis-
specified for the conditional variance:

\[ \sqrt{m} (\hat{\varphi}_m - \varphi^*_m) \approx N(0, C_m^s), \quad \varphi^*_m = (\beta^\prime, \theta^\prime_1, \theta^\prime_2) \]

with

\[ C_m^s = \begin{bmatrix} C_{m_{33}}^s & C_{m_{30}}^s \\ C_{m_{30}}^s & C_{m_{00}}^s \end{bmatrix} = \begin{bmatrix} J_{m_{33}}^{-1} I_{m_{33}}^{-1} J_{m_{33}}^{-1} & J_{m_{33}}^{-1} I_{m_{30}}^{-1} J_{m_{30}}^{-1} \\ J_{m_{30}}^{-1} I_{m_{30}}^{-1} J_{m_{33}}^{-1} & J_{m_{30}}^{-1} I_{m_{00}}^{-1} J_{m_{00}}^{-1} \end{bmatrix}, \theta' = (\theta_1', \theta_2') \]

where

\[ J_{m_{33}}^s = \frac{1}{m} \sum_{i=1}^m E (H_{i33}^s), \quad I_{m_{33}}^s = \frac{1}{m} \sum_{i=1}^m E (s_{i3}^s s_{i3}'^s) \]
\[ J_{m_{30}}^s = \frac{1}{m} \sum_{i=1}^m E (H_{i30}^s), \quad I_{m_{30}}^s = I_{m_{03}}^s = \frac{1}{m} \sum_{i=1}^m E (s_{i3}^s s_{i0}'^s) \]
\[ I_{m_{00}}^s = \frac{1}{m} \sum_{i=1}^m E (s_{i0}^s s_{i0}'^s) - U_{m_{00}}^s, \quad U_{m_{00}}^s = \frac{1}{m} \sum_{i=1}^m E (s_{i0}^s E (s_{i0}^s)') \]

Consistent covariance matrix estimator \((E_o (.) = \text{conditional expect.})\):

\[ \hat{C}_{m_{33}}^s = \left( \frac{1}{m} \sum_{i=1}^m E_o (\hat{H}_{i33}) \right)^{-1} \left( \frac{1}{m} \sum_{i=1}^m \hat{s}_{i3} \hat{s}_{i3}' \right) \left( \frac{1}{m} \sum_{i=1}^m E_o (\hat{H}_{i33}) \right)^{-1} \]
\[ = m \left( \sum_{i=1}^m X_i' \hat{\Omega}_{i}^{-1} X_i \right)^{-1} \left( \sum_{i=1}^m X_i' \hat{\Omega}_{i}^{-1} \hat{u}_i \hat{u}_i' \hat{\Omega}_{i}^{-1} X_i \right) \left( \sum_{i=1}^m X_i' \hat{\Omega}_{i}^{-1} X_i \right)^{-1} \]

\[ \hat{C}_{m_{30}}^s = \left( \frac{1}{m} \sum_{i=1}^m E_o (\hat{H}_{i30}) \right)^{-1} \left( \frac{1}{m} \sum_{i=1}^m \hat{s}_{i3} \hat{s}_{i0}' \right) \left( \frac{1}{m} \sum_{i=1}^m \hat{H}_{i0} \right)^{-1} \]
\[ \hat{Q}_{m_{00}}^s = \left( \frac{1}{m} \sum_{i=1}^m \hat{H}_{i0} \right)^{-1} \left( \frac{1}{m} \sum_{i=1}^m \hat{s}_{i0} \hat{s}_{i0}' \right) \left( \frac{1}{m} \sum_{i=1}^m \hat{H}_{i0} \right)^{-1} \]

\[ \hat{Q}_{m_{00}}^a \stackrel{a.s.}{\rightarrow} J_{m_{00}}^{-1} (I_{m_{00}} + U_{m_{00}}^s) J_{m_{00}}^{-1} \Rightarrow C_{m_{00}}^s \text{ (allow conservative test)} \]

where

\[ \hat{H}_{i00} = -\frac{1}{2} \left( \frac{\partial \vec{\hat{\Omega}}_i}{\partial \theta'} \right)' \left( \hat{\Omega}_i^{-1} \otimes \hat{\Omega}_i^{-1} \right) \frac{\partial \vec{\hat{\Omega}}_i}{\partial \theta'} \]
\[ -\frac{1}{2} \left( \left( \vec{\hat{u}}_i \vec{\hat{u}}_i' - \hat{\Omega}_i \right)' \otimes I_0 \right) \left( \partial \vec{\hat{\Omega}}_i^{-1} / \partial \theta' \right) \]
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:

\[
E(Y_i/X_i, Z_i) = X_i \beta \quad i = 1, \ldots, m
\]

\[
V(Y_i/X_i, Z_i) = \text{diag} \left( \sigma_\nu^2 h(\tilde{Z}_i \tilde{\theta}_1) \right) + \sigma_\mu^2 J_{n_i}
\]

⇒ PLM test \( H_0 \): \( \tilde{\theta}_1 = 0 \) and \( \sigma_\mu^2 = 0 \) \( \iff \varepsilon_{ij} \sim \text{i.d.}(0, \sigma_\nu^2) \) under \( H_0 \)

- \( h(.) = \) arbitrary fct. satisfying: \( h(.) > 0, h(0) = 1 \) and \( h'(0) \neq 0 \)
- No variance fct. associated with \( \sigma_\mu^2 \) since we are testing \( \sigma_\mu^2 = 0 \)

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

\[
\frac{1}{m} \left( \sum_{i=1}^{m} \tilde{s}_i \right)^' J_m^{o-1} R' \left( R J_m^{o-1} I_m J_m^{o-1} R' \right)^{-1} R J_m^{o-1} \left( \sum_{i=1}^{m} \tilde{s}_i \right) \to \chi^2(r)
\]

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 \( H_0 \): \( \tilde{\theta}_1 = 0 \) and \( \sigma_\mu^2 = 0 \) may be written as \( (\theta' = (\sigma_\nu^2, \sigma_\mu^2, \tilde{\theta}_1')) \):

\[
\left[ R_{\tilde{\theta}} \left( E \left( \sum_{i=1}^{m} B_i^o B_i^o \right) \right)^{-1} \left( \sum_{i=1}^{m} \tilde{r}_i ^' \tilde{B}_i \right) \right] \to \chi^2 \left( 1 + k_{\tilde{\theta}_1} \right)
\]

where \( B_i = \left( \frac{\partial \text{vec} \Omega}{\partial \tilde{\theta}} \right) \) and \( r_i = \text{vec} \left( u_i u_i' - \sigma_\nu^2 I_{n_i} \right) \)
• Simplifying the above expression, the PLM statistic turns out to be:

\[ \mathcal{PLM}_{I,H} = \mathcal{PLM}_{I,r} + \mathcal{PLM}_{H} \xrightarrow{d} \chi^2 (1 + k\theta_1) \]

where

\[ \mathcal{PLM}_{I,r} = \frac{1}{2} \left( \left( \frac{1}{\hat{\sigma}^2} \sum_{i=1}^{m} (\tilde{u}_i' e_{ni})^2 - N \right) \right)^2 \left( \sum_{i=1}^{m} n_i^2 \right) - N \xrightarrow{d} \chi^2 (1) \]

\[ \mathcal{PLM}_{H} = (\tilde{u}^2 - \tilde{\sigma}^2 e_N)' \tilde{Z} \left[ \tilde{Z}' M_{eN} (\text{diag}(\tilde{u}^4) - \tilde{\sigma}^4 I_N) M_{eN} \tilde{Z} \right]^{-1} \tilde{Z}' (\tilde{u}^2 - \tilde{\sigma}^2 e_N) \xrightarrow{d} \chi^2 (k\theta_1) \]

• 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 (\( = NR^2_u \)) of the OLS regression \( \bar{Z}_{ij} = [M_{eN} \tilde{Z}]_{ij} \):

\[ 1 = \left( (\tilde{u}_{ij}^2 - \tilde{\sigma}^2 e_N) \bar{Z}_{ij} \right) b + \text{residuals}, \quad i = 1, \ldots, m; \quad j = 1, \ldots, n_i \]

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

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

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

\[ \mathcal{PLM}_{H}^{BP} = \frac{1}{2\hat{\sigma}^4} (\tilde{u}^2 - \tilde{\sigma}^2 e_N)' \tilde{Z} \left[ \tilde{Z}' M_{eN} \tilde{Z} \right]^{-1} \tilde{Z}' (\tilde{u}^2 - \tilde{\sigma}^2 e_N) \]
4.1.2. The Bonferroni Multiple Comparison Procedure

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

  \[ \Rightarrow \text{BMCP based on } \mathcal{P} \mathcal{L} \mathcal{M}_I \text{ and } \mathcal{P} \mathcal{L} \mathcal{M}_H \]

- 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 \( \alpha_i \) of the separate tests such that the induced test has an overall well-defined size \( \alpha_I \):

  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: \( \alpha_I = \alpha_1 + \alpha_2 - \alpha_1 \alpha_2 \)

  \[ \Rightarrow \text{Here, a sensible choice: } \alpha_I/2 \]

- Problem:

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

  \[ \Rightarrow \text{must be handle with circumspection} \]
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 perform either:
- An Hausman-type test on the variance parameters
- Or an information matrix test on the covariance matrix of the mean parameters, i.e.,

\[
E (s_{i \beta} s_{i \beta} + H_{i \beta \beta \beta}) = 0, \quad i = 1, ..., m
\]

\[\iff E (X_i' \Omega_i \Big| u_i u_i') X_i - X_i' \Omega_i \Big| u_i u_i') X_i = 0\]

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

\[E \left[ \text{im}_i \right] = 0 \quad \text{im}_i = \begin{pmatrix} \beta' \\ \theta_1' \\ \theta_2' \end{pmatrix}, \quad (im_i \equiv q \times 1, \phi' = (\beta', \theta_1', \theta_2')) \]

\[\iff E \left[ S (X_i' \otimes X_i') (\Omega_i^{-1} \otimes \Omega_i^{-1}) \text{vec} (u_i u_i' - \Omega_i) \right] = 0 \]

where \(S\) is a \(q \times k \beta^2\) selection matrix

\[\therefore\] this is a classical \(m\)-test which may be based on the indicator

\[\hat{IM}_m = \frac{1}{m} \sum_{i=1}^{m} \text{im}_i = \frac{1}{m} \sum_{i=1}^{m} \text{im}_i (Y_i, X_i, Z_i, W_i, \hat{\beta}, \hat{\theta}_1, \hat{\theta}_2)\]

- General form of the \(m\)-test statistic for testing \(H_0\): \(E \left[ im_i^0 \right] = 0 \)

\[\left( \sum_{i=1}^{m} \text{im}_i \right)' \left[ E \left( \sum_{i=1}^{m} \left( im_i^0 - G_m^0 J_m^{0^{-1}} s_i^0 \right) \left( im_i^0 - G_m^0 J_m^{0^{-1}} s_i^0 \right)' \right]^{-1} \right] \left( \sum_{i=1}^{m} \text{im}_i \right) \xrightarrow{d} \chi^2(q) \]

where \(\text{im}_i\) is evaluated at the GPML2 estimator, \(G_m^0 = \frac{1}{m} \sum_{i=1}^{m} E \left( \frac{\partial im_i^0}{\partial \phi} \right)\)

the term between square brackets may be replaced by any consistent estimator under \(H_0\)
In the present case (key features: $G_{m\beta}^o = 0$ and $J_m^o$ is block-diagonal) the information matrix test statistic of the variance specification may be written as $(\theta' = (\theta'_1, \theta'_2))$:

$$\mathcal{IM}_\beta = \left( \sum_{i=1}^{m} \hat{r}_i \hat{\Phi}_i^{-1} \hat{A}_i \right) S'$$

$$= \left[ S \left( \sum_{i=1}^{m} \left( \hat{r}_i \hat{\Phi}_i^{-1} \left( \hat{A}_i - \hat{B}_i \hat{P} \right) \right)' \left( \hat{r}_i \hat{\Phi}_i^{-1} \left( \hat{A}_i - \hat{B}_i \hat{P} \right) \right) \right) S' \right]^{-1}$$

$$S \left( \sum_{i=1}^{m} \hat{r}_i \hat{\Phi}_i^{-1} \hat{A}_i \right)' \xrightarrow{d} \chi^2(q)$$

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

$$B_i = \left( \frac{\partial \text{vec} \Omega_i}{\partial \theta} \right), \quad P = \left( \sum_{i=1}^{m} \hat{B}_i \hat{\Phi}_i^{-1} \hat{B}_i \right)^{-1} \left( \sum_{i=1}^{m} \hat{B}_i \hat{\Phi}_i^{-1} \hat{A}_i \right)$$

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 ($= m R_u^2$) of the OLS regression:

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

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 ($n_i^2 \times k_{i\beta}^2$) 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:

\[
\ln y_{ij} - \ln y_{ij} = \alpha_0 + \alpha_t t + \frac{1}{2} \alpha_{tt} t^2 + \sum_{k=1}^{3} \alpha_{tk} t (\ln x_{ij}^k - \ln x_{ij}^k) \\
+ \sum_{k=1}^{3} \beta_k (\ln x_{ij}^k - \ln x_{ij}^k) \\
+ \frac{1}{2} \sum_{k=1}^{3} \sum_{l=1}^{3} \beta_{kl} (\ln x_{ij}^k - \ln x_{ij}^k)(\ln x_{ij}^l - \ln x_{ij}^l) + \varepsilon_{ij}
\]

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: \( Z_{ij}^k = (\ln x_{ij}^k - \ln \bar{x}_{ij}^k), \ k = 1, 2, 3 \)

⇒ allow variances to change according to both size and input ratios

Table 1: Sector-specific PLM test statistics

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

* 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 \( (\ln x_{ij}^k - \ln \bar{x}_{ij}^k), \ k = 1, 2, 3 \)
- \( W_i = \) intercept and \( \frac{1}{n_i} \sum_{j=1}^{n_i} (\ln x_{ij}^k - \ln \bar{x}_{ij}^k), \ k = 1, 2, 3 \)
- Std. errors computed assuming misspecified conditional variance
Table 2: Sector 10 (Textile industry) GPML2 estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>Std. error</th>
<th>t-stat.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.20938</td>
<td>0.01580</td>
<td>-13.256</td>
<td>0.0000</td>
</tr>
<tr>
<td>Trend</td>
<td>0.04926</td>
<td>0.00561</td>
<td>8.777</td>
<td>0.0000</td>
</tr>
<tr>
<td>Trend²</td>
<td>-0.00358</td>
<td>0.00062</td>
<td>-5.756</td>
<td>0.0000</td>
</tr>
<tr>
<td>K x Trend</td>
<td>0.00003</td>
<td>0.00121</td>
<td>0.027</td>
<td>0.9783</td>
</tr>
<tr>
<td>L x Trend</td>
<td>0.00583</td>
<td>0.00243</td>
<td>2.401</td>
<td>0.0164</td>
</tr>
<tr>
<td>M x Trend</td>
<td>-0.00882</td>
<td>0.00235</td>
<td>-3.754</td>
<td>0.0002</td>
</tr>
<tr>
<td>K</td>
<td>0.02105</td>
<td>0.00858</td>
<td>2.452</td>
<td>0.0142</td>
</tr>
<tr>
<td>L</td>
<td>0.20494</td>
<td>0.01358</td>
<td>15.090</td>
<td>0.0000</td>
</tr>
<tr>
<td>M</td>
<td>0.73676</td>
<td>0.01548</td>
<td>47.602</td>
<td>0.0000</td>
</tr>
<tr>
<td>K²</td>
<td>0.00437</td>
<td>0.00282</td>
<td>1.549</td>
<td>0.1214</td>
</tr>
<tr>
<td>L²</td>
<td>0.05252</td>
<td>0.01093</td>
<td>4.804</td>
<td>0.0000</td>
</tr>
<tr>
<td>M²</td>
<td>0.06138</td>
<td>0.00909</td>
<td>6.754</td>
<td>0.0000</td>
</tr>
<tr>
<td>L x K</td>
<td>-0.00295</td>
<td>0.00802</td>
<td>-0.368</td>
<td>0.7130</td>
</tr>
<tr>
<td>M x K</td>
<td>0.00203</td>
<td>0.00544</td>
<td>0.372</td>
<td>0.7096</td>
</tr>
<tr>
<td>L x M</td>
<td>-0.09520</td>
<td>0.01719</td>
<td>-5.538</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

\[ \sigma^2_{\nu_{ij}} = \exp(,) \]

- \( \text{Intercept} \) -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

\[ \sigma^2_{\mu_i} = \exp(,) \]

- \( \text{Intercept} \) -4.09073 0.10683 -38.292 0.0000
- \( \bar{K} \) 0.24047 0.09857 2.439 0.0147
- \( \bar{L} \) -0.27645 0.13238 -2.088 0.0368
- \( \bar{M} \) -0.49759 0.09501 -5.237 0.0000

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

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