A full heteroscedastic one-way error components model: pseudo-maximum likelihood estimation and specification testing

Bernard Lejeune

HEC-University of Liège, CORE and ERUDITE

Boulevard du Rectorat, 3, B33
4000 Liège Belgium
E-mail: B.Lejeune@ulg.ac.be

May, 2005 (Final version)

Abstract

This paper proposes an extension of the standard one-way error components model allowing for heteroscedasticity in both the individual-specific and the general error terms, as well as for unbalanced panel. On the grounds of its computational convenience, its potential efficiency, its robustness to non-normality and its robustness to possible misspecification of the assumed scedastic structure of the data, we argue for estimating this model by Gaussian pseudo-maximum likelihood of order two. Further, we review how, taking advantage of the powerful m-testing framework, the correct specification of the prominent aspects of the model may be tested. We survey potentially useful nested, non-nested, Hausman and information matrix type diagnostic tests of both the mean and the variance specification of the model. Finally, we illustrate the usefulness of our proposed model and estimation and diagnostic testing procedures through an empirical example.

Keywords: Error components model, Heteroscedasticity, Unbalanced panel data, Pseudo-maximum likelihood estimation, m-testing.
J.E.L. classification: C12, C22, C52.

Acknowledgements

I wish to thank Laurence Broze, Patrick Sevestre, Alain Trognon, Jean-Pierre Urbain, James Lindsey, Badi Baltagi, Tim Coelli, Muriel Fonder, Arie Preminger and two anonymous referees for useful comments and suggestions on earlier versions of this paper. Errors, of course, remain my sole responsability.

1. Introduction

As largely acknowledged, heteroscedasticity is endemic when working with microeconomic cross-section data. One of its common sources is differences in size (the level of the variables) across individuals. This kind of heteroscedasticity is mechanical. It is simply a consequence of the additive disturbance structure of the classical regression model. It is generally tackled by performing a logarithmic transformation of the dependent variable. However, even after accounting in such a way for differences in size, numerous cases remain where we cannot expect the error variance to be constant. On one hand, there is a priori no reason to believe that the logarithmic specification postulating similar percentage variations across observations is relevant. In the production field for example, observations for lower outputs firms seem likely to evoke larger variances (see Baltagi and Griffin (1988)). On the other hand, the error variance may also vary across observations of similar size. For example, the variance of firms outputs might depend upon their capital intensity.

Obviously, there is no reason to expect the heteroscedasticity problems associated with microeconomic panel data to be markedly different from those encountered in work with cross-section data. Nonetheless, the issue of heteroscedasticity has received somewhat limited attention in the literature related to panel data error components models.

Seemingly, the first authors who dealt with the problem were Mazodier and Trognon (1978). Subsequent contributions include Verbon (1980), Rao, Kaplan and Cochran (1981), Magnus (1982), Baltagi (1988), Baltagi and Griffin (1988), Randolph (1988), Wansbeek (1989), Li and Stengos (1994), Holly and Gardiol (2000), Roy (2002), Phillips (2003), Baltagi, Bresson and Pirotte (2004) and Lejeune (2004).

Within the framework of the classical one-way error components regression model, the issues considered by these papers can be summarized as follows. Both Mazodier and Trognon (1978) and Baltagi and Griffin (1988) are concerned with estimating a model allowing for changing variances of the individual-specific error term across individuals, i.e. they assume that we may write the composite error as $\varepsilon_{it} = \mu_i + \nu_{it}$, $\nu_{it} \sim (0, \sigma^2_\nu)$ while $\mu_i \sim (0, \sigma^2_{\mu_i})$. Phillips (2003) considers a similar model where heteroscedasticity occurs only through individual-specific variances changing across strata of individuals. Rao, Kaplan and Cochran (1981), Magnus (1982), Baltagi (1988) and Wansbeek (1989) adopt a different specification, allowing for changing variances of the general error term across individuals, i.e. assume that $\nu_{it} \sim (0, \sigma^2_{\nu_i})$ while $\mu_i \sim (0, \sigma^2_\mu)$. Verbon (1980) is interested in Lagrange Multiplier (LM) testing of the standard normally distributed homoscedastic one-way error components model against the heteroscedastic alternative $\nu_{it} \sim N(0, \sigma^2_{\nu_i})$ and $\mu_i \sim N(0, \sigma^2_{\mu_i})$, where $\sigma^2_{\nu_i}$ and $\sigma^2_{\mu_i}$ are, up to a multiplicative constant, identical parametric functions of a vector of time-invariant explanatory variables Z_i, i.e. $\sigma^2_{\nu_i} = \sigma^2_\nu \phi(Z_i\gamma)$ and $\sigma^2_{\mu_i} = \sigma^2_\mu \phi(Z_i\gamma)$. Baltagi, Bresson and Pirotte (2004) consider a joint LM test of the same null hypothesis but against the more general heteroscedastic alternative $\nu_{it} \sim N(0, \sigma^2_{\nu_{it}})$ and $\mu_i \sim N(0, \sigma^2_{\mu_i})$, where $\sigma^2_{\nu_{it}}$ and $\sigma^2_{\mu_i}$ are, up to a multiplicative constant, possibly different parametric functions of vectors of explanatory variables Z^1_{it} and Z^2_i, i.e. $\sigma^2_{\nu_{it}} = \sigma^2_\nu \phi_\nu(Z^1_{it}\gamma_1)$ and $\sigma^2_{\mu_i} = \sigma^2_\mu \phi_\mu(Z^2_i\gamma_2)$. They further consider "marginal" LM tests of again the same null hypothesis but against the "marginal" heteroscedastic alternatives, on one hand, $\nu_{it} \sim N(0, \sigma^2_{\nu_{it}})$ and $\mu_i \sim N(0, \sigma^2_\mu)$, and on the other hand, $\nu_{it} \sim N(0, \sigma^2_\nu)$ and $\mu_i \sim N(0, \sigma^2_{\mu_i})$. The latter test was previously obtained by Holly and Gardiol (2000). Lejeune (2004) provides a distribution-free joint test and robust one-directional tests of the null hypothesis of no individual effect and heteroscedasticity. These tests allow one to detect, from preliminary (pooled) OLS estimation of the model, the possible simultaneous presence of both individual effects and heteroscedasticity. Randolph (1988) concentrates on supplying an observation-by-observation data transformation for a full heteroscedastic error components model assuming that $\nu_{it} \sim (0, \sigma^2_{\nu_{it}})$ and $\mu_i \sim (0, \sigma^2_{\mu_i})$. Provided that the variances $\sigma^2_{\nu_{it}}$ and $\sigma^2_{\mu_i}$ are known, this transformation allows generalized least squares estimates to be obtained from ordinary least squares. Li and Stengos (1994) deal with adaptive estimation of an error components model supposing heteroscedasticity of unknown form for the general error term, i.e. assume that $\mu_i \sim (0, \sigma^2_\mu)$ while $\nu_{it} \sim (0, \sigma^2_{\nu_{it}})$, where $\sigma^2_{\nu_{it}}$ is a non-parametric function $\phi(Z_{it})$ of a vector of explanatory variables Z_{it}. Likewise, Roy (2003) considers adaptive estimation of a error components model also assuming heteroscedasticity of unknown form, but for the individual-specific error term, i.e. supposes that $\mu_i \sim (0, \sigma^2_{\mu_i})$ while $\nu_{it} \sim (0, \sigma^2_\nu)$. Except Rao, Kaplan and Cochran (1981), Randolph (1988) and Lejeune (2004), all these papers consider balanced panels.

In this paper, we are concerned with estimation and specification testing of a full heteroscedastic one-way error components linear regression model specified in the spirit of Randolph (1988) and Baltagi, Bresson and Pirotte (2004). In short, we assume that the (conditional) variances $\sigma^2_{\nu_{it}}$ and $\sigma^2_{\mu_i}$ are distinct parametric functions of, respectively, vectors of explanatory variables Z^1_{it} and Z^2_i, i.e. $\sigma^2_{\nu_{it}} = \phi_\nu(Z^1_{it}\gamma_1)$ and $\sigma^2_{\mu_i} = \phi_\mu(Z^2_i\gamma_2)$. Further, we treat the model in the context of unbalanced panels. This specification differs from the previously proposed formulations of estimable heteroscedastic error components models as it simultaneously embodies three characteristics. First, heteroscedasticity distinctly applies to both individual-specific and general error components. Second, (nonlinear) variance functions are parametrically specified. Finally, the model allows for unbalanced panels.

Explicitly allowing for unbalanced panels is obviously desirable. Indeed, at least for micro-data, incompleteness is rather the rule than the exception. Specifying parametric variance functions is also attractive. First, this strategy avoids incidental parameter (and thus consistency) problems arising from any attempt to model changing variances by grouped heteroscedasticity when the number of individual units is large but the number of observations per individual is small, i.e. in typical microeconomic panel datasets. Second, provided that the functional forms of the variance functions are judiciously chosen, it prevents problems due to estimated variances being negative or zero. Finally, since the variance estimates may have intrinsic values of their own as indicators of the between and within individual heterogeneity, parametric forms are convenient for ease of interpretation.

The heuristic background for allowing heteroscedasticity to distinctly apply to both individual-specific and general error components is the following. In essence, except for the fact that it may be broken down into an individual-specific and a general component, the composite error term in panel data is not different from a cross-section error term. Accordingly, all we said about the possible sources of heteroscedasticity in cross-section may be roughly applied to the panel data composite error term. The only new issue is to assess the plausible origin — between and/or within, i.e. the individual-specific error and/or the general error — of any given cross-section like heteroscedasticity in the composite error term. Clearly, the answer depends upon the situation at hand. When heteroscedasticity arises from differences in size, both error terms may be expected to be heteroscedastic, presumably according to parallel patterns. As a matter of fact, this is implicitly acknowledged whenever a transformation of the dependent variable is used for solving heteroscedasticity problems (the transformation alters the distribution of both error terms). Likewise, if size-related heteroscedasticity still prevails after having transformed the dependent variable, the same should hold. When heteroscedasticity is not directly associated with size, it seems much more difficult to say anything general : depending on the situation, either only one or both error terms may be heteroscedastic, and when both are, their scedastic pattern may further be different. Be that as it may, as a general setting, it thus appears sensible to allow heteroscedasticity to distinctly apply to both individual-specific and general error components.

For estimating our proposed full heteroscedastic one-way error components model, we argue for resorting to a Gaussian pseudo-maximum likelihood of order 2 estimator (Gourieroux, Monfort and Trognon (1984, 1993), Bollerslev and Wooldridge (1992), Wooldridge (1994)). This estimator has indeed numerous nice properties : it is computationally convenient, it allows one to straightforwardly handle unbalanced panels, it is efficient under normality but robust to non-normality, and last but not least, in the present context, it is also robust to possible misspecification of the assumed scedastic structure of the data.

Further, we outline how, taking advantage of the powerful m-testing framework (Newey (1985), Tauchen (1985), White (1987, 1994), Wooldridge (1990, 1991a, 1991b)), the correct specification of the prominent aspects of our proposed model may be tested. We consider potentially useful nested, non-nested, Hausman and information matrix type diagnostic tests of both the mean and the variance specifications. Joined to the Gaussian pseudo-maximum likelihood of order 2 (GPML2) estimator, this set of diagnostic tests provides a complete statistical tool-box for estimating and evaluating the empirical relevance of our proposed model. For Gauss users, an easy-to-use package implementing this complete statistical tool-box may be obtained (free of charge) upon request from the author.

The rest of the paper proceeds as follows. Section 2 describes our proposed full heteroscedastic one-way error components model. Section 3 considers GPML2 estimation of the model and outlines its asymptotic properties. Section 4 deals with specification testing of the model. Section 5 provides an empirical illustration of the practical usefulness of our suggested model and estimation and specification testing procedures. Finally, Section 6 concludes.

2. The model

We consider the following one-way error components linear regression model

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

where Y_{it}, ε_{it}, μ_i and ν_{it} are scalars, X_{it} is a $1 \times k$ vector of strictly exogenous explanatory variables (the first element being a constant) and β is a $k \times 1$ vector of parameters. The index i refers to the individuals and the index t to the (repeated) observations (over time) of each individual i. Each individual i is assumed to be observed a fixed number of times T_i. The unbalanced structure of the panel is supposed to be ignorable in the sense of Wooldridge (1995). The total number of observations is $N = \sum_{i=1}^{n} T_i$. The observations are assumed to be independently (but not necessarily identically) distributed across individuals.

Stacking the T_i observations of each individual i, (1) yields the multivariate linear regression model

$$Y_i = X_i\beta + \varepsilon_i, \quad \varepsilon_i = e_{T_i}\mu_i + \nu_i, \qquad i = 1, 2, ..., n \tag{2}$$

where e_{T_i} is a $T_i \times 1$ vector of ones, Y_i, ν_i and ε_i are $T_i \times 1$ vectors, and X_i is a $T_i \times k$ matrix.

Let Z_i^1 denote a $T_i \times l_1$ matrix of strictly exogenous explanatory variables (the first column being a constant), Z_{it}^1 stand for the t-th row of Z_i^1, and Z_i^2 be a $1 \times l_2$ vector of strictly exogenous explanatory variables (the first element being again a constant). For all i, t and t', the error terms ν_{it} and μ_i are assumed to satisfy the assumptions

$$E(\nu_{it}|X_i, Z_i^1, Z_i^2) = 0, \;\; E(\mu_i|X_i, Z_i^1, Z_i^2) = 0 \tag{3}$$

$$E\left(\nu_{it}\nu_{it'}|X_i, Z_i^1, Z_i^2\right) = 0 \;(t' \neq t), \;\; E\left(\mu_i\nu_{it}|X_i, Z_i^1, Z_i^2\right) = 0 \tag{4}$$

$$V\left(\nu_{it}|X_i, Z_i^1, Z_i^2\right) = \sigma_{\nu_{it}}^2 = \phi_\nu(Z_{it}^1\gamma_1) \text{ and } V\left(\mu_i|X_i, Z_i^1, Z_i^2\right) = \sigma_{\mu_i}^2 = \phi_\mu(Z_i^2\gamma_2) \tag{5}$$

where $\phi_\nu(.)$ and $\phi_\mu(.)$ are (strictly) positive twice continuously differentiable functions while γ_1 and γ_2 are, respectively, $l_1 \times 1$ and $l_2 \times 1$ vectors of parameters which vary independently of each other and independently of β. Hereafter, we will denote by $\gamma = \left(\gamma_1', \gamma_2'\right)'$ the vector of variance-specific parameters, and $\theta = (\beta', \gamma')'$ will stand for the entire set of parameters.

The regressors appearing in the conditional variances (5) may (and usually will) be related to the X_i variables. Different choices are possible for the variance functions $\phi_\nu(.)$ and $\phi_\mu(.)$, see for example Breusch and Pagan (1979) and Harvey (1976). Among them, the multiplicative heteroscedasticity formulation investigated in Harvey (1976) appears particularly attractive. It simply means taking $\phi_\nu(.) = \phi_\mu(.) = \exp(.)$.

Under (3)-(5), ε_i is easily seen to satisfy

$$\begin{aligned} &E(\varepsilon_i|X_i, Z_i^1, Z_i^2) = 0, \qquad\qquad i = 1, 2, ..., n \\ &V(\varepsilon_i|X_i, Z_i^1, Z_i^2) = \Omega_i = \operatorname{diag}\left(\phi_\nu(Z_i^1\gamma_1)\right) + J_{T_i}\,\phi_\mu(Z_i^2\gamma_2) \end{aligned} \tag{6}$$

where $J_{T_i} = e_{T_i}e_{T_i}'$ is a $T_i \times T_i$ matrix of ones, and, for a $T_i \times 1$ vector x, the functions $\phi_\nu(x)$ and $\phi_\mu(x)$ denote $T_i \times 1$ vectors containing the element-by-element transformations $\phi_\nu(x)$ and $\phi_\mu(x)$ of the elements of x, $\operatorname{diag}\left(\phi_\nu(x)\right)$ further standing for a diagonal $T_i \times T_i$ matrix containing $\phi_\nu(x)$ as diagonal elements and zeros elsewhere.

The model may thus be written as

$$\begin{aligned} E(Y_i|X_i, Z_i^1, Z_i^2) &= X_i\beta, \qquad\qquad i = 1, 2, ..., n \\ V(Y_i|X_i, Z_i^1, Z_i^2) &= \Omega_i = \operatorname{diag}\left(\phi_\nu(Z_i^1\gamma_1)\right) + J_{T_i}\, \phi_\mu(Z_i^2\gamma_2) \end{aligned} \tag{7}$$

This model obviously contains the standard homoscedastic one-way error components linear regression model as a special case: it is simply obtained by letting the Z_i^1 and Z_i^2 variables only contain an intercept.

In practice, model (7) may or may not be correctly specified. It will be correctly specified for the conditional mean if the observations are indeed such that $E(Y_i|X_i, Z_i^1, Z_i^2) = X_i\beta^o$, $i = 1, 2, ..., n$, for some true value β^o. Likewise, it will be correctly specified for the conditional variance if the observations are indeed such that $V(Y_i|X_i, Z_i^1, Z_i^2) = \Omega_i^o = \operatorname{diag}\left(\phi_\nu(Z_i^1\gamma_1^o)\right) + J_{T_i}\, \phi_\mu(Z_i^2\gamma_2^o)$, $i = 1, 2, ..., n$, for some true-value $\gamma^o = (\gamma_1^{o\prime}, \gamma_2^{o\prime})'$.

3. Pseudo-maximum likelihood estimation

The most popular procedure for estimating the standard homoscedastic one-way error components model consists in first estimating the mean parameters of the model by OLS, then in estimating the variance of the error components based on the residuals obtained in the first step, and finally, for efficiency, in re-estimating the mean parameters by feasible generalized least squares (FGLS).

Pursuing a similar multiple-step procedure for estimating our proposed full heteroscedastic model does not appear very attractive. Indeed, if in the standard homoscedastic model it is straightforward to consistently estimate the variance of the error components based on first step regression residuals, it is no longer the case in our proposed full heteroscedastic model: given the general functional forms adopted for the variance functions[1], no simple — i.e. avoiding nonlinear optimization — procedure for consistently estimating the variance parameters appearing in Ω_i seems conceivable.

As nonlinear optimization appears unavoidable, we argue for estimating our proposed model by Gaussian pseudo-maximum likelihood of order two (Gourieroux, Monfort and Trognon (1984, 1993), Bollerslev and Wooldridge (1992), Wooldridge (1994)). This GPML2 estimator has numerous attractive properties. First, if it requires nonlinear optimization, it is a one-step estimator, simultaneously providing mean and variance parameters estimates. Second, as developed below, while fully efficient if normality holds, it is not only robust to non-normality (i.e. its consistency does not rely on normality) but also to possible misspecification of the conditional variance (i.e. it remains consistent for the mean parameters even if the assumed scedastic structure of the data is misspecified). Finally, it readily allows one to handle unbalanced panels.

[1] The problem would be different if the variance functions were assumed linear. Specifying linear variance functions is however not a good idea as it may result in estimated variances being negative or zero.

3.1. The GPML2 estimator

The GPML2 estimator $\hat{\theta}_n = \left(\hat{\beta}_n', \hat{\gamma}_{1_n}', \hat{\gamma}_{2_n}'\right)'$ of model (7) is defined as a solution of

$$\mathrm{Max}_{\theta \in \Theta}\, L_n(\beta, \gamma_1, \gamma_2) = \frac{1}{n} \sum_{i=1}^{n} L_i(Y_i | X_i, Z_i^1, Z_i^2; \beta, \gamma_1, \gamma_2) \tag{8}$$

where Θ denotes the parameter space and the (conditional) pseudo log-likelihood functions $L_i(Y_i|X_i, Z_i^1, Z_i^2; \beta, \gamma_1, \gamma_2)$ are

$$L_i(Y_i|X_i, Z_i^1, Z_i^2; \beta, \gamma_1, \gamma_2) = -\frac{T_i}{2} \ln 2\pi - \frac{1}{2} \ln |\Omega_i| - \frac{1}{2} u_i' \Omega_i^{-1} u_i$$

with $u_i = Y_i - X_i\beta$.

Closed-form expressions are available for $|\Omega_i|$ and Ω_i^{-1}. These are given in Appendix A, where we also provide expressions for the first derivatives, Hessian matrix and expected Hessian matrix of the pseudo log-likelihood function $L_n(\beta, \gamma_1, \gamma_2)$.

If one checks the first-order conditions defining $\hat{\theta}_n$, it is evident that the GPML2 mean-specific estimator $\hat{\beta}_n$ is nothing but a FGLS estimator where the variance parameters appearing in Ω_i are jointly estimated. Additionally, the GPML2 variance-specific estimator $\hat{\gamma}_n = \left(\hat{\gamma}_{1_n}', \hat{\gamma}_{2_n}'\right)'$ may be interpreted as a weighted nonlinear least squares estimator in the multivariate nonlinear regression model $\mathrm{vec}(u_i u_i') = \mathrm{vec}\, \Omega_i(\gamma_1, \gamma_2) +$ residuals, $i = 1, 2, ..., n$, where the errors u_i and the weights $\Gamma_i^{-1} = (\Omega_i^{-1} \otimes \Omega_i^{-1})$ are likewise jointly estimated.

Practical guidelines for computing the GPML2 estimator $\hat{\theta}_n$, including a numerical algorithm and starting values, are discussed in Appendix B.

3.2. Asymptotic properties of the GPML2 estimator

Beyond its computational convenience and its ability to readily handle unbalanced panel, the most attractive feature of the GPML2 estimator is its statistical properties, namely its potential efficiency and its robustness.

Obviously, when the model is correctly specified for both the conditional mean and the conditional variance and when in addition normality also holds, the GPML2 estimator is just a standard maximum likelihood estimator. According to standard maximum likelihood theory, we then have that $\hat{\theta}_n$ is consistent and asymptotically normal,

$$\hat{\theta}_n \xrightarrow{p} \theta^o \quad \text{and} \quad \sqrt{n}(\hat{\theta}_n - \theta^o) \approx \mathcal{N}(0, \bar{C}_n^o), \text{ as } n \to \infty \ (T_i \text{ bounded})$$

with an asymptotic covariance matrix given by

$$\bar{C}_n^o = \begin{bmatrix} -\bar{A}_{\beta\beta}^{-1} & 0 \\ 0 & -\bar{A}_{\gamma\gamma}^{-1} \end{bmatrix}$$

where

$$\bar{A}_{\beta\beta} = \frac{1}{n}\sum_{i=1}^{n} E\left[\underline{h}_i^{\beta\beta}\right]_{\theta=\theta^o}, \quad \bar{A}_{\gamma\gamma} = \frac{1}{n}\sum_{i=1}^{n} E\left[\underline{h}_i^{\gamma\gamma}\right]_{\theta=\theta^o}, \quad \underline{h}_i^{\gamma\gamma} = \begin{bmatrix} \underline{h}_i^{\gamma_1\gamma_1} & \underline{h}_i^{\gamma_1\gamma_2} \\ \underline{h}_i^{\gamma_2\gamma_1} & \underline{h}_i^{\gamma_2\gamma_2} \end{bmatrix}$$

and $\underline{h}_i^{\beta\beta}$ and $\underline{h}_i^{\gamma\gamma}$ refer to the expected Hessian of L_i and are defined in Appendix A.

In this favorable situation, the GPML2 estimator is fully efficient, both for the mean and the variance parameters. However, since in practice normality may at best be expected to only very approximately hold, this result must essentially be viewed as a benchmark result.

As for all pseudo-maximum likelihood estimators, the distributional normality assumption underlying the GPML2 estimator is purely nominal. As a matter of fact, according to second order Gaussian pseudo-maximum likelihood theory (Gourieroux, Monfort and Trognon (1984, 1993), Bollerslev and Wooldridge (1992), Wooldridge (1994)), if the model is correctly specified for the conditional mean and the conditional variance but normality does not hold, we still have that $\hat{\theta}_n$ is consistent and asymptotically normal,

$$\hat{\theta}_n \xrightarrow{p} \theta^o \quad \text{and} \quad \sqrt{n}(\hat{\theta}_n - \theta^o) \approx \mathcal{N}(0, C_n^o), \text{ as } n \to \infty \ (T_i \text{ bounded})$$

but with a more complicated asymptotic covariance matrix given by

$$C_n^o = \begin{bmatrix} -\bar{A}_{\beta\beta}^{-1} & \bar{A}_{\beta\beta}^{-1} B_{\beta\gamma} \bar{A}_{\gamma\gamma}^{-1} \\ \bar{A}_{\gamma\gamma}^{-1} B_{\gamma\beta} \bar{A}_{\beta\beta}^{-1} & \bar{A}_{\gamma\gamma}^{-1} B_{\gamma\gamma} \bar{A}_{\gamma\gamma}^{-1} \end{bmatrix}$$

where

$$B_{\beta\gamma} = \frac{1}{n}\sum_{i=1}^{n} E\left[s_i^{\beta} s_i^{\gamma\prime}\right]_{\theta=\theta^o} = B'_{\gamma\beta}, \quad B_{\gamma\gamma} = \frac{1}{n}\sum_{i=1}^{n} E\left[s_i^{\gamma} s_i^{\gamma\prime}\right]_{\theta=\theta^o}, \quad s_i^{\gamma} = \left(s_i^{\gamma_1\prime}, s_i^{\gamma_2\prime}\right)'$$

and s_i^{β} and s_i^{γ} refer to the first derivatives of L_i and are again defined in Appendix A.

Note that non-normality does not affect the asymptotic covariance matrix of the GPML2 mean-specific estimator $\hat{\beta}_n$. It is still given by $-\bar{A}_{\beta\beta}^{-1}$, which, since $\hat{\beta}_n$ is in fact nothing but a FGLS estimator, is actually equal to the asymptotic covariance matrix of the usual FGLS estimator (implemented using any consistent estimator of the variance parameters appearing in Ω_i). Of course, in this situation, the GPML2 estimator is no longer fully efficient. It is clearly not efficient regarding the variance parameters. Regarding the mean parameters, as FGLS, it is however still efficient in a semi-parametric sense[2].

Besides being robust to non-normality, the GPML2 estimator has an additional nice property in that it is also robust to conditional variance misspecification, i.e. to misspecification of the assumed scedastic structure of the data. Since the GPML2 mean-specific estimator $\hat{\beta}_n$ is a FGLS estimator, this should not be surprising[3].

[2] The asymptotic covariance matrix of $\hat{\beta}_n$ attains the well-known semi-parametric efficiency bound (Chamberlain (1987), Newey (1990, 1993), Wooldridge (1994)) associated with optimal GMM estimation based on the first order conditional moments of the data.

[3] It is well-known that conditional variance misspecification does not affect the consistency of the FGLS

According to Lejeune (1998), if the model is correctly specified for the conditional mean but misspecified for the conditional variance, it indeed turns out that $\hat{\beta}_n$ is still consistent for its true value β^o while $\hat{\gamma}_n$ is now consistent for some pseudo-true value $\gamma_n^* = \left(\gamma_{1_n}^{*\prime}, \gamma_{2_n}^{*\prime}\right)'$,

$$\hat{\beta}_n \xrightarrow{p} \beta^o \quad \text{and} \quad \hat{\gamma}_n - \gamma_n^* \xrightarrow{p} 0, \text{ as } n \to \infty \ (T_i \text{ bounded})$$

and that $\hat{\theta}_n$ remains jointly asymptotically normal

$$\sqrt{n}(\hat{\theta}_n - \theta_n^{o*}) \approx \mathcal{N}(0, C_n^{o*}), \text{ as } n \to \infty \ (T_i \text{ bounded}), \text{ where } \theta_n^{o*} = \left(\beta^{o\prime}, \gamma_n^{*\prime}\right)'$$

with an asymptotic covariance matrix given by

$$C_n^{o*} = \begin{bmatrix} A_{\beta\beta}^{-1} B_{\beta\beta} A_{\beta\beta}^{-1} & A_{\beta\beta}^{-1} B_{\beta\gamma} A_{\gamma\gamma}^{-1} \\ A_{\gamma\gamma}^{-1} B_{\gamma\beta} A_{\beta\beta}^{-1} & A_{\gamma\gamma}^{-1} \ddot{B}_{\gamma\gamma} A_{\gamma\gamma}^{-1} \end{bmatrix}$$

where

$$A_{\beta\beta} = \frac{1}{n} \sum_{i=1}^{n} E\left[h_i^{\beta\beta}\right]_{\theta=\theta_n^{o*}}, \quad A_{\gamma\gamma} = \frac{1}{n} \sum_{i=1}^{n} E\left[h_i^{\gamma\gamma}\right]_{\theta=\theta_n^{o*}}, \quad h_i^{\gamma\gamma} = \begin{bmatrix} h_i^{\gamma_1\gamma_1} & h_i^{\gamma_1\gamma_2} \\ h_i^{\gamma_2\gamma_1} & h_i^{\gamma_2\gamma_2} \end{bmatrix}$$

$$B_{\beta\beta} = \frac{1}{n} \sum_{i=1}^{n} E\left[s_i^{\beta} s_i^{\beta\prime}\right]_{\theta=\theta_n^{o*}}, \quad B_{\beta\gamma} = \frac{1}{n} \sum_{i=1}^{n} E\left[s_i^{\beta} s_i^{\gamma\prime}\right]_{\theta=\theta_n^{o*}} = B_{\gamma\beta}'$$

$$\ddot{B}_{\gamma\gamma} = \frac{1}{n} \sum_{i=1}^{n} E\left[s_i^{\gamma} s_i^{\gamma\prime}\right]_{\theta=\theta_n^{o*}} - U_{\gamma\gamma}, \quad U_{\gamma\gamma} = \frac{1}{n} \sum_{i=1}^{n} E\left[s_i^{\gamma}\right]_{\theta=\theta_n^{o*}} E\left[s_i^{\gamma}\right]_{\theta=\theta_n^{o*}}'$$

and $h_i^{\beta\beta}$ and $h_i^{\gamma\gamma}$ refer to the Hessian of L_i and are again defined in Appendix A.

Of course, in this latter situation, the GPML2 mean-specific estimator $\hat{\beta}_n$ is no longer efficient. However, as its asymptotic covariance matrix $A_{\beta\beta}^{-1} B_{\beta\beta} A_{\beta\beta}^{-1}$ collapses to the semi-parametric efficiency bound $-\bar{A}_{\beta\beta}^{-1}$ outlined above when the conditional variance is correctly specified, we may intuitively expect that the more the specified conditional variance is close to the actual scedastic structure of the data, the more the covariance matrix of $\hat{\beta}_n$ will be close to this lower bound, i.e. $\hat{\beta}_n$ will be close to semi-parametric efficiency. From a empirical point of view, this in particular implies that it makes sense to consider using our proposed full heteroscedastic model, even if possibly misspecified, whenever the homoscedasticity assumption of the standard one-way error components model does not appear to hold: some efficiency benefits may indeed generally be expected from taking into account even approximately the actual scedastic structure of the data.

In practical applications, the extent to which our assumed full heteroscedastic model is actually correctly specified is of course a priori unknown. This may nevertheless be checked through diagnostic tests, as discussed in Section 4 below. Once this is done, a consistent estimate of the asymptotic covariance matrix of the estimated parameters may then be straightforwardly computed by taking, as usual,

estimator.

the empirical counterpart of the relevant theoretical asymptotic covariance matrix[4]. There is one exception however: due to the term $U_{\gamma\gamma}$, unless the observations are IID and the panel dataset is balanced (in which case $U_{\gamma\gamma} = 0$), a consistent estimate of the asymptotic covariance matrix $A_{\gamma\gamma}^{-1}\ddot{B}_{\gamma\gamma}A_{\gamma\gamma}^{-1}$ of the GPML2 variance-specific estimator $\hat{\gamma}_n$ under correct conditional mean specification but conditional variance misspecification may in general not be obtained. A consistent estimate of an upper bound of this asymptotic covariance matrix, upper bound given by $A_{\gamma\gamma}^{-1}B_{\gamma\gamma}A_{\gamma\gamma}^{-1}$ where $B_{\gamma\gamma} = \frac{1}{n}\sum_{i=1}^{n} E\left[s_i^{\gamma}s_i^{\gamma\prime}\right]_{\theta=\theta_n^{o*}}$, may nevertheless be computed in the usual way. Interestingly, based on this estimated upper bound, a conservative — i.e. with asymptotic true size necessarily inferior to its specified nominal size — (joint) Wald test of the null hypothesis that the non-intercept parameters of γ_1 and γ_2 are zero may then be validly performed. In other words, a valid conservative test which checks that, as assumed, the observations indeed exhibit some heteroscedasticity-like pattern related to the Z_i^1 and Z_i^2 explanatory variables may then readily be carried out, and this is regardless of possible conditional variance misspecification.

4. Specification testing

The GPML2 estimator of model (7) always delivers a consistent estimate of the mean parameters if the model is correctly specified for the conditional mean, and consistent estimates of both the mean and variance parameters if the model is correctly specified for both the conditional mean and the conditional variance. But nothing a priori guarantees that the model is indeed correctly specified.

Hereafter, we outline how, taking advantage of the powerful m-testing framework (Newey (1985), Tauchen (1985), White (1987, 1994), Wooldridge (1990, 1991a, 1991b)), the conditional mean and the conditional variance specification of our proposed full heteroscedastic one-way error components model may be checked. We first consider conditional mean diagnostic tests, and then conditional variance diagnostic tests.

4.1. Conditional mean diagnostic tests

Having estimated our proposed model (7), the first thing to consider is to check its conditional mean specification. Testing the null hypothesis that the conditional mean is correctly specified means testing

$$\mathrm{H}_0^m : E(Y_i|X_i, Z_i^1, Z_i^2) = X_i\beta^o, \text{ for some } \beta^o, \quad i = 1, 2, ..., n$$

Following White (1987, 1994), Wooldridge (1990, 1991a, 1991b) and Lejeune (1998), based on the GPML2 estimator $\hat{\theta}_n$, H_0^m may efficiently be tested by checking, for appropriate choices of $T_i \times q$ indicator matrices $\hat{W}_i^m$ (which may depend on

[4] For example, a consistent estimate of the asymptotic covariance matrix $A_{\beta\beta}^{-1}B_{\beta\beta}A_{\beta\beta}^{-1}$ of the GPML2 mean-specific estimator $\hat{\beta}_n$ under correct conditional mean specification but conditional variance misspecification may be computed as $\hat{A}_{\beta\beta}^{-1}\hat{B}_{\beta\beta}\hat{A}_{\beta\beta}^{-1}$, where $\hat{A}_{\beta\beta} = \frac{1}{n}\sum_{i=1}\hat{h}_i^{\beta\beta}$, $\hat{B}_{\beta\beta} = \frac{1}{n}\sum_{i=1}^{n}\hat{s}_i^{\beta}\hat{s}_i^{\beta\prime}$ and the superscript '$\hat{\ }$' denotes quantities evaluated at $\hat{\theta}_n$.

the conditioning variables (X_i, Z_i^1, Z_i^2) as well as on additional estimated nuisance parameters), that $q \times 1$ misspecification indicators of the form

$$\hat{\Phi}_n^m = \frac{1}{n} \sum_{i=1}^{n} \hat{W}_i^{m\prime} \hat{\Omega}_i^{-1} \hat{u}_i \tag{9}$$

are not significantly different from zero.

Given the assumed statistical setup, a relevant statistic for checking that $\hat{\Phi}_n^m$ is not significantly different from zero is given by the asymptotic chi-squared statistic[5]

$$\begin{aligned}\mathcal{M}_n^m \;=\; & \left(\sum_{i=1}^{n} \hat{W}_i^{m\prime} \hat{\Omega}_i^{-1} \hat{u}_i \right)' \\ & \left(\sum_{i=1}^{n} \left(\hat{W}_i^m - X_i \hat{P}^m \right)' \hat{\Omega}_i^{-1} \hat{u}_i \hat{u}_i' \hat{\Omega}_i^{-1} \left(\hat{W}_i^m - X_i \hat{P}^m \right) \right)^{-1} \\ & \left(\sum_{i=1}^{n} \hat{W}_i^{m\prime} \hat{\Omega}_i^{-1} \hat{u}_i \right) \xrightarrow{d} \chi^2(q)\end{aligned}$$

where

$$\hat{P}^m = \left(\sum_{i=1}^{n} X_i' \hat{\Omega}_i^{-1} X_i \right)^{-1} \sum_{i=1}^{n} X_i' \hat{\Omega}_i^{-1} \hat{W}_i^m$$

By suitably choosing the $T_i \times q$ indicator matrices $\hat{W}_i^m$ in (9), as detailed below, H_0^m may be tested against nested alternatives, non-nested alternatives, or without resorting to explicit alternatives through Hausman and information matrix type tests.

A prominent characteristic of all conditional mean diagnostic tests implemented through the $\mathcal{M}_n^m$ statistic is that they yield valid tests of H_0^m regardless of whether or not the assumed scedastic pattern of the data is correct and whether or not normality holds. Consequently, since they do not rely on assumptions other than H_0^m itself, a rejection may always be unambiguously attributed to a failure of H_0^m to hold. Interestingly, another important characteristic of diagnostic tests implemented through $\mathcal{M}_n^m$ is that they will have optimal properties if the conditional variance is actually correctly specified and normality holds.

Following Wooldridge (1990, 1991a, 1991b) and Lejeune (1998), for testing H_0^m against a nested alternative of the form

$$\mathrm{H}_1^m : E(Y_i | X_i, Z_i^1, Z_i^2) = m_i^a(X_i, Z_i^1, Z_i^2, \beta^o, \alpha^o), \text{ for some } (\beta^{o\prime}, \alpha^{o\prime})', \; i = 1, 2, ..., n$$

where $m_i^a(X_i, Z_i^1, Z_i^2, \beta, \alpha)$ denotes some alternative conditional mean specification such that for some value $\alpha = c$ of the $q \times 1$ vector of additional parameters α we

[5] Note that $\mathcal{M}_n^m$ may in practice be computed as n minus the residual sum of squares $(= nR_u^2$, R_u^2 being the uncentered R-squared) of the artificial OLS regression $1 = \left[\hat{u}_i' \hat{\Omega}_i^{-1} \left(\hat{W}_i^m - X_i \hat{P}^m \right) \right] b +$ residuals, $i = 1, 2, ..., n$.

have

$$m_i^a(X_i, Z_i^1, Z_i^2, \beta, c) = X_i\beta, \quad i = 1, 2, ..., n$$

the appropriate choice of $\hat{W}_i^m$ is given by

$$\hat{W}_i^m = \frac{\partial m_i^a(X_i, Z_i^1, Z_i^2, \hat{\beta}_n, c)}{\partial \alpha'}$$

When the considered alternative conditional mean specification takes the simple linear form

$$m_i^a(X_i, Z_i^1, Z_i^2, \beta, \alpha) = X_i\beta + G_i\alpha, \quad i = 1, 2, ..., n$$

where G_i is a $T_i \times q$ matrix of variables which are functions of the set of conditioning variables $CV_i \equiv (X_i, Z_i^1, Z_i^2)$, $\hat{W}_i^m$ is simply equal to G_i and the test corresponds to a standard variable addition test. We may for example check in this way the linearity of the assumed conditional mean by setting G_i equal to (some of) the squares and/or the cross-products of (some of) the X_i variables.

On the other hand, for testing H_0^m against a non-nested alternative such as

$$\mathrm{H}_1^m : E(Y_i|X_i, Z_i^1, Z_i^2) = g_i^a(X_i, Z_i^1, Z_i^2, \delta^o), \text{ for some } \delta^o, \quad i = 1, 2, ..., n$$

where $g_i^a(X_i, Z_i^1, Z_i^2, \delta)$ denotes some alternative conditional mean specification which does not contain the null conditional mean specification $X_i\beta$ as a special case and δ is a vector of parameters, an appropriate choice of $\hat{W}_i^m$ is given by

$$\hat{W}_i^m = g_i^a(X_i, Z_i^1, Z_i^2, \hat{\delta}_n) - X_i\hat{\beta}_n$$

where $\hat{\delta}_n$ is any consistent estimator of δ^o under H_1^m. This yields a Davidson and MacKinnon (1981) type test of a non-nested alternative. Because obvious choices of $g_i^a(.)$ are in practice rarely available, this kind of test of H_0^m is unlikely to be routinely performed. It may however be useful in some situations.

By construction, diagnostic tests against nested or non-nested alternatives have power against the specific alternative they consider, but may be expected to have limited power against other (if weakly related) alternatives. General purpose diagnostic tests with expected power against a broader range of alternatives are provided by Hausman and information matrix type tests.

One of the equivalent forms of the popular Hausman specification test of the standard homoscedastic one-way error components model is based on comparing the (non-intercept) FGLS and OLS estimators of β^o (see for example Baltagi (1995)). This strongly suggests considering a generalized (i.e. allowing for any choice of S and robust to conditional variance misspecification) Hausman type test of H_0^m based on checking, for some chosen selection matrix S, the closeness to zero of the misspecification indicator

$$\hat{\Phi}_n^m = S(\hat{\beta}_n - \hat{\beta}_n^{OLS})$$

Following the lines of White (1994) and Lejeune (1998), a test that is asymptotically equivalent to checking the above misspecification indicator is obtained by setting

$$\hat{W}_i^m = \hat{\Omega}_i X_i \hat{Q}^{-1} S'$$

where $\hat{Q} = \sum_{i=1}^{n} X_i' X_i$. As is the case with the standard textbook Hausman test (to which it is asymptotically equivalent under standard textbook homoscedasticity conditions), this test will have power against any alternative H_1^m for which $\hat{\beta}_n$ and $\hat{\beta}_n^{OLS}$ converge to different pseudo-true values. Note by the way that, contrary to the standard textbook case, heteroscedasticity (and incompleteness) usually allows one to include all β parameters as part of this Hausman test without yielding a singular statistic.

On the other hand, following again the lines of White (1994) and Lejeune (1998), an information matrix type test of H_0^m may be based on checking, for some chosen selection matrix S, the closeness to zero of the misspecification indicator

$$\hat{\Phi}_n^m = S \frac{1}{n} \sum_{i=1}^{n} \operatorname{vec} \hat{h}_i^{\beta\gamma}, \qquad h_i^{\beta\gamma} = \left[\begin{array}{cc} h_i^{\beta\gamma_1} & h_i^{\beta\gamma_2} \end{array} \right]$$

where $h_i^{\beta\gamma}$ refers to cross-derivatives of L_i and is defined in Appendix A. Such a test essentially involves checking the block diagonality between mean and variance parameters of the expected Hessian matrix of the GPML2 estimator, which must hold under correct conditional mean specification (regardless of the correctness of the conditional variance specification). It is obtained by setting

$$\hat{W}_i^m = \hat{F}_i S'$$

where

$$\hat{F}_i = \left[\begin{array}{cccccc} \frac{\partial \hat{\Omega}_i}{\partial \gamma_1^1} \hat{\Omega}_i^{-1} X_i & \cdots & \frac{\partial \hat{\Omega}_i}{\partial \gamma_1^{l_1}} \hat{\Omega}_i^{-1} X_i & \frac{\partial \hat{\Omega}_i}{\partial \gamma_2^1} \hat{\Omega}_i^{-1} X_i & \cdots & \frac{\partial \hat{\Omega}_i}{\partial \gamma_2^{l_2}} \hat{\Omega}_i^{-1} X_i \end{array} \right]$$

and $\frac{\partial \Omega_i}{\partial \gamma_p^r}$ $(p = 1, 2)$ is again defined in Appendix A. This test, which will have power against any alternative H_1^m for which the block diagonality of the expected Hessian matrix the GPML2 estimator fails, is a quite natural complement to the above Hausman test for testing H_0^m without resorting to explicit alternatives. Note that if the multiplicative heteroscedasticity formulation is adopted for both $\phi_\nu(.)$ and $\phi_\mu(.)$, one of the two matrix elements $\frac{\partial \hat{\Omega}_i}{\partial \gamma_1^1} \hat{\Omega}_i^{-1} X_i$ and $\frac{\partial \hat{\Omega}_i}{\partial \gamma_2^1} \hat{\Omega}_i^{-1} X_i$ of $\hat{F}_i$ is redundant (yielding a singular statistic for S being set to an identity matrix) and must thus be discarded.

When a test against a specific nested or non-nested alternative rejects the null hypothesis H_0^m, it is natural to then consider modifying the originally assumed conditional mean specification in the direction of the considered alternative. When a Hausman or information matrix type test rejects H_0^m, the way that one should react is less obvious and depends on the situation at hand. In all cases, considering further diagnostic tests against various nested or non-nested alternatives should help one to identify the source(s) of rejection of H_0^m.

To conclude this brief review of conditional mean diagnostic m-tests, we make one additional remark. In empirical practice, it is not unusual for one to test the null model against an explicit alternative which includes variables which are not functions of the original set of conditioning variables $CV_i \equiv (X_i, Z_i^1, Z_i^2)$. This does not modify the way in which testing against explicit alternatives is implemented. It is however important to be aware that, in such a case, we are no longer only testing the null H_0^m but instead the null $\mathrm{H}_0^{m\prime}$: H_0^m holds and $E(Y_i|X_i, Z_i^1, Z_i^2, \underline{G}_i) = E(Y_i|X_i, Z_i^1, Z_i^2)$, $i = 1, 2, ..., n$, where $\underline{G}_i$ denotes the variables which are not functions of CV_i. In other words, we are jointly testing that H_0^m holds and that the additional $\underline{G}_i$ variables are irrelevant as conditioning variables for the expectation of Y_i. We thus must be careful in interpreting such a specification test given that H_0^m might well hold while $\mathrm{H}_0^{m\prime}$ does not.

4.2. Conditional variance diagnostic tests

Having tested — and if needed adjusted — the conditional mean specification of the model, we may then check its conditional variance specification. Testing the null hypothesis that the conditional variance is correctly specified entails testing the null

$$\mathrm{H}_0^v : \begin{cases} \mathrm{H}_0^m \text{ holds and, for some } \gamma^o, \\ V(Y_i|X_i, Z_i^1, Z_i^2) = \mathrm{diag}\,(\phi_\nu(Z_i^1\gamma_1^o)) + J_{T_i}\,\phi_\mu(Z_i^2\gamma_2^o), \quad i = 1, 2, ..., n \end{cases}$$

Note that H_0^v embodies H_0^m: there is indeed no way to test the conditional variance specification without simultaneously assuming that the conditional mean is correctly specified. This is however not a real problem since, using the above diagnostic tests, the conditional mean specification may in a first step be checked without having to assume correct conditional variance specification.

Following again White (1987, 1994), Wooldridge (1990, 1991a, 1991b)) and Lejeune (1998), based on the GPML2 estimator $\hat{\theta}_n$, H_0^v may efficiently be tested by checking, for appropriate choices of $T_i^2 \times q$ indicator matrices $\hat{W}_i^v$ (which may depend on the conditioning variables (X_i, Z_i^1, Z_i^2) as well as on additional estimated nuisance parameters), that $q \times 1$ misspecification indicators which similarly are of the form

$$\hat{\Phi}_n^v = \frac{1}{n}\sum_{i=1}^n \hat{W}_i^{v\prime}\hat{\Gamma}_i^{-1}\hat{v}_i \tag{10}$$

where

$$\hat{\Gamma}_i^{-1} = \left(\hat{\Omega}_i^{-1} \otimes \hat{\Omega}_i^{-1}\right) \quad \text{and} \quad \hat{v}_i = \mathrm{vec}(\hat{u}_i\hat{u}_i' - \hat{\Omega}_i)$$

are not significantly different from zero.

Given the assumed statistical setup, a relevant statistic for checking that $\hat{\Phi}_n^v$ is not significantly different from zero is given by the asymptotic chi-squared statistic[6]

$$\mathcal{M}_n^v = \left(\sum_{i=1}^n \hat{W}_i^{v\prime}\hat{\Gamma}_i^{-1}\hat{v}_i\right)'$$

[6] Note that $\mathcal{M}_n^v$ may in practice be computed as n minus the residual sum of squares ($= nR_u^2$, R_u^2 being the uncentered R-squared) of the artificial OLS regression $1 = \left[\hat{v}_i'\hat{\Gamma}_i^{-1}\left(\hat{W}_i^v - \frac{\partial\,\mathrm{vec}\,\hat{\Omega}_i}{\partial\gamma'}\hat{P}^v\right)\right] b$ + residuals, $i = 1, 2, ..., n$.

$$\left(\sum_{i=1}^{n}\left(\hat{W}_i^v - \tfrac{\partial\,\mathrm{vec}\,\hat{\Omega}_i}{\partial\gamma'}\hat{P}^v\right)'\hat{\Gamma}_i^{-1}\hat{v}_i\hat{v}_i'\hat{\Gamma}_i^{-1}\left(\hat{W}_i^v - \tfrac{\partial\,\mathrm{vec}\,\hat{\Omega}_i}{\partial\gamma'}\hat{P}^v\right)\right)^{-1}$$
$$\left(\sum_{i=1}^{n}\hat{W}_i^{v\prime}\hat{\Gamma}_i^{-1}\hat{v}_i\right) \xrightarrow{d} \chi^2(q)$$

where

$$\frac{\partial\,\mathrm{vec}\,\hat{\Omega}_i}{\partial\gamma'} = \left[\begin{array}{cc}\dfrac{\partial\,\mathrm{vec}\,\hat{\Omega}_i}{\partial\gamma_1'} & \dfrac{\partial\,\mathrm{vec}\,\hat{\Omega}_i}{\partial\gamma_2'}\end{array}\right]$$

$$\hat{P}^v = \left(\sum_{i=1}^{n}\left(\frac{\partial\,\mathrm{vec}\,\hat{\Omega}_i}{\partial\gamma'}\right)'\hat{\Gamma}_i^{-1}\frac{\partial\,\mathrm{vec}\,\hat{\Omega}_i}{\partial\gamma'}\right)^{-1}\sum_{i=1}^{n}\left(\frac{\partial\,\mathrm{vec}\,\hat{\Omega}_i}{\partial\gamma'}\right)'\hat{\Gamma}_i^{-1}\hat{W}_i^v$$

and $\frac{\partial\,\mathrm{vec}\,\Omega_i}{\partial\gamma'}$ is defined in Appendix A.

As was the case with the conditional mean diagnostic tests, by suitably choosing the $T_i^2 \times q$ indicator matrices $\hat{W}_i^v$ in (10), as detailed below, H_0^v may be tested against nested alternatives, non-nested alternatives, or without resorting to explicit alternatives through Hausman and information matrix type tests.

A prominent characteristic of all conditional variance diagnostic tests implemented through the $\mathcal{M}_n^v$ statistic is that they yield valid tests of H_0^v whether or not normality holds. Consequently, since they do not rely on assumptions other than H_0^v itself, a rejection may always be unambiguously attributed to a failure of H_0^v to hold. Further, given the nested nature of H_0^m and H_0^v and the robustness to possible conditional variance misspecification of the diagnostic tests of H_0^m, if no misspecification has been detected by conditional mean diagnostic tests, a rejection of H_0^v may then sensibly be attributed to conditional variance misspecification: situations where conditional variance diagnostic tests detect a misspecification in the mean which has not been detected by conditional mean diagnostic tests are indeed likely to be rare in practice. Interestingly, another important characteristic of diagnostic tests implemented through the $\mathcal{M}_n^v$ is that they will have optimal properties if normality actually holds.

Following White (1994), Wooldridge (1990, 1991a, 1991b) and Lejeune (1998), for testing H_0^v against a nested alternative of the form

$$\mathrm{H}_1^v : \begin{cases} \mathrm{H}_0^m \text{ holds and, for some } (\gamma^{o\prime}, \alpha^{o\prime}) \\ V(Y_i|X_i, Z_i^1, Z_i^2) = \Omega_i^a(X_i, Z_i^1, Z_i^2, \gamma^o, \alpha^o), \quad i = 1, 2, ..., n \end{cases}$$

where $\Omega_i^a(X_i, Z_i^1, Z_i^2, \gamma, \alpha)$ denotes some alternative conditional variance specification such that for some value $\alpha = c$ of the $q \times 1$ vector of additional parameters α we have

$$\Omega_i^a(X_i, Z_i^1, Z_i^2, \gamma, c) = \mathrm{diag}\left(\phi_\nu(Z_i^1\gamma_1)\right) + J_{T_i}\,\phi_\mu(Z_i^2\gamma_2), \quad i = 1, 2, ..., n$$

the appropriate choice of $\hat{W}_i^v$ is given by

$$\hat{W}_i^v = \frac{\partial\,\mathrm{vec}\,\Omega_i^a(X_i, Z_i^1, Z_i^2, \hat{\gamma}_n, c)}{\partial\alpha'}$$

If the considered nested alternative takes the simple semi-linear form

$$\Omega_i^a(X_i, Z_i^1, Z_i^2, \gamma, \alpha) = \text{diag}\left(\phi_\nu(Z_i^1\gamma_1 + G_i^1\alpha_1)\right) + J_{T_i}\,\phi_\mu(Z_i^2\gamma_2 + G_i^2\alpha_2)$$

where $\alpha = (\alpha_1', \alpha_2')'$ and G_i^1 and G_i^2 are respectively $T_i \times q_1$ matrices and $1 \times q_2$ vectors ($q_1 + q_2 = q$) of variables which are functions of the set of conditioning variables $CV_i \equiv (X_i, Z_i^1, Z_i^2)$, the test corresponds to a variable addition test and $\hat{W}_i^v$ is equal to

$$\hat{W}_i^v = \left[\begin{array}{cc} \hat{W}_i^{v1} & \hat{W}_i^{v2} \end{array}\right]$$

with

$$\begin{aligned}
\hat{W}_i^{v1} &= \text{diag}\left(\text{vec}\left(\text{diag}\left(\phi_\nu'(Z_i^1\hat{\gamma}_{1_n})\right)\right)\right)\left(G_i^1 \otimes e_{T_i}\right) \\
&= \sum_{r=1}^{q_1} \text{vec}\left(\text{diag}\left(\phi_\nu'(Z_i^1\hat{\gamma}_{1_n}) \odot G_i^{1^r}\right)\right) e_{q_1}^{r'} \\
\hat{W}_i^{v2} &= \phi_\mu'(Z_i^2\hat{\gamma}_{2_n})\,\text{vec}\left(J_{n_i}\right) G_i^2 = \sum_{r=1}^{q_2} \text{vec}\left(\phi_\mu'(Z_i^2\hat{\gamma}_{2_n}) G_i^{2^r} J_{T_i}\right) e_{q_2}^{r'}
\end{aligned}$$

where $\phi_\nu'(.)$ and $\phi_\mu'(.)$ stand for the first derivatives of $\phi_\nu(.)$ and $\phi_\mu(.)$, $G_i^{1^r}$ and $G_i^{2^r}$ denote the r-th column of respectively G_i^1 and G_i^2, $e_{q_1}^r$ and $e_{q_2}^r$ are respectively $q_1 \times 1$ and $q_2 \times 1$ vectors with a one in the r-th place and zeros elsewhere, and $\odot$ stands for the Hadamard product, i.e. an element-by-element multiplication. As for the conditional mean, we may for example check in this way the semi-linearity of the assumed conditional variance by setting G_i^1 and G_i^2 equal to (some of) the squares and/or the cross-products of (some of) the Z_i^1 and Z_i^2 variables.

On the other hand, for testing H_0^v against a non-nested alternative such as

$$\text{H}_1^v : \left\{ \begin{array}{l} \text{H}_0^m \text{ holds and, for some } \delta^o, \\ V(Y_i|X_i, Z_i^1, Z_i^2) = \Sigma_i^a(X_i, Z_i^1, Z_i^2, \delta^o), \quad i = 1, 2, ..., n \end{array} \right.$$

where $\Sigma_i^a(X_i, Z_i^1, Z_i^2, \delta)$ denotes some alternative conditional variance specification which does not contain the null conditional variance specification Ω_i as a special case and δ is a vector of parameters, appropriate choices of $\hat{W}_i^v$ are given by

$$\hat{W}_i^v = \text{vec}(\hat{\Sigma}_i^a - \hat{\Omega}_i) \tag{11}$$

and

$$\hat{W}_i^v = \text{vec}(\hat{\Omega}_i \hat{\Sigma}_i^{a^{-1}} \hat{\Omega}_i - \hat{\Omega}_i) \tag{12}$$

where $\hat{\Sigma}_i^a = \Sigma_i^a(X_i, Z_i^1, Z_i^2, \hat{\delta}_n)$ and $\hat{\delta}_n$ is any consistent estimator of δ^o under H_1^v. The first possible choice (11) of $\hat{W}_i^v$ yields a Davidson and MacKinnon (1981) type test of a non-nested alternative while the second one (12) corresponds to a Cox (1961, 1962) type test of a non-nested alternative. It seems that the Cox-like form of the test is generally more powerful than the Davidson-like form. Be that as it may, such tests may for example be used for checking the chosen variance functions $\phi_\nu(.)$ and $\phi_\mu(.)$ against some other possible functional forms, or more generally for checking the assumed heteroscedastic model against any other non-nested specification for the scedastic structure of the data.

As was the case in our discussion of conditional mean testing, when a test against a specific nested or non-nested alternative rejects the null hypothesis H_0^v, it is natural for one to consider modifying the originally assumed conditional variance specification in the direction of the considered alternative. Likewise, in both the nested and non-nested cases, the way to perform the tests is unchanged if the alternative includes variables which are not functions of the original set of conditioning variables $CV_i \equiv (X_i, Z_i^1, Z_i^2)$. But similarly, the tested null hypothesis is modified. It here takes the form $\mathrm{H}_0^{v\prime}$: H_0^v holds and, both $E(Y_i|X_i, Z_i^1, Z_i^2, \underline{G}_i) = E(Y_i|X_i, Z_i^1, Z_i^2)$ and $V(Y_i|X_i, Z_i^1, Z_i^2, \underline{G}_i) = V(Y_i|X_i, Z_i^1, Z_i^2)$, $i = 1, 2, ..., n$, where $\underline{G}_i$ denotes the variables which are not functions of CV_i. In other words, besides H_0^v, $\mathrm{H}_0^{v\prime}$ further assumes that the additional variables $\underline{G}_i$ are irrelevant as conditioning variables for the variance but also for the expectation of Y_i.

Beside tests against nested and non-nested alternatives, general purpose diagnostic tests with expected power against a broader range of alternatives may be performed through Hausman and information matrix type tests.

Testing H_0^v through a Hausman type test requires one to choose a consistent estimator of γ^o alternative to $\hat{\gamma}_n$. As already suggested, the GPML2 estimator $\hat{\gamma}_n$ may be shown to be asymptotically equivalent to the weighted nonlinear least squares (NLS) estimator with weights $\left\{\tilde{\Gamma}_i^{-1}\right\}$ of the multivariate nonlinear regression $\mathrm{vec}(\tilde{u}_i\tilde{u}_i') = \mathrm{vec}(\mathrm{diag}(\phi_\nu(Z_i^1\gamma_1)) + J_{T_i}\, \phi_\mu(Z_i^2\gamma_2)) +$ residuals, $i = 1, 2, ..., n$, where the superscript '$\sim$' denotes quantities evaluated at any preliminary consistent estimator of β^o and γ^o. A straightforward and natural alternative to it is hence to use the standard (i.e. unweighted) NLS estimator, say $\underline{\hat{\gamma}}_n$, of the same nonlinear regression. Accordingly, a relevant Hausman type test of H_0^v may be obtained by checking, for some chosen selection matrix S, the closeness to zero of the misspecification indicator

$$\hat{\Phi}_n^v = S(\hat{\gamma}_n - \underline{\hat{\gamma}}_n)$$

Following the lines of White (1994) and Lejeune (1998), a test asymptotically equivalent to checking the above misspecification indicator is obtained by setting

$$\hat{W}_i^v = \hat{\Gamma}_i \frac{\partial\, \mathrm{vec}\, \hat{\Omega}_i}{\partial \gamma'} \hat{Q}^{-1} S'$$

where $\hat{Q} = \sum_{i=1}^n \left(\frac{\partial\, \mathrm{vec}\, \hat{\Omega}_i}{\partial \gamma'}\right)' \frac{\partial\, \mathrm{vec}\, \hat{\Omega}_i}{\partial \gamma'}$. As with all Hausman type tests, this test will have power against any alternative H_1^v for which $\hat{\gamma}_n$ and $\underline{\hat{\gamma}}_n$ converge to different pseudo-true values.

On the other hand, following again the lines of White (1994) and Lejeune (1998), an information matrix type test of H_0^v may be based on checking, for some chosen selection matrix S which at least removes its otherwise obvious redundant elements, the closeness to zero of the misspecification indicator

$$\hat{\Phi}_n^v = S \frac{1}{n} \sum_{i=1}^n \mathrm{vec}\left(\hat{s}_i^\beta \hat{s}_i^{\beta\prime} + \underline{\hat{h}}_i^{\beta\beta}\right)$$

Such a test basically means checking the information matrix equality $B_{\beta\beta} = -\bar{A}_{\beta\beta}$ for the mean parameters, which must hold under correct conditional mean and conditional variance specification. It is obtained by setting

$$\hat{W}_t^v = (X_i \otimes X_i)S'$$

This latter way of testing H_0^v without resorting to explicit alternatives, which seems generally more powerful than the above Hausman type test, will clearly have power against any alternative H_1^v for which the mean parameters information matrix equality fails.

As in conditional mean testing, when a Hausman or information matrix type test rejects H_0^v, the way to react is not obvious and depends on the situation at hand. But in all cases, considering further diagnostic tests against various nested or non-nested alternatives should likewise help to identify the source(s) of rejection of H_0^v.

5. An empirical illustration

We hereafter illustrate the potential usefulness of our proposed full heteroscedastic model and its accompanying robust inferential methods through an empirical example which involves estimating and testing at an inter-sectorial level the correctness of the specification of a transcendental logarithmic (translog) production model for a sample of 824 French firms observed over the period 1979 - 1988. As we will see, the results of this exercise suggest (a) that, as argued in Baltagi and Griffin (1988), heteroscedasticity-related problems are likely to be present when estimating this kind of production model, (b) that our proposed full heteroscedastic model and its accompanying robust inferential methods offer a sensible, although imperfect, way to deal with it, and (c) that a judicious use of the set of proposed specification tests allows one to obtain very informative insights regarding the empirical correctness of this simple production model.

5.1. Data and model

The data originally came from a panel dataset constructed by the "Marchés et Stratégie d'Entreprises" division of INSEE. It contains 5 201 observations and involves in an unbalanced panel of 824 French firms from 9 sectors[7] of the NAP 15 Classification observed over the period 1979 - 1988[8]. Available data include the value added (va) of the firms deflated by an NAP 40 sector-specific price index (base: 1980), their stock of capital (k) and their labor force (l). The stock of capital has been constructed by INSEE and the labor force is the number of workers expressed in full-time units.

As is usual in this kind of dataset, the variability of the observations essentially

[7] Agricultural and food industries, energy production and distribution, intermediate goods industries, equipment goods industries, consumption goods industries, construction and civil engineering, trade, transport and telecommunications, and market services.

[8] I wish to thank Patrick Sevestre for giving me the opportunity to use this dataset.

lies in the between (across individuals) dimension and is very important : the number of workers ranges from 19 to almost 32 000 and the capital intensity (k/l) varies from a factor of 1 to more than 320. Globally, large firms are over-represented.

For this dataset, we considered estimating and testing the following full heteroscedastic one-way error components translog production function model :

$$V_{it} = \beta_{(sc\times t)} + \beta_k K_{it} + \beta_l L_{it} + \beta_{kk} K_{it}^2 + \beta_{ll} L_{it}^2 + \beta_{kl} K_{it} L_{it} + \mu_i + \nu_{it} \qquad (13)$$

with

$$\sigma^2_{\nu_{it}} = \exp\left(\gamma_1^c + \gamma_1^k K_{it} + \gamma_1^l L_{it}\right) \qquad (14)$$

$$\sigma^2_{\mu_i} = \exp\left(\gamma_2^c + \gamma_2^k \bar{K}_i + \gamma_2^l \bar{L}_i\right) \qquad (15)$$

where

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

$$\bar{K}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} K_{it} \quad \text{and} \quad \bar{L}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} L_{it}$$

The subscript '$(sc \times t)$' attached to the intercept parameter $\beta_{(sc\times t)}$ means that we actually let the intercept be sectorial and time-period specific. The model thus contains 90 dummies (9 sectors $\times$ 10 periods). This allows for sector-specific productivity growth patterns.

The explanatory variables are centered so that the estimated values of β_k and β_l reported below may directly be interpreted as the elasticities of the value added with respect to capital and labor at $k = k^*$ and $l = l^*$. We set k^* and l^* at their entire sample means.

For both the individual-specific and general error variance functions, we adopted Harvey's (1976) multiplicative heteroscedasticity formulation. In the general error variance function, the explanatory variables are simply taken as the (log of the) capital and labor inputs. Taking the individual mean values of the (log of the) capital and labor inputs as explanatory variables in the individual-specific variance function is mainly a pragmatic choice. It appears sensible as far as the observations variability prominently lies in the between dimension. Be that as it may, these choices allow the variances to change according to both size and input ratios.

5.2. Estimation and specification testing

The results of GPML2 estimation of model (13)-(15) are reported in Table 1[9]. As it seems natural when first estimating the model, the covariance matrix of the parameters was first computed supposing correct conditional mean specification but possibly misspecified conditional variance, i.e. as the empirical counterpart of C_n^{o*}, or more precisely as the empirical counterpart of C_n^{o*} for the mean parameters and as the empirical counterpart of the outlined upper bound (thus allowing Wald conservative tests) of C_n^{o*} for the variance parameters (see Section 3.2). The standard

[9] For conciseness, the dummy parameter estimates are not reproduced.

errors reported in Table 1 are derived from this first estimated covariance matrix.

Table 1 : GPML2 estimates and diagnostic tests

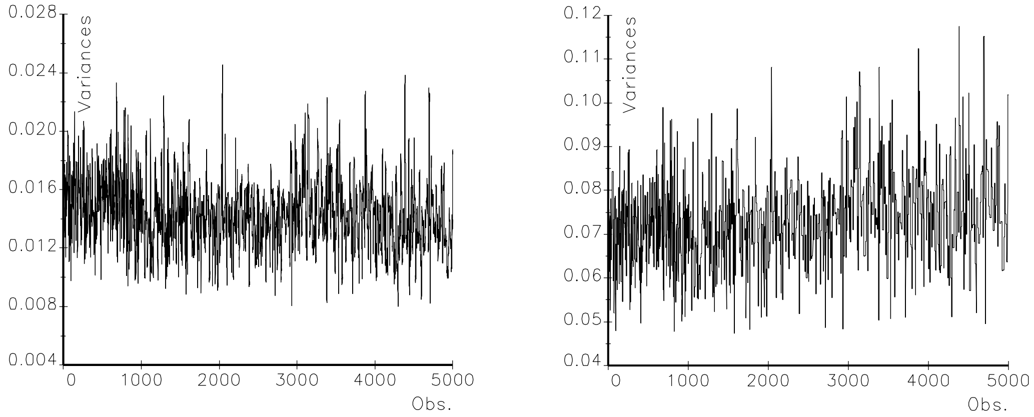
Variable	Coefficient	Std. error*	t-ratio	P-value
K	0.2487	0.0188	13.26	0.0000
L	0.7367	0.0244	30.21	0.0000
K^2	0.0547	0.0072	7.58	0.0000
L^2	0.0572	0.0132	4.35	0.0000
KL	-0.1137	0.0176	-6.48	0.0000
$\sigma^2_{\nu_{it}} = \exp(.)$				
const.	-4.1997	0.0541	-77.65	0.0000
K	0.1870	0.0582	3.21	0.0013
L	-0.2482	0.0849	-2.92	0.0035
$\sigma^2_{\mu_i} = \exp(.)$				
const.	-2.5213	0.0732	-34.43	0.0000
$\bar{K}$	0.1676	0.0610	2.74	0.0060
$\bar{L}$	-0.1709	0.0799	-2.14	0.0325
		Stat.	D.f.	P-value
Conditional mean tests				
(1) Hausman		5.9	5	0.3180
(2) Information matrix		33.7	25	0.1141
(3) H_1: non-neutral TP		8.4	2	0.0146
(4) H_1: third power		2.8	4	0.5961
(5) H_1: time heterogeneity		57.1	45	0.1064
(6) H_1: sectorial heterogeneity		41.0	40	0.4249
Conditional variance tests				
(7) Hausman		18.4	6	0.0052
(8) Information matrix		45.6	15	0.0001
(9) H_1: second power		2.2	6	0.9015
(10) H_1: sectorial heterogeneity		98.6	48	0.0000

*Standard errors computed assuming correct conditional mean specification but possibly misspecified conditional variance

As is apparent from Table 2 and confirmed when formally performing a (conservative) Wald test of the null hypothesis that the non-intercept parameters of both individual-specific and general variance functions are zero (P-value of the test : 0.0008), it appears that heteroscedasticity-like patterns are effectively present in both the individual-specific and general errors of the model. In both cases, heteroscedasticity seems to be related to input ratios : more capital intensive firms tend to achieve more heterogeneous outputs both in the between and within dimensions relative to the more labor intensive firms. The captured heteroscedasticity does not however seem to be notably related to size. Figure 1 portrays this latter point. In this figure, estimated general error and individual-specific error variances are graphed against the observations sorted in ascending order according to individual means of the fitted dependent variable and, within each individual, according to the values of the fitted dependent variable itself.

Figure 1 : Estimated variances versus size

General error variances : $\hat{\sigma}^2_{\nu_{it}}$ Individual error variances : $\hat{\sigma}^2_{\mu_i}$



Neither of these plots reveal notable links between variances and size. They do however outline two other points. First, variations in the observed inputs ratios imply variations in the estimated variances — identified by the difference between the lower and upper levels of the estimated variances — of more than a factor 2. Second, the estimated individual-specific variances are roughly 5 - 6 times higher than the estimated general error variances.

Having estimated the model, we next checked the correctness of its specification, considering first its conditional mean specification. To this end, we performed both Hausman and information matrix type tests and tests against nested alternatives. For the record, Hausman and information matrix type tests may be viewed as general purpose diagnostic tests allowing one to in particular detect unforeseen forms of misspecification, while tests against nested alternatives constitute a standard device for detecting a priori well-defined and plausible forms of misspecification.

In the present case, we considered a Hausman test based on comparing the GPML2 and OLS estimators of all mean parameters (excepted the dummies) and an information matrix test based on checking the closeness to zero of the sub-block of the Hessian corresponding to the cross-derivatives between the non-intercept mean parameters and all variance parameters (except for the intercept of the individual-specific variance function, to avoid singularity (cf. Section 4.1)). On the other hand, we considered tests against nested alternatives checking for possible non-neutral technical progress (the alternative model including as additional variables the interactions between a trend and the first order terms of the tranlog function[10]), for a possible more general functional form (the alternative model including terms of third power[11] as additional variables to the null translog specification), for possible time heterogeneity (the alternative model allowing for the non-intercept mean parameters to be time-period specific), and finally for possible sectorial heterogeneity (the alternative model allowing for the non-intercept mean parameters to be

[10] Non-neutral technical progress is typically modelled by considering a trend, a trend-squared and interaction terms between the trend and the first order terms of the translog function as additional inputs. The trend and trend-squared terms being already captured by the set of dummies, it thus remains to test for the interaction terms between a trend and the first order terms of the translog function.

[11] i.e. K^3, L^3, KL^2 and K^2L.

sector-specific).

Table 1 reports the results obtained from the computation of these conditional mean diagnostic tests[12]. As may be seen, it appears that the conditional mean does not exhibit patent misspecification. The only statistic which indicates some possible deviation from correct specification is the one of test (3). Its P-value is however not really worrying: from a formal point of view, according to a standard Bonferroni approach, for rejecting at 5% the null hypothesis that the conditional mean is correctly specified, we "need" that at least one of the 6 separate tests rejects the null at 0.83 % ($0.05/6 \simeq 0.0083$). Viewed in a less formal way, it is normal to find that some statistics (moderately) deviate when multiplying the number of diagnostic tests. The model may thus sensibly be viewed as a satisfactory statistical representation — on which for example productivity growth measurements could be based — of the available data for the conditional mean.

Taking correct conditional mean specification of the model for granted, we then examined its conditional variance. To this end, as for the conditional mean, we performed general purpose Hausman and information matrix type tests and tests against nested alternatives. Practically, we considered a Hausman test based on comparing the GPML2 and (unweighted) NLS estimators of all variance parameters and an information matrix test based on checking the closeness to zero of the non-redundant elements of the sub-block of the information matrix equality associated with the non-intercept mean parameters. On the other hand, we considered tests against nested alternatives checking for a possible more general functional form (the alternative model specifying both the individual-specific and general error variances as (the exponential of) translog functions instead of Cobb-Douglas like functions) and for possible sectorial heterogeneity (the alternative model allowing for all variance parameters to be sector-specific).

Before examining the results of these tests[13], note that the fact of finding no patent misspecification in the conditional mean supports the validity of the (conservative) standard errors of the variance parameter estimates reported in Table 1. These standard errors — and further the result of the outlined formal (conservative) Wald test of the null hypothesis that the non-intercept parameters of the individual-specific and general variance functions are zero — undoubtedly indicate that a heteroscedasticity-like pattern is effectively present in the errors of the model. However, according to the conditional variance tests reported in the same table, the assumed specification for this heteroscedasticity-like pattern turns out to be seriously misspecified. Test (9) suggests that relaxing the functional form would not really help. On the other hand, test (10) points out that a problem of sectorial heterogeneity might be involved.

To shed light on the latter point as well as to gauge the sensibility of the conditional mean estimates and diagnostic tests to the specification of the conditional variance, Table 2 reports GPML2 estimates and diagnostic tests — the same tests

[12] Note that none of these diagnostic tests involves variables which are not a function of the original set of conditioning variables (i.e. K, L, sector dummies and time dummies). The null hypothesis of these tests is thus never more than H_0^m itself (cf. Section 4.1).

[13] Note again that none of these diagnostic tests involves variables which are not a function of the original set of conditioning variables. The null hypothesis of these tests is thus again never more than H_0^v itself (cf. Section 4.2).

as above — of an extension of model (13)-(15), where both the individual-specific and the general error variance parameters are allowed to be sector-specific.

As may be seen from Table 2, the obtained mean parameter estimates are not very different from those obtained under the assumption of identical variances across sectors (cf. Table 1). For conciseness, the variance parameter estimates are not reported. But, as expected, they unambiguously confirm both that a heteroscedasticity-like pattern related to input ratios is present, and that this heteroscedasticity-like pattern is indeed sector-specific.

Table 2: GPML2 estimates and diagnostic tests with sector-specific conditional variances

Variable	Coefficient	Std. error*	t-ratio	P-value
K	0.2455	0.0169	14.54	0.0000
L	0.7519	0.0210	35.77	0.0000
K^2	0.0557	0.0062	9.03	0.0000
L^2	0.0639	0.0101	6.29	0.0000
KL	-0.1165	0.0148	-7.87	0.0000
		Stat.	D.f.	P-value
Conditional mean tests				
(1) Hausman		6.5	5	0.2579
(2) Information matrix		38.7	25	0.0396
(3) H_1: non-neutral TP		3.9	2	0.1446
(4) H_1: third power		3.3	4	0.5061
(5) H_1: time heterogeneity		55.6	45	0.1341
(6) H_1: sectorial heterogeneity		36.0	40	0.6505
Conditional variance tests				
(7) Hausman		72.1	50	0.0221
(8) Information matrix		52.8	15	0.0000

*Standard errors computed assuming correct conditional mean specification but possibly misspecified conditional variance

The diagnostic tests reported in Table 2 corroborate our result that the conditional mean of the model does not exhibit patent misspecification. However, they also show that allowing for sector-specific variance functions did not solve our misspecification problem in the conditional variance. How to fix this misspecification does not appear to be a trivial exercise.

Note nevertheless that, even if misspecified, these sector-specific variance functions are not useless. Comparing the standard errors of the mean parameters reported in Table 1 and 2, it may indeed be seen that allowing for this more flexible conditional variance specification has entailed (moderate) efficiency gains: the reduction of the standard errors ranges from -10.1% to -23.4%. This illustrates that, as argued in Section 3.2, a misspecified conditional variance may get efficiency benefits — for estimation but also testing of the conditional mean — from taking into account even approximately the actual scedastic structure of the data.

6. Conclusion

This paper proposed an extension of the standard one-way error components model allowing for heteroscedasticity in both the individual-specific and the general error terms, as well as for unbalanced panel. On the grounds of its computational convenience, its ability to straightforwardly handle unbalanced panels, its potential efficiency, its robustness to non-normality and its robustness to possible misspecification of the assumed scedastic structure of the data, we argued for estimating this model by Gaussian pseudo-maximum likelihood of order two. We further reviewed how, taking advantage of the powerful m-testing framework, the correct specification of the prominent aspects of the assumed full heteroscedastic model may be tested. We finally illustrated the practical relevance of our proposed model and estimation and diagnostic testing procedures through an empirical example.

To conclude, note that, since our proposed model contains as a special case the standard one-way error components model (just let the Z_i^1 and Z_i^2 variables only contain an intercept), our proposed integrated statistical tool-box, for which an easy-to-use Gauss package is available upon request from the author, may actually also be used for estimating and checking the specification of this standard model. On the other hand, remark that, following the lines of this paper, our proposed integrated statistical tool-box may readily be adapted to handle a more general model, for example allowing for a nonlinear (instead of linear) specification in the conditional mean and/or any fully nonlinear (instead of semi-linear) specification in the conditional variance.

Appendix A

Closed-form expressions for $|\Omega_i|$ and Ω_i^{-1} are given by

$$\begin{aligned}
|\Omega_i| &= (b_i)^{T_i}\,|C_i|\left(1+\operatorname{tr}C_i^{-1}\right) = \left(\prod_{t=1}^{T_i} a_{it}\right)\left(1+e_{T_i}'\bar{c}_i\right) \\
\Omega_i^{-1} &= \frac{1}{b_i}\left(C_i^{-1} - \frac{1}{1+\operatorname{tr}C_i^{-1}}\left(C_i^{-1}J_{T_i}C_i^{-1}\right)\right) \\
&= \operatorname{diag}\left(\bar{a}_i\right) - \frac{1}{b_i\left(1+e_{T_i}'\bar{c}_i\right)}\bar{c}_i\bar{c}_i'
\end{aligned}$$

where

$$\begin{array}{lll}
b_i = \phi_\mu(Z_i^2\gamma_2) & c_i = \dfrac{1}{b_i}\phi_\nu(Z_i^1\gamma_1) & a_i = \phi_\nu(Z_i^1\gamma_1) \\
C_i = \operatorname{diag}\left(c_i\right) & \bar{c}_i = e_{T_i} \div c_i & \bar{a}_i = e_{T_i} \div a_i
\end{array}$$

a_{it} being the t-th element of a_i and $\div$ indicating an element-by-element division. Note that according to this notation, $\Omega_i = b_i\left(C_i + J_{T_i}\right)$.

Following Magnus (1978, 1988), the first derivatives of $L_n(\beta, \gamma_1, \gamma_2)$ may be written

$$\frac{\partial L_n}{\partial \theta} = \frac{1}{n}\sum_{i=1}^{n} s_i^{\theta}, \qquad s_i^{\theta} = \begin{bmatrix} s_i^{\beta} \\ s_i^{\gamma_1} \\ s_i^{\gamma_2} \end{bmatrix}$$

with

$$\begin{aligned} s_i^{\beta} &= X_i'\Omega_i^{-1}u_i && \text{(A-1)} \\ s_i^{\gamma_p} &= \frac{1}{2}\left(\frac{\partial \operatorname{vec}\Omega_i}{\partial \gamma_p'}\right)'\left(\Omega_i^{-1}\otimes\Omega_i^{-1}\right)\operatorname{vec}(u_iu_i' - \Omega_i) \qquad (p=1,2) && \text{(A-2)} \\ &= \frac{1}{2}\sum_{r=1}^{l_p} \operatorname{tr}\left(\Omega_i^{-1}\frac{\partial \Omega_i}{\partial \gamma_p^r}\Omega_i^{-1}(u_iu_i' - \Omega_i)\right) e_{l_p}^r \end{aligned}$$

where $e_{l_p}^r$ is a $l_p \times 1$ vector with a one in the r-th place and zeros elsewhere, i.e. the r-th column of a $l_p \times l_p$ identity matrix, γ_p^r is the r-th component of γ_p, and the derivatives of $\operatorname{vec}\Omega_i$ with respect to γ_p' $(p = 1, 2)$ are

$$\begin{aligned} \frac{\partial \operatorname{vec}\Omega_i}{\partial \gamma_1'} &= \operatorname{diag}\left(\operatorname{vec}\left(\operatorname{diag}\left(\phi_\nu'(Z_i^1\gamma_1)\right)\right)\right)\left(Z_i^1 \otimes e_{T_i}\right) && \text{(A-3)} \\ \frac{\partial \operatorname{vec}\Omega_i}{\partial \gamma_2'} &= \phi_\mu'(Z_i^2\gamma_2)\operatorname{vec}\left(J_{n_i}\right) Z_i^2 = \phi_\mu'(Z_i^2\gamma_2)\left(e_{T_i}\otimes e_{T_i}\right) Z_i^2 && \text{(A-4)} \end{aligned}$$

while the derivatives of Ω_i with respect to γ_p^r $(p = 1, 2)$ are

$$\frac{\partial \Omega_i}{\partial \gamma_1^r} = \operatorname{diag}\left(\phi_\nu'(Z_i^1\gamma_1)\odot Z_i^{1^r}\right) \quad \text{and} \quad \frac{\partial \Omega_i}{\partial \gamma_2^r} = \phi_\mu'(Z_i^2\gamma_2)Z_i^{2^r}J_{T_i} \qquad \text{(A-5)}$$

where $\phi_\nu'(.)$ and $\phi_\mu'(.)$ denote the first derivatives of $\phi_\nu(.)$ and $\phi_\mu(.)$, $Z_i^{1^r}$ is the r-th column of the matrix of explanatory variables Z_i^1, $\odot$ stands for the Hadamard product, i.e. an element-by-element multiplication, and $Z_i^{2^r}$ is the r-th column of the row vector of explanatory variables Z_i^2. Note that if the multiplicative heteroscedasticity formulation is adopted for both $\phi_\nu(.)$ and $\phi_\mu(.)$, then, in (A-3)-(A-5), $\phi_\nu'(.)$ and $\phi_\mu'(.)$ are simply equal to exp(.).

Following again Magnus (1978, 1988), the Hessian matrix of $L_n(\beta, \gamma_1, \gamma_2)$ may be written

$$\frac{\partial^2 L_n}{\partial \theta \partial \theta'} = \frac{1}{n}\sum_{i=1}^{n} h_i^{\theta\theta}, \qquad h_i^{\theta\theta} = \begin{bmatrix} h_i^{\beta\beta} & h_i^{\beta\gamma_1} & h_i^{\beta\gamma_2} \\ h_i^{\gamma_1\beta} & h_i^{\gamma_1\gamma_1} & h_i^{\gamma_1\gamma_2} \\ h_i^{\gamma_2\beta} & h_i^{\gamma_2\gamma_1} & h_i^{\gamma_2\gamma_2} \end{bmatrix}$$

with

$$\begin{aligned} h_i^{\beta\beta} &= \frac{\partial s_i^{\beta}}{\partial \beta'} = -X_i'\Omega_i^{-1}X_i && \text{(A-6)} \\ h_i^{\beta\gamma_p} &= \frac{\partial s_i^{\beta}}{\partial \gamma_p'} = \left(\frac{\partial s_i^{\gamma_p}}{\partial \beta'}\right)' = h_i^{\gamma_p\beta\prime} \qquad (p=1,2) && \text{(A-7)} \end{aligned}$$

$$
\begin{aligned}
&= -\left(u_i'\Omega_i^{-1} \otimes X_i'\Omega_i^{-1}\right) \frac{\partial \operatorname{vec} \Omega_i}{\partial \gamma_p'} \\
&= -\sum_{r=1}^{l_p} \left(X_i'\Omega_i^{-1} \frac{\partial \Omega_i}{\partial \gamma_p^r} \Omega_i^{-1} u_i \right) e_{l_p}^{r\prime} \\
h_i^{\gamma_p \gamma_q} &= \frac{\partial s_i^{\gamma_p}}{\partial \gamma_q'} = \left(\frac{\partial s_i^{\gamma_q}}{\partial \gamma_p'} \right)' = h_i^{\gamma_q \gamma_p \prime} \qquad (p = 1, 2 \; ; \quad q = 1, 2) \qquad \text{(A-8)} \\
&= -\frac{1}{2} \left(\frac{\partial \operatorname{vec} \Omega_i}{\partial \gamma_p'} \right)' \left(\Omega_i^{-1} \otimes \Omega_i^{-1} \right) \frac{\partial \operatorname{vec} \Omega_i}{\partial \gamma_q'} - \frac{1}{2} \left((\operatorname{vec}(u_i u_i' - \Omega_i))' \otimes I_{l_p} \right) \Upsilon_i^{\gamma_p \gamma_q} \\
&= -\frac{1}{2} \sum_{r=1}^{l_p} \sum_{s=1}^{l_q} \left(\operatorname{tr} \left(\Omega_i^{-1} \frac{\partial \Omega_i}{\partial \gamma_p^r} \Omega_i^{-1} \frac{\partial \Omega_i}{\partial \gamma_q^s} \right) + \operatorname{tr} \left((u_i u_i' - \Omega_i) \frac{\partial^2 \Omega_i^{-1}}{\partial \gamma_p^r \partial \gamma_q^s} \right) \right) e_{l_p}^r e_{l_q}^{s\prime}
\end{aligned}
$$

where I_{l_p} is a $l_p \times l_p$ identity matrix,

$$
\Upsilon_i^{\gamma_p \gamma_q} = \frac{\partial \operatorname{vec} \left(\frac{\partial \operatorname{vec} \Omega_i^{-1}}{\partial \gamma_p'} \right)'}{\partial \gamma_q'} = \sum_{r=1}^{l_p} \sum_{s=1}^{l_q} \operatorname{vec} \left(\frac{\partial^2 \Omega_i^{-1}}{\partial \gamma_p^r \partial \gamma_q^s} \right) \otimes \left(e_{l_p}^r e_{l_q}^{s\prime} \right)
$$

i.e. a $T_i^2 l_p \times l_q$ matrix,

$$
\frac{\partial^2 \Omega_i^{-1}}{\partial \gamma_p^r \partial \gamma_q^s} = \Omega_i^{-1} \left(2 \frac{\partial \Omega_i}{\partial \gamma_p^r} \Omega_i^{-1} \frac{\partial \Omega_i}{\partial \gamma_q^s} - \frac{\partial^2 \Omega_i}{\partial \gamma_p^r \partial \gamma_q^s} \right) \Omega_i^{-1}
$$

and the needed derivatives not yet given are

$$
\begin{aligned}
\frac{\partial^2 \Omega_i}{\partial \gamma_1^r \partial \gamma_1^s} &= \operatorname{diag} \left(\phi_\nu'' \left(Z_i^1 \gamma_1 \right) \odot Z_i^{1^r} \odot Z_i^{1^s} \right) \\
\frac{\partial^2 \Omega_i}{\partial \gamma_1^r \partial \gamma_2^s} &= 0 = \frac{\partial^2 \Omega_i}{\partial \gamma_2^r \partial \gamma_1^s} \\
\frac{\partial^2 \Omega_i}{\partial \gamma_2^r \partial \gamma_2^s} &= \phi_\mu'' \left(Z_i^2 \gamma_2 \right) Z_i^{2^r} Z_i^{2^s} J_{T_i}
\end{aligned}
$$

where $\phi_\nu''(.)$ and $\phi_\mu''(.)$ denote the second derivatives of $\phi_\nu(.)$ and $\phi_\mu(.)$. If the multiplicative heteroscedasticity formulation is adopted for both $\phi_\nu(.)$ and $\phi_\mu(.)$, $\phi_\nu''(.)$ and $\phi_\mu''(.)$ are again simply equal to $\exp(.)$.

Under conditional mean and conditional variance correct specification, we have $E(u_i^o | X_i, Z_i^1, Z_i^2) = 0$ and $E\left((u_i^o u_i^{o\prime} - \Omega_i^o) | X_i, Z_i^1, Z_i^2\right) = 0$, so that using the law of iterated expectation it is easily checked that the expected Hessian matrix of $L_n(\beta, \gamma_1, \gamma_2)$ may be written

$$
E \left[\frac{\partial^2 L_n}{\partial \theta \partial \theta'} \right]_{\theta = \theta^o} = \frac{1}{n} \sum_{i=1}^n E \left[h_i^{\theta\theta} \right]_{\theta = \theta^o} = \frac{1}{n} \sum_{i=1}^n E \left[\underline{h}_i^{\theta\theta} \right]_{\theta = \theta^o}
$$

where

$$\underline{h}_i^{\theta\theta} = \begin{bmatrix} \underline{h}_i^{\beta\beta} & 0 & 0 \\ 0 & \underline{h}_i^{\gamma_1\gamma_1} & \underline{h}_i^{\gamma_1\gamma_2} \\ 0 & \underline{h}_i^{\gamma_2\gamma_1} & \underline{h}_i^{\gamma_2\gamma_2} \end{bmatrix}$$

and

$$\underline{h}_i^{\beta\beta} = h_i^{\beta\beta} = -X_i'\Omega_i^{-1}X_i \tag{A-9}$$

$$\begin{aligned} \underline{h}_i^{\gamma_p\gamma_q} &= \underline{h}_i^{\gamma_q\gamma_p\prime} \qquad (p = 1,2\ ;\quad q = 1,2) \qquad\qquad (\text{A-10}) \\ &= -\frac{1}{2}\left(\frac{\partial\,\mathrm{vec}\,\Omega_i}{\partial\gamma_p'}\right)'\left(\Omega_i^{-1}\otimes\Omega_i^{-1}\right)\frac{\partial\,\mathrm{vec}\,\Omega_i}{\partial\gamma_q'} \\ &= -\frac{1}{2}\sum_{r=1}^{l_p}\sum_{s=1}^{l_q}\mathrm{tr}\left(\Omega_i^{-1}\frac{\partial\Omega_i}{\partial\gamma_p^r}\Omega_i^{-1}\frac{\partial\Omega_i}{\partial\gamma_q^s}\right)e_{l_p}^r e_{l_q}^{s\prime} \end{aligned}$$

Note that contrary to the Hessian which depends on first and second derivatives, the expected Hessian is block-diagonal (between mean and variance parameters) and only depends on first derivatives.

Appendix B

For Gaussian maximum likelihood estimation of the standard (homoscedastic) one-way error components model, Breusch (1987) suggests an iterated GLS procedure. Although applicable in very general situations (see Magnus (1978)), in the present case it is not very attractive since it implies at each step the (numerical) resolution of a set of nonlinear equations defined by the first-order conditions $\frac{\partial L_n}{\partial\gamma_p} = 0$ $(p = 1,2)$.

As alternatives, we can use either a Newton or quasi-Newton (secant methods) algorithm. While the former requires the computation of the first and second derivatives, the latter (for example, the so-called Davidson-Fletcher-Powell and Broyden-Fletcher-Goldfard-Shanno methods) requires only the computation of the first derivatives (see Quandt (1983)). In the present case, a variant of the Newton method appears particularly appealing, namely the scoring method. This variant simply involves substituting the Hessian $\frac{\partial^2 L_n}{\partial\theta\partial\theta'} = \frac{1}{n}\sum_{i=1}^n h_i^{\theta\theta}$ used in the Newton algorithm by the empirical counterpart of its expectation under conditional mean and conditional variance correct specification, i.e. by $\frac{1}{n}\sum_{i=1}^n \underline{h}_i^{\theta\theta}$. As noted above in Appendix A, the latter is considerably simpler: it is block-diagonal and only involves first derivatives. It will be a good approximation of the Hessian if the model is correctly specified and θ is not too far from θ^o. According to our experience, even under quite severe misspecification, provided that all quantities are analytically computed, the scoring method generally converges in less time (more computation time per iteration but fewer iterations) than the secant methods. Further, since the empirical expected Hessian is always negative semidefinite, it is numerically stable.

A sensible set of starting values for the above algorithm may be computed by proceeding as follows:

1- Obtain the $\hat{\underline{\beta}}$ and $\widehat{\alpha} = (\widehat{\alpha}_1, ..., \widehat{\alpha}_i, ..., \widehat{\alpha}_n)$ OLS estimates of the dummy variables model $Y_i = \alpha_i + \underline{X}_i\underline{\beta}+$ residuals $(i = 1, 2, ..., n)$, where $\underline{X}_i$ is the same as X_i except its dropped first column. At this stage, $\hat{\underline{\beta}}$ and the mean of the $\widehat{\alpha}_i$, i.e. $\overline{\alpha} = \frac{1}{n}\sum_{i=1}^{n}\hat{\alpha}_i$, provide initial values for β. Note that in practice $\hat{\underline{\beta}}$ and $\widehat{\alpha}_i$ may be computed as $\hat{\underline{\beta}} = (\sum_{i=1}^{n}\underline{X}_i' M_{T_i}\underline{X}_i)^{-1}\sum_{i=1}^{n}\underline{X}_i' M_{T_i} Y_i$ (within OLS estimator) and $\widehat{\alpha}_i = \frac{1}{T_i}e_{T_i}'(Y_i - \underline{X}_i\hat{\underline{\beta}})$, where $M_{T_i} = I_{T_i} - \frac{1}{T_i}J_{T_i}$, i.e. the within transformation matrix. See Balestra (1996) for details.

2- Run the OLS regression $\phi_\nu^{-1}(\hat{\underline{u}}_{it}^2) = Z_{it}^1\gamma_1+$ residuals $(i = 1, 2, ..., n; t = 1, 2, ..., T_i)$, where $\hat{\underline{u}}_{it} = Y_{it} - \hat{\alpha}_i - \underline{X}_{it}\hat{\underline{\beta}}$ and $\phi_\nu^{-1}(.)$ is the (supposed well-defined) inverse function of $\phi_\nu(.)$. The non-intercept parameters of $\hat{\gamma}_1$ and the intercept parameter of $\hat{\gamma}_1$ minus γ_{1_c}, where γ_{1_c} is an intercept correction term, give initial values for γ_1. The desirability of an intercept correction of $\hat{\gamma}_1$ arises from the fact that, even if we suppose that $\hat{\underline{u}}_{it}$ is equal to the true disturbance ν_{it}, the (conditional) expectation of the error term in the above OLS regression is usually not zero (and even not necessarily a constant). The "optimal" value of the intercept correction term γ_{1_c} depends upon the functional form $\phi_\nu^{-1}(.)$ and the actual distribution of the ν_{it}. In the case of the multiplicative heteroscedasticity formulation where $\phi_\nu^{-1}(.)$ is simply equal to $\ln(.)$, a sensible choice is $\gamma_{1_c} = -1.2704$. This follows from the fact that $E[\ln(\nu_{it}^2) - \ln(\sigma_{\nu_{it}}^2)] = E[\ln(\nu_{it}^2/\sigma_{\nu_{it}}^2)] = -1.2704$ if $\nu_{it} \sim N(0, \sigma_{\nu_{it}}^2)$; see Harvey (1976).

3- Finally, run the OLS regression $\phi_\mu^{-1}((\hat{\alpha}_i - \overline{\alpha})^2) = Z_i^2\gamma_2+$ residuals $(i = 1, 2, ..., n)$, where $\phi_\mu^{-1}(.)$ is the (supposed well-defined) inverse function of $\phi_\mu(.)$. According to the same reasoning as above, the non-intercept parameters of $\hat{\gamma}_2$ and the intercept parameter of $\hat{\gamma}_2$ minus γ_{2_c}, where γ_{2_c} is an intercept correction term, give initial values for γ_2. In the case of the multiplicative heteroscedasticity formulation where $\phi_\mu^{-1}(.)$ is again equal to $\ln(.)$, γ_{2_c} should also be set to -1.2704.

Note that a simpler alternative to the step 2 and 3 is workable. It merely consists in computing the "mean variance components" $\widehat{\sigma}_\nu^2 = \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T_i}\hat{\underline{u}}_{it}^2$ and $\widehat{\sigma}_\mu^2 = \frac{1}{n}\sum_{i=1}^{n}(\hat{\alpha}_i - \overline{\alpha})^2$. The inverse function values $\phi_\nu^{-1}(\widehat{\sigma}_\nu^2)$ and $\phi_\mu^{-1}(\widehat{\sigma}_\mu^2)$ may then be used for the first elements (intercepts) of γ_1 and γ_2, their remaining elements being simply set equal to zero.

References

Balestra P. (1996), "Fixed effect models and fixed coefficient models", Ch. 3 in *The Econometrics of Panel Data*, eds. Mátyás L. and Sevestre P., Kluwer Academic Publishers, Dordrecht.

Baltagi B.H. (1988), "An alternative heteroscedastic error components model", Problem 88.2.2, *Econometric Theory*, 4, 349-350.

Baltagi B.H and Griffin J.M. (1988), "A generalized error component model with heteroscedastic disturbances", *International Economic Review*, 29 (4), 745-753.

Baltagi B.H. (1995), *Econometric Analysis of Panel Data*, John Wiley & Sons, New York.

Baltagi B.H., Bresson G. and Pirotte A. (2004), "Joint LM test for homoscedasticity in a one-way error component model", Working Paper, mimeo, Texas A&M University.

Bollerslev T. and Wooldridge J. (1992), "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances", *Econometric Reviews*, 11 (2), 143-172.

Breusch T.S. and Pagan A.R. (1979), "A simple test for heteroscedasticity and random coefficient variation", *Econometrica*, 47 (5), 1287-1297.

Breusch T.S. (1987), "Maximum likelihood estimation of random effects models", *Journal of Econometrics*, 36, 383-389.

Chamberlain G. (1987), "Asymptotic efficiency in estimation with conditional moment restrictions", *Journal of Econometrics*, 34, 305-334.

Cox D.R., (1961), "Tests of separate families of hypotheses", in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, Berkeley, 105-123.

Cox D.R., (1962), "Further results on tests of separate families of hypotheses", *Journal of the Royal Statistical Society*, series B, 406-424.

Davidson R. and MacKinnon J.G. (1981), "Several tests for model specification in the presence of alternative hypotheses", *Econometrica*, 49 (3), 781-793.

Gourieroux C., Monfort A. and Trognon A. (1984), "Pseudo-maximum likelihood methods: theory", *Econometrica*, 52 (3), 681-700.

Gourieroux C. and Monfort A. (1993), "Pseudo-likelihood methods", Ch. 12 in *Handbook of Statistics*, Vol. 11, eds. Maddala G.S., Rao C.R. and Vinod H.D., North Holland, Amsterdam.

Harvey A.C. (1976), "Estimating regression models with multiplicative heteroscedasticity", *Econometrica*, 44 (3), 461-465.

Holly A. and Gardiol L. (2000), "A score test for individual heteroscedasticity in a one-way error components model", Ch. 10 in *Panel Data Econometrics: Future Directions*, eds Kirshnakumar J. and Ronchetti E., Elsevier, Amsterdam.

Lejeune B. (1998), "Error components models and variable heterogeneity: modelisation, second order pseudo-maximum likelihood estimation and specification testing", Ph.D. dissertation, Université de Liège and Université de Paris XII - Val de Marne.

Lejeune B. (2004), "A distribution-free joint test and one-directional robust tests for random individual effects and heteroscedasticity allowing for unbalanced panels", Working paper, mimeo, University of Liège, Liège.

Li Q. and Stengos T. (1994), "Adaptive estimation on the panel data error component model with heteroscedasticity of unknown form", *International Economic Review*, 35, 981-1000.

Magnus J.R. (1978), "Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix", *Journal of Econometrics*, 7, 281-312. Corrigenda, *Journal of Econometrics*, 10, 261.

Magnus J.R. (1982), "Multivariate error components analysis of linear and non-linear regression models by maximum likelihood", *Journal of Econometrics*, 19, 239-285.

Magnus J.R. and Neudecker H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons, New York.

Mazodier P. and Trognon A. (1978), "Heteroscedasticity and stratification in error components models", *Annales de l'INSEE*, 30-31, 451-482.

Newey W. (1985), "Maximum likelihood specification testing and conditional moment tests", *Econometrica*, 53 (5), 1047-1070.

Newey W. (1990), "Semiparametric efficiency bounds", *Journal of Applied Econometrics*, 5, 99-135.

Newey W. (1993), "Efficient estimation of models with conditional moment restrictions", Ch. 16 in *Handbook of Statistics*, Vol. 11, eds. Maddala G.S., Rao C.R. and Vinod H.D., North Holland, Amsterdam.

Phillips R.L. (2003), "Estimation of a stratified error components model", *International Economic Review*, 44, 501-521.

Quandt R.E. (1983), "Computational problems and methods", Ch. 12 in *Handbook of Econometrics*, Vol. 1, eds. Griliches Z. and Intriligator M.D., North Holland, Amsterdam.

Randolph W.C. (1988), "A transformation for heteroscedastic error components regression models", *Economics Letters*, 27, 349-354.

Rao S.R.S., Kaplan J. and Cochran W.G. (1981), "Estimators for the one-way random effects model with unequal error variances", *Journal of the American Statistical Association*, 76, 89-97.

Roy N. (2002), "Is adaptive estimation useful for panel models with heteroscedasticity in the individual specific error component ? Some Monte Carlo evidence", *Econometric Reviews*, 21, 189-203.

Tauchen G. (1985), "Diagnostic testing and evaluation of maximum likelihood models", *Journal of Econometrics*, 30, 415-443.

Verbon H.A.A. (1980), "Testing for heteroscedasticity in a model of seemingly unrelated regression equations with variance components", *Economics Letters*, 5, 149-153.

Wansbeek T. (1989), "An alternative heteroscedastic error components model", Solution 88.1.1, *Econometric Theory*, 5, 326.

White H. (1987), "Specification testing in dynamic models", Ch. 1 in *Advances in Econometrics - Fifth World Congress*, Vol. 1, eds. Bewley T., Cambridge University Press, New-York.

White H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge University Press, Cambridge.

Wooldridge J. (1990), "A unified approach to robust, regression-based specification tests", *Econometric Theory*, 6, 17-43.

Wooldridge J. (1991a), "On the application of robust, regression-based diagnostics to models of conditional means and conditional variances", *Journal of Econometrics*, 47, 5-46.

Wooldridge J. (1991b), "Specification testing and quasi-maximum likelihood estimation", *Journal of Econometrics*, 48, 29-55.

Wooldridge J. (1994), "Estimation and inference for dependent processes", Ch. 45 in *Handbook of Econometrics*, Vol. 4, eds. Engle R. and McFadden D., North Holland, Amsterdam.

Wooldridge J. (1995), "Selection corrections for panel data models under conditional mean independence assumptions", *Journal of Econometrics*, 68, 115-132.