Journal of Econometrics 154 (2010) 122–124

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

# Testing for heteroskedasticity and serial correlation in a random effects panel data model

Badi H. Baltagi [a,*], Byoung Cheol Jung [b], Seuck Heun Song [c]

[a] Syracuse University, Department of Economics and Center for Policy Research, Syracuse, NY 13244–1020, USA

[b] Department of Statistics, University of Seoul, Dongdaemun-Gu, Seoul, 130-743, Republic of Korea

[c] Department of Statistics, Korea University, Sungbuk-Ku, Seoul, 136-701, Republic of Korea

ARTICLE INFO

*Article history:*
Received 23 January 2009
Received in revised form
23 January 2009
Accepted 21 April 2009
Available online 28 July 2009

*JEL classification:*
C23

*Keywords:*
Panel data
Heteroskedasticity
Serial correlation
Lagrange multiplier tests
Likelihood ratio
Random effects

ABSTRACT

This paper considers a panel data regression model with heteroskedastic as well as serially correlated disturbances, and derives a joint LM test for homoskedasticity and no first order serial correlation. The restricted model is the standard random individual error component model. It also derives a conditional LM test for homoskedasticity given serial correlation, as well as, a conditional LM test for no first order serial correlation given heteroskedasticity, all in the context of a random effects panel data model. Monte Carlo results show that these tests along with their likelihood ratio alternatives have good size and power under various forms of heteroskedasticity including exponential and quadratic functional forms.



## 1. Introduction

The standard error component panel data model assumes that the disturbances have homoskedastic variances and constant serial correlation through the random individual effects, see Hsiao (2003). These may be restrictive assumptions for a lot of panel data applications. For example, the cross-sectional units may be varying in size and as a result may exhibit heteroskedasticity. Also, for investment behavior of firms, for example, an unobserved shock this period may affect the behavioral relationship for at least the next few periods. In fact, the standard error component model has been extended to take into account serial correlation by Baltagi and Li (1995), Galbraith and Zinde-Walsh (1995) and Bera et al. (2001) to mention a few. This model has also been generalized to take into account hetetoskedasticity by Li and Stengos (1994), Lejeune (1996), Holly and Gardiol (2000), Roy (2002) and Baltagi et al. (2006) to mention a few. For a review of these papers, see Baltagi (2008). However, these strands of literature are almost separate in the panel data error component literature. When one deals with heteroskedasticity, serial correlation is ignored, and when one deals with serial correlation, heteroskedasticity is ignored. Exceptions are robust estimation of the variance–covariance matrix of the reported estimates.

Baltagi and Li (1995) for example, derived a Lagrange Multiplier (LM) test which jointly tests for the presence of serial correlation as well as random individual effects assuming homoskedasticity of the disturbances. While, Holly and Gardiol (2000), for example, derived an LM statistic which tests for homoskedasticity of the disturbances in the context of a one-way random effects panel data model. The latter LM test assumes no serial correlation in the remainder disturbances. This paper extends the Holly and Gardiol (2000) model to allow for first order serial correlation in the remainder disturbances as described in Baltagi and Li (1995). It derives a joint LM test for homoskedasticity and no first order serial correlation. The restricted model is the standard random effects error component model. It also derives a conditional LM test for homoskedasticity given serial correlation, as well as, a conditional LM test for no first order serial correlation given heteroskedasticity. Monte Carlo results show that these tests along with their likelihood ratio alternatives have good size and power under various forms of heteroskedasticity including exponential and quadratic functional forms.

* Tel.: +1 315 443 1630; fax: +1 315 443 1081.
*E-mail address:* bbaltagi@maxwell.syr.edu (B.H. Baltagi).

0304-4076/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2009.04.009

## 2. The model and LM tests

Consider the following panel data regression model :

$$y_{it} = x_{it}'\beta + u_{it}, \quad i = 1, \ldots, N, \text{ and } t = 1, \ldots, T, \tag{1}$$

where $y_{it}$ is the observation on the dependent variable for the $i$th individual at the $t$th time period, $x_{it}$ denotes the $k \times 1$ vector of observations on the nonstochastic regressors. The regression disturbances of (1) are assumed to follow a one-way error component model

$$u_{it} = \mu_i + \nu_{it}, \tag{2}$$

where $\mu_i$ denote the random individual effects which are assumed to be normally and independently distributed with mean 0 and variance

$$\mathrm{Var}(\mu_i) = h(z_i'\alpha), \tag{3}$$

the function $h(\cdot)$ is an arbitrary non-indexed (strictly) positive twice continuously differentiable function, see Breusch and Pagan (1979). $\alpha$ is a $p \times 1$ vector of unrestricted parameters and $z_i$ is a $p \times 1$ vector of strictly exogenous regressors which determine the heteroskedasticity of the individual specific effects. The first element of $z_i$ is one, and without loss of generality, $h(\alpha_1) = \sigma_\mu^2$. Therefore, when the model is homoskedastic with $\alpha_2 = \alpha_3 = \cdots = \alpha_p = 0$, this model reduces to the standard random effects model, as in Holly and Gardiol (2000). In addition, we allow the remainder disturbances to follow an $AR(1)$ process: $\nu_{it} = \rho\nu_{i,t-1} + \epsilon_{it}$, with $|\rho| < 1$ and $\epsilon_{it} \sim IIN(0, \sigma_\epsilon^2)$, as described in Baltagi and Li (1995). The $\mu_i$'s are independent of the $\nu_{it}$'s, and $\nu_{i,0} \sim N(0, \sigma_\epsilon^2/(1-\rho^2))$.

The model considered generalizes the one-way error component model to allow for heteroskedastic individual effects a la Holly and Gardiol (2000) *and* for first order serially correlated remainder disturbances a la Baltagi and Li (1995). The model (1) can be rewritten in matrix notation as

$$y = X\beta + u, \tag{4}$$

where $y$ is of dimension $NT \times 1$, $X$ is $NT \times k$, $\beta$ is $k \times 1$ and $u$ is a $NT \times 1$. $X$ is assumed to be of full column rank. The disturbance in Eq. (2)can be written in vector form as:

$$u = (I_N \otimes \iota_T)\mu + \nu, \tag{5}$$

where $\iota_T$ is a vector of ones of dimension $T$, $I_N$ is an identity matrix of dimension $N$, $\mu' = (\mu_1, \ldots, \mu_N)$ and $\nu' = (\nu_{11}, \ldots, \nu_{1T}, \ldots, \nu_{N1}, \ldots, \nu_{NT})$. Under these assumptions, the variance–covariance matrix of $u$ can be written as

$$\Omega = E(uu') = \mathrm{diag}[h(z_i'\alpha)] \otimes J_T + I_N \otimes V, \tag{6}$$

where $J_T$ is a matrix of ones of dimension $T$, and $\mathrm{diag}[h(z_i'\alpha)]$ is a diagonal matrix of dimension $N \times N$ and $V$ is the familiar AR(1) covariance matrix. It is well established that the matrix

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

transforms the usual AR(1) model into serially uncorrelated disturbances. For panel data, this has to be applied for $N$ individuals, see Baltagi and Li (1995). The transformed regression disturbances are given by :

$$\begin{aligned} u^* &= (I_N \otimes C)u = (I_N \otimes C\iota_T)\mu + (I_N \otimes C)\nu \\ &= (1-\rho)(I_N \otimes \iota_T^\delta)\mu + (I_N \otimes C)\nu, \end{aligned} \tag{7}$$

where $C\iota_T = (1-\rho)\iota_T^\delta$ with $\iota_T^\delta = (\delta, \iota_{T-1}')$, and $\delta = \sqrt{\frac{1+\rho}{1-\rho}}$.

Therefore,the variance–covariance matrix of transformed model is given by

$$\begin{aligned} \Omega^* &= E(u^*u^{*\prime}) \\ &= \mathrm{diag}[h(z_i'\alpha)(1-\rho)^2] \otimes \iota_T^\delta\iota_T^{\delta\prime} + \mathrm{diag}[\sigma_\epsilon^2] \otimes I_T, \end{aligned} \tag{8}$$

since $(I_N \otimes C)E(\nu\nu')(I_N \otimes C') = \mathrm{diag}[\sigma_\epsilon^2] \otimes I_T$. Replace $J_T^\delta = \iota_T^\delta\iota_T^{\delta\prime}$ by its idempotent counterpart $d^2\bar{J}_T^\delta$, where $d^2 = \iota_T^{\delta\prime}\iota_T^\delta = \delta^2 + T - 1$, and $\bar{J}_T^\delta$ is by definition $J_T^\delta/d^2$. Also, replace $I_T$ by $E_T^\delta + \bar{J}_T^\delta$, where $E_T^\delta$ is by definition $I_T - \bar{J}_T^\delta$, and collect like terms, we get

$$\Omega^* = \mathrm{diag}[\lambda_i^2] \otimes \bar{J}_T^\delta + \mathrm{diag}[\sigma_\epsilon^2] \otimes E_T^\delta, \tag{9}$$

where $\lambda_i^2 = d^2(1-\rho)^2h(z_i'\alpha) + \sigma_\epsilon^2$, from which it is easy to infer, see Baltagi and Li (1995) that

$$\Omega^{*r} = \mathrm{diag}[(\lambda_i^2)^r] \otimes \bar{J}_T^\delta + \mathrm{diag}[(\sigma_\epsilon^2)^r] \otimes E_T^\delta, \tag{10}$$

where $r$ is an arbitrary scalar. $r = -1$ obtains the inverse, while $r = -\frac{1}{2}$ obtains $\Omega^{*-\frac{1}{2}}$. In addition, one gets, $|\Omega^*| = \Pi_{i=1}^N(\lambda_i^2)(\sigma_\epsilon^2)^{T-1}$.

The null hypothesis for jointly testing for no heteroskedasticity and no serial correlation in a random effects panel data model is given by

$$H_0^a : \alpha_2 = \cdots = \alpha_p = 0 \quad \text{and} \quad \rho = 0. \tag{11}$$

Under the null hypothesis $H_0^a$, the model reduces to the familiar one-way random effects error component model with no serial correlation nor heteroskedasticity, see Baltagi (2008). Due to space limitations, the accompanying working paper, Baltagi et al. (2008), contains all the proofs and derivations of the score and information matrix under normality of the disturbances. The resulting joint LM statistic for the hypothesis (11) is given by:

$$LM_a = \frac{1}{2}f'Z(Z'Z)^{-1}Z'f + \frac{T^2}{T^2C_{\rho\rho} - 2N(T-1)}D_a(\rho)^2, \tag{12}$$

where $Z$ is an $N \times p$ matrix of observations on the $p$ variables $z_k$, $k = 1, 2, \ldots, p$, each of dimension $N \times 1$, and $f = (f_1, \ldots, f_N)'$ with $f_i = [(\sum_{t=1}^T \tilde{u}_{it})^2/T\tilde{\sigma}_1^2] - 1$. $\tilde{u} = y - X\tilde{\beta}_{MLE}$ denote the restricted maximum likelihood residuals under the null hypothesis $H_0^a$, i.e., under a standard random effects panel data model with no serial correlation or heteroskedasticity. Hence, $\tilde{\beta}_{MLE}$ is the standard random effects maximum likelihood (ML) estimator of $\beta$, and $\tilde{\sigma}_\epsilon^2$ and $\tilde{\sigma}_\mu^2$ are the corresponding ML estimates of the variance components. Also, $\tilde{\sigma}_1^2 = T\tilde{\sigma}_\mu^2 + \tilde{\sigma}_\epsilon^2$; $C_{\rho\rho} = J_{\rho\rho} - \frac{2N(T-1)^2}{T^2}\frac{\tilde{\sigma}_\epsilon^4}{\tilde{\sigma}_1^4}$, and $J_{\rho\rho} = N[2a^2(T-1)^2 + 2a(2T-3) + T - 1]$, with $a = \frac{\tilde{\sigma}_\epsilon^2 - \tilde{\sigma}_1^2}{T\tilde{\sigma}_1^2}$.

$$\begin{aligned} D_a(\tilde{\rho}) = {} & \frac{N(T-1)}{T}\frac{\tilde{\sigma}_1^2 - \tilde{\sigma}_\epsilon^2}{\tilde{\sigma}_1^2} \\ & + \frac{\tilde{\sigma}_\epsilon^2}{2}\tilde{u}'[I_N \otimes (\bar{J}_T/\tilde{\sigma}_1^2 + E_T/\tilde{\sigma}_\epsilon^2)G(\bar{J}_T/\tilde{\sigma}_1^2 + E_T/\tilde{\sigma}_\epsilon^2)]\tilde{u} \end{aligned} \tag{13}$$

and $G$ is a bidiagonal matrix with zero elements everywhere including the main diagonal, but with the two adjacent diagonal elements all equal to one. Under the null hypothesis $H_0^a$, the LM statistic of (12) is asymptotically distributed as $\chi_p^2$.

The joint LM test is useful especially when one does not reject the null hypothesis $H_0^a$. However, if the null hypotheses is rejected, one can not infer whether the presence of heteroskedasticity, or the presence of serial correlation, or both factors caused this rejection. Here, we derive two conditional LM tests. The first one tests for the absence of serial correlation of the first order assuming that heteroskedasticity of the individual effects might be present. The second one tests for homoskedasticity assuming that serial correlation of the first order might be present. All in the context of a random effects panel data model.

$$\widehat{J}_b(\theta)=\begin{bmatrix} \frac{1}{2}\sum_{i=1}^{N}\Big(\frac{1}{\hat{w}_i^4}+\frac{T-1}{\hat{\sigma}_\epsilon^4}\Big) & \frac{(T-1)\hat{\sigma}_\epsilon^2}{T}\sum_{i=1}^{N}\Big(\frac{1}{\hat{w}_i^4}-\frac{1}{\hat{\sigma}_\epsilon^4}\Big) & \frac{T}{2}\iota_N' HW^{-2}Z \\ \frac{(T-1)\hat{\sigma}_\epsilon^2}{T}\sum_{i=1}^{N}\Big(\frac{1}{\hat{w}_i^4}-\frac{1}{\hat{\sigma}_\epsilon^4}\Big) & \hat{a}_{\rho\rho} & (T-1)\hat{\sigma}_\epsilon^2\iota_N' HW^{-2}Z \\ \frac{T}{2}Z'W^{-2}H\iota_N & (T-1)\hat{\sigma}_\epsilon^2 Z'W^{-2}H\iota_N & \frac{T^2}{2}Z'W^{-2}H^2Z \end{bmatrix},$$

where $\hat{a}_{\rho\rho} = \frac{2(T-1)^2}{T^2}\sum_{i=1}^{N}(\hat{\sigma}_\epsilon^2/\hat{w}_i^2-1)^2+\frac{2(2T-3)}{T}\sum_{i=1}^{N}(\hat{\sigma}_\epsilon^2/\hat{w}_i^2-1)+N(T-1)$, $W=\mathrm{diag}(\hat{w}_1^2,\ldots,\hat{w}_N^2)$ and $H=\mathrm{diag}(h'(z_1'\hat{\alpha}),\ldots,h'(z_N'\hat{\alpha}))$.

**Box I.**

The first conditional LM test, tests the null hypothesis

$$H_0^b:\rho=0 \quad \text{(assuming some elements of } \alpha \text{ may not be zero)}. \tag{14}$$

This LM statistic is derived in the accompanying working paper Baltagi et al. (2008), and is given by

$$LM_b=\widehat{J}_b(\theta)^{\rho\rho}D_b(\hat{\rho})^2 \tag{15}$$

where

$$D_b(\hat{\rho})=\frac{T-1}{T}\sum_{i=1}^{N}\Big(\frac{\hat{w}_i^2-\hat{\sigma}_\epsilon^2}{\hat{w}_i^2}\Big) + \frac{\hat{\sigma}_\epsilon^2}{2}\hat{u}'\Big[\Big(\mathrm{diag}(1/\hat{w}_i^2)\otimes\bar{J}_T+1/\hat{\sigma}_\epsilon^2 I_N\otimes E_T\Big) \times\Big(I_N\otimes G\Big)\Big(\mathrm{diag}(1/\hat{w}_i^2)\otimes\bar{J}_T+1/\hat{\sigma}_\epsilon^2 I_N\otimes E_T\Big)\Big]\hat{u} \tag{16}$$

with $\widehat{u}=y-X\widehat{\beta}_{MLE}$ denoting the restricted MLE residuals under $H_0^b$. Also, $\widehat{w}_i^2=Th(z_i'\widehat{\alpha})+\widehat{\sigma}_\epsilon^2$, where $\widehat{\alpha}$ and $\widehat{\sigma}_\epsilon^2$ are the restricted MLE of $\alpha$ and $\sigma_\epsilon^2$ under $H_0^b$. The information matrix with respect to $\theta=(\sigma_\epsilon^2,\rho,\alpha')'$ under $H_0^b$ is given in Box I. $\widehat{J}_b(\theta)^{\rho\rho}$ is the element of the inverse of the information matrix $\widehat{J}_b(\theta)$ corresponding to $\rho$ evaluated under $H_0^b$. Under the null hypothesis, $LM_b$ is asymptotically distributed as $\chi_1^2$.

The second conditional LM test, tests the null hypothesis

$$H_0^c:\alpha_2=\cdots=\alpha_p=0 \quad (\text{given } \sigma_\mu^2>0 \text{ and } \rho\neq 0). \tag{17}$$

This LM statistic is derived in the accompanying working paper Baltagi et al. (2008), and is given by

$$LM_c=\frac{1}{2}f^{c\prime}Z(Z'Z)^{-1}Z'f^c \tag{18}$$

$LM_c$ is the familiar LM test used in testing the heteroskedasticity by Breusch and Pagan (1979). However, this one uses the restricted maximum likelihood residuals under the null hypothesis $H_0^c$ rather than OLS residuals. In this case, $f_i^c=\frac{\hat{\lambda}^2}{\hat{d}^2(1-\hat{\rho})^2\hat{\sigma}_\epsilon^4}\hat{u}_i'\hat{A}\hat{u}_i-1$, with $d^2=\delta^2+T-1$ and $\lambda^2=d^2\sigma_\mu^2(1-\rho)^2+\sigma_\epsilon^2$ and $\delta=\sqrt{\frac{1+\rho}{1-\rho}}$.

$$\hat{A}=\hat{\Sigma}^{-1}J_T\hat{\Sigma}^{-1}-2\frac{\hat{\sigma}_\mu^2}{\hat{\lambda}^2}\hat{\Sigma}^{-1}J_T\hat{\Sigma}^{-1}J_T\hat{\Sigma}^{-1} +\frac{\hat{\sigma}_\mu^4}{\hat{\lambda}^4}\hat{\Sigma}^{-1}J_T\hat{\Sigma}^{-1}J_T\hat{\Sigma}^{-1}J_T\hat{\Sigma}^{-1},$$

where $\Sigma=\frac{1}{1-\rho^2}R$, and $R$ is the AR(1) correlation matrix. $\widehat{u}=y-X\widehat{\beta}_{MLE}$ denotes the restricted maximum likelihood residuals under the null hypothesis $H_0^c$. Also, $\hat{\rho}$, $\hat{\sigma}_\epsilon^2$, $\hat{\sigma}_\mu^2$ and $\hat{\alpha}_1$ are the restricted ML estimates of $\rho$, $\sigma_\epsilon^2$, $\sigma_\mu^2$ and $\alpha_1$, under $H_0^c$. Under the null hypothesis $H_0^c$, $LM_c$ is asymptotically distributed as $\chi_{p-1}^2$.

The Monte Carlo results, described in detail in the companion working paper Baltagi et al. (2008), show that these tests along with their likelihood ratio alternatives have good size and power under various forms of heteroskedasticity including exponential and quadratic functional forms. This confirms the general asymptotic theory that the proposed LM tests have the correct size asymptotically and their powers converge to one as the sample size increases, see Bera and Bilias (2001) for an excellent review of this literature.

## Acknowledgement

This work was supported by KOSEF (R01-2006-000-10563-0). We dedicate this paper in memory of our colleague and co-author Seuck Heun Song who passed away March, 2008. We would like to thank the referees, the Associate editor and Cheng Hsiao for their helpful comments and suggestions.

## References

Baltagi, B.H., 2008. Econometrics Analysis of Panel Data. Wiley, Chichester.
Baltagi, B.H., Bresson, G., Pirotte, A., 2006. Joint LM test for heteroskedasticity in a one-way error component model. Journal of Econometrics 134, 401–417.
Baltagi, B.H., Li, Q., 1995. Testing AR(1) against MA(1) disturbances in an error component model. Journal of Econometrics 68, 133–151.
Baltagi, B.H., Jung, B.C., Song, S.H., 2008. Testing for heteroskedasticity and serial correlation in a random effects panel data model, Center for Policy Research Working Paper, Syracuse University, Syracuse, New York.
Bera, A.K., Bilias, Y., 2001. Rao's score, Neyman's $C(\alpha)$ and Silvey's LM tests: An essay on historical developments and some new results. Journal of Statistical Planning and Inference 97, 9–44.
Bera, A.K., Sosa-Escudero, W., Yoon, M., 2001. Tests for the error component model in the presence of local misspecification. Journal of Econometrics 101, 1–23.
Breusch, T.S., Pagan, A.R., 1979. A simple test for heteroskedasticity and random coefficient variation. Econometrica 47, 1287–1294.
Galbraith, J.W., Zinde-Walsh, V., 1995. Transforming the error-component model for estimation with general ARMA disturbances. Journal of Econometrics 66, 349–355.
Holly, A., Gardiol, L., 2000. A score test for individual heteroscedasticity in a one-way error components model. In: Krishnakumar, J., Ronchetti, E. (Eds.), Panel Data Econometrics: Future Directions. North-Holland, Amsterdam, pp. 199–211. Chapter 10.
Hsiao, C., 2003. Analysis of Panel Data. Cambridge University Press, Cambridge.
Lejeune, B., 1996. A full heteroscedastic one-way error components model for incomplete panel: Maximum likelihood estimation and Lagrange multiplier testing, CORE discussion paper 9606, Universite Catholique de Louvain, pp. 1–28.
Li, Q., Stengos, T., 1994. Adaptive estimation in the panel data error component model with heteroskedasticity of unknown form. International Economic Review 35, 981–1000.
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.