The Stata Journal (2008)
8, Number 1, pp. 68–78

Tests for unbalanced error-components models under local misspecification

Walter Sosa-Escudero
Department of Economics
Universidad de San Andrés
Buenos Aires, Argentina
wsosa@udesa.edu.ar

Anil K. Bera
Department of Economics
University of Illinois at Urbana–Champaign
Champaign, IL

Abstract. This paper derives *unbalanced* versions of the test statistics for first-order serial correlation and random individual effects summarized in Sosa-Escudero and Bera (2001, *Stata Technical Bulletin Reprints*, vol. 10, pp. 307–311), and updates their `xttest1` routine. The derived test statistics should be useful for applied researchers faced with the increasing availability of panel information where not every individual or country is observed for the full time span. The test statistics proposed here are based on ordinary least-squares residuals and hence are computationally very simple.

Keywords: sg164_1, xttest1, error-components model, unbalanced panel data, testing, misspecification

1 Introduction

A standard specification check that accompanies the output of almost every estimated error-components model is a simple test for the presence of random individual effects. The well-known Breusch–Pagan statistic (Breusch and Pagan 1980), based on the Rao-score (RS) principle, is a frequent choice. Bera, Sosa-Escudero, and Yoon (2001) demonstrated that, in the presence of first-order serial correlation, the test too often rejects the correct null hypothesis of no random effects. Consequently, they propose a modified version that is not affected by the presence of local serial correlation. A similar concern affects the standard test for first-order serial correlation derived by Baltagi and Li (1991), which overrejects the true null hypothesis when random effects are present. For this case, an adjusted RS test was also derived by Bera, Sosa-Escudero, and Yoon (2001). These test statistics, along with their `xttest1` routine in Stata and some empirical illustrations, are presented in Sosa-Escudero and Bera (2001). For a textbook exposition, see Baltagi (2005, 96–97).

These test procedures were originally derived for the *balanced* case, that is, in the panel-data terminology, the case where all individuals are observed for the same number of periods, and in every period all individuals are observed. On the other hand, in applied work the availability of *unbalanced* panels is far from being an uncommon situation. Though in some cases statistical procedures designed for the balanced case can be straightforwardly extended to accommodate unbalanced panels, many estimation or test procedures require less trivial modifications.

Baltagi and Li (1990) derived an unbalanced version of the Breusch–Pagan statistic. The purpose of this paper is to derive unbalanced versions of the test for first-order serial correlation originally proposed by Baltagi and Li (1991) and of the modified tests proposed by Bera, Sosa-Escudero, and Yoon (2001). As a simple extension, we also derive an unbalanced version of the joint test of serial correlation and random effects proposed by Baltagi and Li (1991). The derived test statistics, being based on ordinary least-squares residuals after pooled estimation, are computationally very simple. Finally, the Sosa-Escudero and Bera (2001) `xttest1` routine is appropriately updated to handle unbalanced panels.

2 Tests for the unbalanced case

Consider a simple linear model for panel data allowing for the presence of random individual effects and first-order serial correlation:

$$\begin{aligned} y_{it} &= x'_{it}\beta + u_{it} \\ u_{it} &= \mu_i + \nu_{it} \\ \nu_{it} &= \lambda\nu_{i,t-1} + \epsilon_{it}, \qquad |\lambda| < 1 \end{aligned}$$

where x_{it} is a $k \times 1$ vector of explanatory variables with 1 in its first position, β is a $k \times 1$ vector of parameters including an intercept, $\mu_i \sim N(0, \sigma^2_\mu)$, and $\epsilon_{it} \sim N(0, \sigma^2_\epsilon)$. We will assume $\nu_{i,0} \sim N\left\{0, \sigma^2_\epsilon/(1-\lambda^2)\right\}$.

We will be interested in testing for the absence of random effects ($H_0 : \sigma^2_\mu = 0$) and/or first-order serial correlation ($H_0 : \lambda = 0$). The panel will be unbalanced in the sense that for every individual $i = 1, \dots, N$ we will observe, possibly, a different number of time observations T_i. We will restrict the analysis to the cases where missing observations occur either at the beginning or at the end of the sample period for each individual (that is, there are no "gaps" in the series), and the starting and final periods are determined randomly. Hence, without loss of generality and to avoid complicating the notation too much, we can safely assume that the series for each individual starts at the same period ($t = 1$) and finish randomly at period $t = T_i$.

Let $m = \sum_{i=1}^N T_i$ be the total number of observations. Let u be an $m \times 1$ vector with typical element u_{it} where observations are sorted first by individuals and then by time, so the time index is the faster one. Then in our setup, $V(u) \equiv \Omega$ can be written as

$$V(u) = \sigma^2_\mu \widetilde{H} + \sigma^2_\epsilon \widetilde{V}$$

where $\widetilde{H}$ is an $m \times m$ block diagonal matrix with blocks H_i equal to matrices of ones, each with dimensions $T_i \times T_i$. Similarly, $\widetilde{V}$ will be a block diagonal $m \times m$ matrix with blocks V_i equal to

$$V_i = \begin{bmatrix} 1 & \lambda & \lambda^2 & \dots & \lambda^{T_i-1} \\ \lambda & 1 & \lambda & \dots & \lambda^{T_i-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \lambda^{T_i-1} & \lambda^{T_i-2} & \lambda^{T_i-3} & \dots & 1 \end{bmatrix}$$

For the purpose of deriving the test statistics, the log-likelihood function will be

$$L(\beta, \lambda, \sigma_\epsilon^2, \sigma_\mu^2) = \text{constant} - \frac{1}{2}\log|\Omega| - \frac{1}{2}u'\Omega^{-1}u$$

The information matrix for this problem is known to be block diagonal between β and the remaining parameters. Therefore, for the purposes of this paper, we will concentrate only on the parameters λ, σ_μ^2, and σ_ϵ^2. Under a more general setup, suppose the log likelihood can be characterized by a three-parameter vector $\theta = (\psi, \phi, \gamma)'$. Let $d(\theta)$ be the score vector and $J(\theta)$ the information matrix. If it can be assumed that $\phi = 0$, the standard Rao-score (RS) test statistic for the null hypothesis $H_0: \psi = 0$ is given by

$$\text{RS}_\psi = d_\psi(\widehat{\theta}) J_{\psi\cdot\gamma}^{-1}(\widehat{\theta}) d_\psi(\widehat{\theta}) \tag{1}$$

where d_ψ is the element of the score corresponding to the parameter ψ, $J_{\psi\cdot\gamma}(\theta) = J_\psi - J_{\psi\gamma}J_\gamma^{-1}J_{\gamma\psi}$, and $\widehat{\theta}$ is the maximum likelihood estimator (MLE) of θ under the restriction implied by the null hypothesis and the assumption $\phi = 0$. Asymptotically, this test statistic under the null hypothesis H_0: $\psi = 0$ is known to have a central chi-squared distribution. In the context of our error-components model, if $\gamma = \sigma_\epsilon^2$ and if we set $\psi = \sigma_\mu^2$ and $\phi = \lambda$, (1) is a test for random effects assuming no serial correlation; and if we set $\psi = \lambda$ and $\phi = \sigma_\mu^2$, (1) gives a test for serial correlation assuming no random effects. The standard Breusch–Pagan test for random effects (assuming no serial correlation) and the Baltagi–Li test for first-order serial correlation (assuming no random effects) are derived from this principle.

Bera and Yoon (1993) showed that the test statistic (1) is invalid when $\phi \neq 0$, in the sense that the test tends to reject the null hypothesis too frequently even when it is correct. More specifically, the RS_ψ statistic is found to have an asymptotic *noncentral* chi-squared distribution under $H_0: \psi = 0$, when $\phi = \delta/\sqrt{n}$, that is, when the alternative is *locally misspecified*. In particular, this implies that when the null is correct, the Breusch–Pagan test tends to reject the true null of absence of random effects if the error term is serially correlated, even in a local sense. A similar situation arises for the test for serial correlation of Baltagi and Li (1991) in the local presence of random effects. In order to remedy this problem, Bera and Yoon (1993) proposed the following modified RS statistic:

$$\begin{aligned}\mathrm{RS}^*_\psi \;=\;& \frac{1}{n}\left\{d_\psi(\widehat{\theta}) - J_{\psi\phi\cdot\gamma}(\widehat{\theta})J^{-1}_{\phi\cdot\gamma}(\widehat{\theta})d_\phi(\widehat{\theta})\right\}' \\ & \left\{J_{\psi\cdot\gamma}(\widehat{\theta}) - J_{\psi\phi\cdot\gamma}(\widehat{\theta})J^{-1}_{\phi\cdot\gamma}(\widehat{\theta})J_{\phi\psi\cdot\gamma}(\widehat{\theta})\right\}^{-1} \\ & \left\{d_\psi(\widehat{\theta}) - J_{\psi\phi\cdot\gamma}(\widehat{\theta})J^{-1}_{\phi\cdot\gamma}(\widehat{\theta})d_\phi(\widehat{\theta})\right\} \end{aligned} \tag{2}$$

where $\widehat{\theta}$ is the MLE of θ under the joint null $\psi = \phi = 0$. This modified test statistic has an asymptotic *central* χ^2_1 distribution under the null hypothesis $H_0 : \psi = 0$ and when $\phi = \delta/\sqrt{n}$, that is, the modified test statistic has the correct size even when the underlying model is locally misspecified. Based on this principle, Bera, Sosa-Escudero, and Yoon (2001) derived modified tests for random effects (serial correlation), which are valid in the presence of local first-order serial correlation (random effects) assuming that the panel is balanced.

To derive tests for the unbalanced case, let $\theta = (\lambda, \sigma^2_\mu, \sigma^2_\epsilon)'$ and $\widehat{\theta} = (0, 0, \widehat{\sigma}^2_\epsilon)'$ be the MLE of θ under the joint null hypothesis $H_0 : \lambda = \sigma^2_\mu = 0$. The following formula by Hemmerle and Hartley (1973) will be useful to derive the score vector for the problem:

$$d_{\theta_r} \equiv \frac{\partial L}{\partial \theta_r} = -\frac{1}{2}\,\mathrm{tr}\left(\Omega^{-1}\frac{\partial\Omega}{\partial\theta_r}\right) + \frac{1}{2}\left(u'\Omega^{-1}\frac{\partial\Omega}{\partial\theta_r}\Omega^{-1}u\right) \tag{3}$$

where θ_r denotes the rth element of θ, $r = 1, 2, 3$. Note that $\partial\Omega/\partial\sigma^2_\mu = \widetilde{H}$ with $\mathrm{tr}(\widetilde{H}) = m$. Similarly, $\partial\Omega/\partial\sigma^2_\epsilon = \widetilde{V}$, which under the restricted MLE is an $m \times m$ identity matrix with trace equal to m. Also $\partial\Omega/\partial\lambda = \sigma^2_\epsilon\widetilde{G}$, where $\widetilde{G}$ is a block diagonal matrix with blocks equal to G_i, with $G_i = \partial V_i/\partial\lambda$ given by

$$G_i = \begin{bmatrix} 0 & 1 & 2\lambda & \cdots & (T_i-1)\lambda^{T_i-2} \\ 1 & 0 & 1 & \cdots & (T_i-2)\lambda^{T_i-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & 1 & 0 & 1 \\ (T_i-1)\lambda^{T_i-2} & \cdots & \cdots & 1 & 0 \end{bmatrix}$$

Under the restricted MLE, G_i is a bidiagonal matrix as follows:

$$G_i(\widehat{\theta}) = \begin{bmatrix} 0 & 1 & 0 & & \cdots & 0 \\ 1 & 0 & 1 & & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & \cdots & & 1 & 0 & 1 \\ 0 & \cdots & & \cdots & 1 & 0 \end{bmatrix}$$

Hence, $\mathrm{tr}\left\{G_i(\widehat{\theta})\right\} = 0$. Replacing these results in (3) and evaluating the expression under the restricted MLE, we obtain

$$
\begin{aligned}
d_{\sigma^2_\mu}(\widehat{\theta}) &= -\frac{1}{2}\mathrm{tr}\left(\frac{1}{\widehat{\sigma}^2_\epsilon}I_m\widetilde{H}\right)+\frac{1}{2}e'\frac{1}{\widehat{\sigma}^2_\epsilon}I_m\widetilde{H}\frac{1}{\widehat{\sigma}^2_\epsilon}e \\
&= -\frac{1}{2}\frac{1}{\widehat{\sigma}^2_\epsilon}m+\frac{1}{2}\frac{1}{\widehat{\sigma}^4_\epsilon}e'\widetilde{H}e=-\frac{m}{\widehat{\sigma}^2_\epsilon}\ A
\end{aligned}
$$

where e is an $m\times 1$ vector with typical element $e_{it}=x'_{it}\widehat{\beta}$, and $\widehat{\beta}$ is the restricted MLE of β. Similarly, $\widehat{\sigma}^2_\epsilon = e'e/m$ is the restricted MLE of σ^2_ϵ, and $A\equiv 1-e'\widetilde{H}e/(e'e)$. In a similar fashion,

$$
\begin{aligned}
d_\lambda(\widehat{\theta}) &= -\frac{1}{2}\mathrm{tr}\left\{\frac{1}{\widehat{\sigma}^2_\epsilon}\widehat{\sigma}^2_\epsilon\widetilde{G}(\widehat{\theta})\right\}+\frac{1}{2}2\frac{1}{\widehat{\sigma}^2_\epsilon}e'\widetilde{G}(\widehat{\theta})e \\
&= \frac{1}{\widehat{\sigma}^2_\epsilon}e'\widetilde{G}(\widehat{\theta})\ e = m\ B
\end{aligned}
$$

where $B\equiv e'\widetilde{G}e/e'e$.

To derive the elements of the information matrix, we will use the following formula from Baltagi (2005, 59–60):

$$
J_{r,s}(\theta)=E\left(-\frac{\partial^2 L}{\partial\theta_r\partial\theta_s}\right)=\frac{1}{2}\mathrm{tr}\left(\Omega^{-1}\frac{\partial\Omega}{\partial\theta_r}\Omega^{-1}\frac{\partial\Omega}{\partial\theta_s}\right)
$$

Then

$$
\begin{aligned}
J_{\sigma^2_\epsilon,\sigma^2_\epsilon}(\widehat{\theta}) &= \frac{1}{2}\mathrm{tr}\left\{\frac{1}{\widehat{\sigma}^2_\epsilon}\widetilde{V}(\widehat{\theta})\frac{1}{\widehat{\sigma}^2_\epsilon}\widetilde{V}(\widehat{\theta})\right\}=\frac{1}{2}\mathrm{tr}\left(\frac{1}{\widehat{\sigma}^4_\epsilon}I_m\right)=\frac{m}{2\widehat{\sigma}^4_\epsilon} \\
J_{\hat{\sigma}^2_\mu,\hat{\sigma}^2_\mu}(\widehat{\theta}) &= \frac{1}{2}\mathrm{tr}\left(\frac{1}{\widehat{\sigma}^2_\epsilon}\widetilde{H}\frac{1}{\widehat{\sigma}^2_\epsilon}\widetilde{H}\right)=\frac{1}{2}\frac{1}{\widehat{\sigma}^4_\epsilon}\mathrm{tr}\left(\widetilde{H}\widetilde{H}\right)=\frac{\sum_{i=1}^N T_i^2}{2\widehat{\sigma}^4_\epsilon} \\
J_{\lambda,\lambda}(\widehat{\theta}) &= \frac{1}{2}\mathrm{tr}\left\{\frac{1}{\widehat{\sigma}^2_\epsilon}\widehat{\sigma}^2_\epsilon\widetilde{G}(\widehat{\theta})\frac{1}{\widehat{\sigma}^2_\epsilon}\widehat{\sigma}^2_\epsilon\widetilde{G}(\widehat{\theta})\right\}=\frac{1}{2}\mathrm{tr}\left\{\widetilde{G}(\widehat{\theta})\widetilde{G}(\widehat{\theta})\right\} \\
&= \frac{1}{2}\sum_{i=1}^N 2(T_i-1)=m-N \\
J_{\hat{\sigma}^2_\epsilon,\hat{\sigma}^2_\mu}(\widehat{\theta}) &= \frac{1}{2}\mathrm{tr}\left\{\frac{1}{\widehat{\sigma}^2_\epsilon}\widetilde{V}(\widehat{\theta})\frac{1}{\widehat{\sigma}^2_\epsilon}\widetilde{V}(\widehat{\theta})\right\}=\frac{1}{2}\frac{1}{\widehat{\sigma}^4_\epsilon}\mathrm{tr}\left(\widetilde{H}\right)=\frac{m}{2\widehat{\sigma}^4_\epsilon} \\
J_{\hat{\sigma}^2_\epsilon,\lambda}(\widehat{\theta}) &= \frac{1}{2}\mathrm{tr}\left\{\frac{1}{\widehat{\sigma}^2_\epsilon}\widetilde{V}(\widehat{\theta})\frac{1}{\widehat{\sigma}^2_\epsilon}\widetilde{G}(\widehat{\theta})\right\}=\frac{1}{2}\frac{1}{\widehat{\sigma}^4_\epsilon}\mathrm{tr}\left\{\widetilde{G}(\widehat{\theta})\right\}=0 \\
J_{\lambda,\hat{\sigma}^2_\mu}(\widehat{\theta}) &= \frac{1}{2}\mathrm{tr}\left\{\frac{1}{\widehat{\sigma}^2_\epsilon}\widehat{\sigma}^2_\epsilon\widetilde{G}(\widehat{\theta})\frac{1}{\widehat{\sigma}^2_\epsilon}\widetilde{H}\right\}=\frac{1}{2}\frac{1}{\widehat{\sigma}^2_\epsilon}\mathrm{tr}\left\{\widetilde{G}(\widehat{\theta})\widetilde{H}\right\} \\
&= \frac{1}{2}\frac{2}{\widehat{\sigma}^2_\epsilon}\left(\sum_{i=1}^N T_i-N\right)=\frac{1}{\widehat{\sigma}^2_\epsilon}(m-N)
\end{aligned}
$$

where we have used the facts that $\operatorname{tr}\left\{\widetilde{G}_i(\widehat{\theta})\widetilde{G}_i(\widehat{\theta})\right\} = \operatorname{tr}\left\{\widetilde{G}_i(\widehat{\theta})\widetilde{H}_i\right\} = 2(T_i - 1)$, and $\operatorname{tr}(\widetilde{H}_i\widetilde{H}_i) = T_i^2$.

Collecting all the elements, the information matrix evaluated at the restricted MLE under the joint null can be expressed as

$$J(\widehat{\theta}) = \frac{1}{2\widehat{\sigma}_\epsilon^4}\begin{bmatrix} m & m & 0 \\ m & a & 2\widehat{\sigma}_\epsilon^2(m-N) \\ 0 & 2\widehat{\sigma}_\epsilon^2(m-N) & 2\widehat{\sigma}_\epsilon^4(m-N) \end{bmatrix}$$

where $a \equiv \sum_{i=1}^N T_i^2$. For the balanced case $T_i = T$, we get exactly the same expression for $J(\widehat{\theta})$ as in Baltagi and Li (1991, 279). From the above expression of $J(\widehat{\theta})$, we can show that

$$\begin{aligned} J_{\mu\lambda\cdot\sigma_\epsilon^2} &= \frac{m-N}{\widehat{\sigma}_\epsilon^2} \\ J_{\mu\cdot\sigma_\epsilon^2} &= \frac{a-m}{2\widehat{\sigma}_\epsilon^4} \\ J_{\lambda\cdot\sigma_\epsilon^2} &= m-N \end{aligned}$$

Substituting these results in (2), we obtain the unbalanced version of the modified test for random effects as

$$\mathrm{RS}_\mu^* = \frac{m^2\left(A+2B\right)^2}{2\left(a-3m+2N\right)}$$

When $T_i = T$ (the balanced case), the above expression boils down to

$$\mathrm{RS}_\mu^* = \frac{NT\left(A+2B\right)^2}{2(T-1)\left\{1-(2/T)\right\}}$$

as in Bera, Sosa-Escudero, and Yoon (2001) for the balanced case.

Similarly, the modified test statistic for serial correlation is

$$\mathrm{RS}_\lambda^* = \left(B + \frac{m-N}{a-m}A\right)^2 \frac{(a-m)m^2}{(m-N)(a-3m+2N)}$$

and when $T_i = T$, we get

$$\mathrm{RS}_\lambda^* = \left(B + \frac{A}{T}\right)^2 \frac{NT^2}{(T-1)(1-2/T)}$$

which is the expression in Bera, Sosa-Escudero, and Yoon (2001) for the balanced case.

For computational purposes, it is interesting to see that

$$A = 1 - \frac{\sum_{i=1}^{N} \left(\sum_{t=1}^{T_i} e_{it}^2 \right)^2}{\sum_{i=1}^{N} \sum_{t=1}^{T_i} e_{it}^2}$$

and

$$B = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T_i} e_{i,t} e_{i,t-1}}{\sum_{i=1}^{N} \sum_{t=1}^{T_i} e_{it}^2}$$

and, therefore, there is no need to construct the $\widetilde{G}$ or $\widetilde{H}$ matrices; hence, the test statistics can be easily computed right after ordinary least-squares estimation without constructing any matrices.

The previous derivations allow us to obtain the unbalanced version of the test for serial correlation assuming no random effects:

$$\mathrm{RS}_\lambda = \frac{m^2 B^2}{m - N}$$

which again reduces to $NT^2B^2/(T-1)$, originally derived by Baltagi and Li (1991) for balanced panels. Also, for completeness, the unbalanced version of the test for random effects assuming no serial correlation is given by

$$\mathrm{RS}_\mu = \frac{\frac{1}{2} m^2 A^2}{a - m}$$

This test statistic is a particular case of the Baltagi–Li test for the two-way error-components model.

Suppose that we are interested in the joint null hypothesis of no random effects and no first-order serial correlation. Let $\mathrm{RS}_{\phi,\psi}$ be the RS test statistic for the joint null hypothesis $H_0 : \phi = \psi = 0$. Bera and Yoon (2001) show that the following identities hold:

$$\mathrm{RS}_{\phi\psi} = \mathrm{RS}_\psi^* + \mathrm{RS}_\phi = \mathrm{RS}_\phi^* + \mathrm{RS}_\psi$$

This simplifies computations, as illustrated in Sosa-Escudero and Bera (2001). Then, as a simple byproduct of the previous derivations, we can obtain a statistic for jointly testing serial correlation and random effects, as

$$\mathrm{RS}_{\lambda\mu} = m^2 \left\{ \frac{A^2 + 4AB + 4B^2}{2(a - 3m + 2N)} \; + \; \frac{B^2}{m - N} \right\}$$

When $T_i = T$, $\mathrm{RS}_{\lambda\mu}$ simplifies to

$$\mathrm{RS}_{\lambda\mu} = \frac{NT^2}{2(T-1)(T-2)} \left(A^2 + 4AB + 2TB^2\right)$$

which is the original joint test statistic of Baltagi and Li (1991).

Finally, because $\sigma_\mu^2 \geq 0$, it is natural to consider one-sided versions of the tests for the null $H_0 : \sigma_\mu^2 = 0$. As in Bera, Sosa-Escudero, and Yoon (2001), appropriate test statistics can be readily constructed by taking the signed square roots of the original two-sided tests RS_μ and RS_μ^*. Denoting their one-sided versions, respectively, as RSO_μ and RSO_μ^*, we have

$$\mathrm{RSO}_\mu = -\sqrt{\frac{\frac{1}{2}\, m^2}{a-m}}\, A$$

and

$$\mathrm{RSO}_\mu^* = -\sqrt{\frac{m^2}{2\,(a-3m+2N)}}\, (A+2B)$$

3 Empirical illustration

As an illustration of these procedures, we provide an empirical exercise that is based on Gasparini, Marchionni, and Sosa-Escudero (2001). It consists of a simple linear panel-data model where the dependent variable is the Gini coefficient for 17 regions of Argentina. The vector of explanatory variables includes mean income and its square (`ie` and `ie2`); proportion of the population employed in the manufacturing industry (`indus`) and in public administration, health, or education (`adpubedsal`); unemployment rate (`desempleo`); activity rate (`tactiv`); public investment as percentage of GDP (`invipib`); degree of openness (`apertura`); social assistance (`pyas4`); proportion of population older than 64 (`e64`); proportion of population that completed high school (`supc`); and average family size (`tamfam`); for details see Gasparini, Marchionni, and Sosa-Escudero (2001). Models of this type have been used extensively in the literature exploring the links between inequality and development, usually to study the so-called "Kuznets hypothesis", which postulates an inverted U-shaped relationship between these two variables (for example, see Anand and Kanbur [1993] and Gustafsson and Johansson [1999]).

Income-related variables, including the Gini coefficients, are constructed using Argentina's *Permanent Household Survey* (Encuesta Permamente de Hogares), which surveys several socioeconomic variables at the household level for several regions of the country. Because of certain administrative deficiencies, the panel is largely unbalanced, so the number of available temporal observations ranges from 5 to 8 years in the period 1992–2000.

First, we `tsset` the data and then use `xtreg` to estimate the parameters of a one-way error-components model with region-specific random effects:

```
. use ginipanel5

. tsset naglo ano
       panel variable:  naglo (unbalanced)
        time variable:  ano, 1992 to 2000, but with a gap
                delta:  1 unit

. xtreg gini ie ie2 indus adpubedsal desempleo tactiv invipib apertura pyas4
> e64 supc tamfam, re i(naglo)

Random-effects GLS regression                   Number of obs      =       128
Group variable: naglo                           Number of groups   =        17

R-sq:  within  = 0.5096                         Obs per group: min =         6
       between = 0.6153                                        avg =       7.5
       overall = 0.5344                                        max =         8

Random effects u_i ~ Gaussian                   Wald chi2(12)      =    121.30
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        gini |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ie |  -.0000995   .0001823    -0.55   0.585    -.0004568    .0002578
         ie2 |   1.64e-08   2.19e-07     0.08   0.940    -4.12e-07    4.45e-07
       indus |   -.041974   .0704982    -0.60   0.552    -.1801478    .0961999
  adpubedsal |  -.0635789   .0531777    -1.20   0.232    -.1678053    .0406475
   desempleo |  -.1177452   .0638999    -1.84   0.065    -.2429868    .0074963
      tactiv |   .0999584   .0737997     1.35   0.176    -.0446864    .2446031
     invipib |  -.3307239   .1912258    -1.73   0.084    -.7055197    .0440718
    apertura |   .4289793   .0768693     5.58   0.000     .2783183    .5796404
       pyas4 |   2.884162   1.626136     1.77   0.076    -.3030061    6.071331
         e64 |  -.1339182   .1505384    -0.89   0.374    -.4289681    .1611316
        supc |   .2417907   .0946423     2.55   0.011     .0562952    .4272861
      tamfam |   .0169905   .0174328     0.97   0.330    -.0171771    .0511581
       _cons |   .3084864   .1031351     2.99   0.003     .1063453    .5106274
-------------+----------------------------------------------------------------
     sigma_u |  .01370805
     sigma_e |  .01377936
         rho |  .49740589   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Next the command `xttest1` with the `unadjusted` option presents the following output:

```
. xttest1, unadjusted
Tests for the error component model:
        gini[naglo,t] = Xb + u[naglo] + v[naglo,t]
           v[naglo,t] = lambda v[naglo,(t-1)] + e[naglo,t]
        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                    gini |   .0006167       .0248335
                       e |   .0001899      .01377936
                       u |   .0001879      .01370805
        Tests:
           Random Effects, Two Sided:
           LM(Var(u)=0)         =   13.50 Pr>chi2(1) =  0.0002
           ALM(Var(u)=0)        =    6.03 Pr>chi2(1) =  0.0141
           Random Effects, One Sided:
           LM(Var(u)=0)         =    3.67 Pr>N(0,1)  =  0.0001
           ALM(Var(u)=0)        =    2.46 Pr>N(0,1)  =  0.0070
           Serial Correlation:
           LM(lambda=0)         =    9.32 Pr>chi2(1) =  0.0023
           ALM(lambda=0)        =    1.86 Pr>chi2(1) =  0.1732
           Joint Test:
           LM(Var(u)=0,lambda=0) =  15.35 Pr>chi2(2) =  0.0005
```

The unadjusted version of the tests for random effects (`LM(Var(u)=0)`) and serial correlation (`LM(lambda=0)`), and the test for the joint null (`LM(Var(u)=0,lambda=0)`) suggest rejecting their nulls at the 5% significance level. Care must be taken in deriving conclusions about the direction of the misspecification because, in light of the results in Bera, Sosa-Escudero, and Yoon (2001), rejections may arise because of the presence of random effects, serial correlation, or both. To explore the possible nature of the misspecification, we restore the modified versions of the test. The adjusted version of the test for serial correlation `ALM(lambda=0)` now fails to reject the null hypothesis while the adjusted version of the test for random effects `ALM(Var(u)=0)` still does. This suggests that the possible misspecification is likely due to the presence of random effects rather than the serial correlation. Consequently, and to stress the main usefulness of these procedures, in this example the presence of the random effects seems to confound the unadjusted test for serial correlation, making it spuriously reject its null.

4 Acknowledgments

We thank an anonymous referee for useful suggestions that helped improve the routine considerably, and we thank Javier Alejo for excellent research assistance.

5 References

Anand, S., and R. Kanbur. 1993. Inequality and development: A critique. *Journal of Development Economics* 41: 19–43.

Baltagi, B., and Q. Li. 1990. A Lagrange multiplier test for the error components model with incomplete panels. *Econometric Reviews* 9: 103–107.

———. 1991. A joint test for serial correlation and random individual effects. *Statistics and Probability Letters* 11: 277–280.

Baltagi, B. H. 2005. *Econometric Analysis of Panel Data*. 3rd ed. New York: Wiley.

Bera, A., W. Sosa-Escudero, and M. Yoon. 2001. Tests for the error component model in the presence of local misspecification. *Journal of Econometrics* 101: 1–23.

Bera, A., and M. Yoon. 1993. Specification testing with locally misspecified alternatives. *Econometric Theory* 9: 649–658.

———. 2001. Adjustments of Rao's score test for distributional and local parametric misspecifications. Mimeo: University of Illinois at Urbana–Champaign.

Breusch, T., and A. Pagan. 1980. The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies* 47: 239–253.

Gasparini, L., M. Marchionni, and W. Sosa-Escudero. 2001. *Distribucion del Ingreso en la Argentina: Perspectivas y Efectos sobre el Bienestar*. Cordoba: Fundacion Arcor-Triunfar.

Gustafsson, B., and M. Johansson. 1999. In search of smoking guns: what makes income inequality vary over time in different countries? *American Sociological Review* 64: 585–605.

Hemmerle, W. J., and H. Hartley. 1973. Computing maximum likelihood estimates for the mixed AOV model using the W transformation. *Technometrics* 15: 819–831.

Sosa-Escudero, W., and A. Bera. 2001. sg164: Specification tests for linear panel data models. *Stata Technical Bulletin* 61: 18–21. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 307–311. College Station, TX: Stata Press.

About the authors

Walter Sosa-Escudero is an associate professor and the chairman of the Department of Economics, Universidad de San Andrés, Buenos Aires, Argentina.

Anil K. Bera is a professor in the Department of Economics, University of Illinois at Urbana–Champaign.