The Stata Journal (2003)
3, Number 2, pp. 168–177

Testing for serial correlation in linear panel-data models

David M. Drukker
Stata Corporation

Abstract. Because serial correlation in linear panel-data models biases the standard errors and causes the results to be less efficient, researchers need to identify serial correlation in the idiosyncratic error term in a panel-data model. A new test for serial correlation in random- or fixed-effects one-way models derived by Wooldridge (2002) is attractive because it can be applied under general conditions and is easy to implement. This paper presents simulation evidence that the new Wooldridge test has good size and power properties in reasonably sized samples.

Keywords: st0039, panel data, serial correlation, specification tests

1 Introduction

Because serial correlation in linear panel-data models biases the standard errors and causes the results to be less efficient, researchers need to identify serial correlation in the idiosyncratic error term in a panel-data model. While a number of tests for serial correlation in panel-data models have been proposed, a new test discussed by Wooldridge (2002) is very attractive because it requires relatively few assumptions and is easy to implement. This article presents the results of a size and power simulation study of this new Wooldridge test. Because the test is so flexible, simulations must be performed for a number of different cases. This paper presents results from simulations for both fixed- and random-effects designs, with and without conditional homoskedasticity in the idiosyncratic error term, with balanced data, and with unbalanced data with and without gaps in the individual series. The power simulations include both autoregressive and moving-average alternatives. The test is found to have good size and power properties with samples of moderate size.

Baltagi (2001) extensively discusses testing for serial correlation in the presence of random and fixed effects. Many of these tests make specific assumptions about the nature of the individual effects or test for the individual-level effects jointly.[1] Some of these tests, such as the Baltagi–Wu test derived in Baltagi and Wu (1999), are optimal within a class of tests. In contrast, because the Wooldridge test is based on fewer assumptions, it should be less powerful than the more highly parameterized tests, but it should be more robust. While the robustness of the test makes it attractive, it is important to verify that it has good size and power properties under these weaker assumptions.

[1]See, for instance, the tests derived in Baltagi and Li (1995).

 st0039

2 Wooldridge's test

Let's begin by reviewing the linear one-way model,

$$y_{it} = \alpha + \mathbf{X}_{it}\boldsymbol{\beta}_1 + \mathbf{Z}_i\boldsymbol{\beta}_2 + \mu_i + \epsilon_{it} \qquad i \in \{1, 2, \ldots, N\}, \quad t \in \{1, 2, \ldots, T_i\} \tag{1}$$

where y_{it} is the dependent variable; $\mathbf{X}_{it}$ is a $(1 \times K_1)$ vector of time-varying covariates; $\mathbf{Z}_i$ is a $(1 \times K_2)$ vector of time-invariant covariates; α, $\boldsymbol{\beta}_1$, and $\boldsymbol{\beta}_2$ are $1 + K_1 + K_2$ parameters; μ_i is the individual-level effect; and ϵ_{it} is the idiosyncratic error. If the μ_i are correlated with the $\mathbf{X}_{it}$ or the $\mathbf{Z}_i$, the coefficients on the time-varying covariates $\mathbf{X}_{it}$ can be consistently estimated by a regression on the within-transformed data or the first-differenced data.[2] If the μ_i are uncorrelated with the $\mathbf{X}_{it}$ and the $\mathbf{Z}_i$, the coefficients on the time-varying and time-invariant covariates can be consistently and efficiently estimated using the feasible generalized least squares method known as random-effects regression.[3] All of these estimators assume that $E[\epsilon_{it}\epsilon_{is}] = 0$ for all $s \neq t$; i.e., that there is no serial correlation in the idiosyncratic errors, which would cause the standard errors to be biased and the estimates to be less efficient.

Wooldridge's method uses the residuals from a regression in first-differences. Note that first-differencing the data in the model in (1) removes the individual-level effect, the term based on the time-invariant covariates and the constant,

$$\begin{aligned} y_{it} - y_{it-1} &= (\mathbf{X}_{it} - \mathbf{X}_{it-1})\,\boldsymbol{\beta}_1 + \epsilon_{it} - \epsilon_{it-1} \\ \Delta y_{it} &= \Delta \mathbf{X}_{it}\boldsymbol{\beta}_1 + \Delta\epsilon_{it} \end{aligned}$$

where Δ is the first-difference operator.

Wooldridge's procedure begins by estimating the parameters $\boldsymbol{\beta}_1$ by regressing Δy_{it} on $\Delta \mathbf{X}_{it}$ and obtaining the residuals $\widehat{e}_{it}$. Central to this procedure is Wooldridge's observation that, if the ϵ_{it} are not serially correlated, then $\mathrm{Corr}(\Delta\epsilon_{it}, \Delta\epsilon_{it-1}) = -.5$. Given this observation, the procedure regresses the residuals $\widehat{e}_{it}$ from the regression with first-differenced variables on their lags and tests that the coefficient on the lagged residuals is equal to $-.5$. To account for the within-panel correlation in the regression of $\widehat{e}_{it}$ on $\widehat{e}_{it-1}$, the VCE is adjusted for clustering at the panel level. Since `cluster()` implies `robust`, this test is also robust to conditional heteroskedasticity.

3 xtserial

This article uses the new Stata command `xtserial`, which implements the Wooldridge test for serial correlation in panel data. The syntax of `xtserial` is

`xtserial` *depvar* [*varlist*] [`if` *exp*] [`in` *range*] [`, output`]

You must `tsset` your data before using `xtserial`. See [TS] **tsset** for more information about `tsset`.

[2]See Baltagi (2001) and Wooldridge (2002) for a discussion of these estimators.
[3]See Baltagi (2001) and Wooldridge (2002) for a discussion of these estimators.

3.1 Options

`output` specifies that the output from the first-differenced regression should be displayed. By default, the first-differenced regression output is not displayed.

3.2 Example

Let's consider an example using an extract from the National Longitudinal Study of women who were 14–26 years old in 1968. Our model supposes that log wages, `ln_wage`, are a linear function of age, `age`; age squared, `age2`; total working experience, `ttl_exp`; tenure at current position and its square, `tenure` and `tenure2`; and a binary indicator for living in the south, `south`.[4,5] In the output below, we use `xtserial` to test the null hypothesis that there is no serial correlation in this specification. We specify the `output` option so that `xtserial` will display the regression with the first-differenced variables.

```
. use http://www.stata-press.com/data/r8/nlswork.dta
(National Longitudinal Survey.  Young Women 14-26 years of age in 1968)

. tsset idcode year
       panel variable:  idcode, 1 to 5159
        time variable:  year, 68 to 88, but with gaps

. gen age2 = age^2
(24 missing values generated)

. gen tenure2 = tenure^2
(433 missing values generated)

. xtserial ln_wage age* ttl_exp tenure* south, output

Regression with robust standard errors                Number of obs =   10528
                                                      F(  6,  3659) =  105.13
                                                      Prob > F      =  0.0000
                                                      R-squared     =  0.0411
Number of clusters (idcode) = 3660                    Root MSE      =  .30724

------------------------------------------------------------------------------
             |               Robust
   D.ln_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
age          |
          D1 |   .0338027    .0161031     2.10   0.036      .0022308    .0653746
age2         |
          D1 |  -.0002561    .0002672    -0.96   0.338      -.00078     .0002679
ttl_exp      |
          D1 |   .0351088    .0099347     3.53   0.000      .0156307     .054587
tenure       |
          D1 |   .0311144    .0055471     5.61   0.000      .0202387    .0419902
tenure2      |
          D1 |  -.0030878    .0007035    -4.39   0.000     -.0044671   -.0017084
south        |
          D1 |  -.0520378    .0278607    -1.87   0.062     -.1066619    .0025863
------------------------------------------------------------------------------

Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
    F(  1,    1472) =     88.485
           Prob > F =      0.0000
```

[4] See [XT] **xt** for more information about this dataset.

[5] Because the measure of education, highest grade completed, is time-invariant, it cannot be included in the model.

Note that the null hypothesis of no serial correlation is strongly rejected. Also, the output from the first-differenced regression includes standard errors that account for clustering within the panels. If there is serial correlation in the idiosyncratic error term, clustering at the panel level will produce consistent estimates of the standard errors, and as discussed by Baltagi (2001) and Wooldridge (2002), other estimators will produce more efficient estimates.

3.3 Saved results

`xtserial` saves in `r()`:

Scalars

`r(corr)`	estimated coefficient on lagged residuals	`r(F)`	F statistic
`r(df)`	numerator degrees of freedom of F statistic	`r(df_r)`	denominator degrees of freedom of F statistic
`r(p)`	p-value		

4 How the data were generated

Consider a linear panel-data model of the form

$$y_{it} = \alpha + x_{1it}\beta_1 + x_{2it}\beta_2 + \mu_i + \epsilon_{it}$$

for $i \in \{1, \ldots, N\}$ $t \in \{1, \ldots, T_i\}$ and where α, β_1, β_2 are parameters to be estimated.

The simulations discussed below investigate the performance of the Wooldridge test over four axes of interest. These axes are

1. sample size and structure,
2. random-effects and fixed-effects designs,
3. designs that simulate both the null hypothesis of no serial correlation and evaluate the power of the test against different levels of autoregressive and moving-average serial correlation, and
4. conditionally homoskedastic and conditionally heteroskedastic idiosyncratic errors.

By varying the time periods at which individuals are observed, we obtain samples of balanced data, unbalanced data without gaps, and unbalanced data with gaps. We consider only the case in which the mechanism selecting the observations is independent of all the variables in the model. The simulations here investigate the properties of the test with moderately sized samples. In the case of balanced data, sample sizes of $N = 500$, $T = 5$; $N = 500$, $T = 10$; $N = 1000$, $T = 5$; and $N = 1000$, $T = 10$ are considered. Unbalanced data without gaps were obtained by randomly dropping some of the first observations from a random selection of the panels out of an initial sample with

$N = 1000$, $T = 10$.[6] Unbalanced data with gaps were obtained by randomly deciding to include or drop the observations at $t = 3$, $t = 6$, and $t = 7$ for some randomly selected panels.[7]

If $E[\mu_i x_{1it}] = E[\mu_i x_{2it}] = 0$, the model is said to be a random-effects model. Alternatively, if these expectations are not restricted to zero, then the model is said to be a fixed-effects model. For both designs, the N μ_i were drawn from a $N(0, 2.5^2)$ distribution. In the random-effects design, the NT x_{1it} were drawn from a $N(0, 1.5^2)$ distribution, and the NT x_{2it} were drawn from a $N(0, 1.8^2)$ distribution. The fixed-effects design was parameterized by drawing the μ_i, x_{1it} and x_{2it} as in the random-effects case and then redefining x_{1it} to be $x_{1it} + .5\mu_i$ and redefining x_{2it} to be $x_{2it} + .5\mu_i$.

The implementations of the size and power simulations differ between the conditionally homoskedastic and conditionally heteroskedastic cases. In the conditionally homoskedastic case, the size simulations were parameterized by drawing the NT ϵ_{it} from a $N(0, 1)$ distribution. The autoregressive power simulations were parameterized by letting $\epsilon_{it} = \rho\epsilon_{it-1} + \xi_{it}$, where ρ is the autoregressive parameter. The N ϵ_{i0} were drawn from a $N(0, 1)$ distribution, and the NT ξ_{it} were drawn from a $N(0, 1)$ distribution.[8] The moving-average power simulations were parameterized by letting $\epsilon_{it} = \xi_{it} + \rho\xi_{it-1}$, where ρ is the moving-average parameter, ξ_{i0} were drawn from a $N(0, 1)$ distribution, and the NT ξ_{it} were drawn from a $N(0, 1)$ distribution.[9]

For the conditionally heteroskedastic case, two modifications were made. For the size simulations, the NT ϵ_{it} were drawn from a $N(0, .64(|x_{1it}| + |x_{2it}|)^2)$ distribution. For the autoregressive and moving-average power simulations, the NT ξ_{it} were drawn from a $N(0, .64(|x_{1it}| + |x_{2it}|)^2)$ distribution.

5 Results

Tables 1–4 contain the results for the four cases: fixed effects and homoskedastic, fixed effects and heteroskedastic, random effects and homoskedastic, and random effects and heteroskedastic. The empirical rejection frequencies over 2,000 runs and their 95% confidence intervals are reported for the 42 different Monte Carlo experiments in each table.[10]

The rows of the tables correspond to different size or power simulations. The columns of the tables correspond to different sample sizes or structures.

[6] The probability that a panel would have some observations dropped was .4, and the initial time period in truncated panels varied uniformly over $\{2, 3, 4, 5, 6, 7\}$.

[7] The probability that a panel was truncated was .4. If a panel was selected to be truncated then each of the three observations had a probability of being dropped of .6.

[8] To burn-in the time-series process, 290 extra pre-observations were drawn and discarded for each panel.

[9] As in the autoregressive case, to burn-in the time-series process, 290 extra pre-observations were drawn and discarded for each panel.

[10] The confidence intervals were calculated using the exact binomial method. For those cases in which the empirical rejection frequency is 1.00, the one-sided 97.5% confidence interval is reported. See the Stata Reference manual entry [R] **ci** for details.

In the Model column, "AR" stands for "autoregressive", and "MA" stands for moving average. The "Corr." column specifies the value of ρ used for the simulations in that row. In each table, the first row has Model = AR and Corr. = 0. Because an autoregressive model with no correlation is a model with no serial correlation, these are the empirical sizes. In all the size experiments, the nominal size was .05.

Eight general observations can be made about these results.

1. The empirical sizes are all reasonably close to their nominal sizes of .05.
2. For the largest sample size considered, $N = 1000$, $T = 10$, the empirical power is always essentially 100% for all the alternatives considered.
3. When the correlation is .2 or higher, the test has nearly 100% power in all cases.
4. When the errors are conditionally homoskedastic, the test performs equally well for the random-effects and fixed-effects designs.
5. When the errors are conditionally heteroskedastic, the test may have less power in the fixed-effects case than in the random-effects case in small samples with low levels of serial correlation.
6. The addition of conditional heteroskedasticity may cause a small loss of power in small samples for low levels of serial correlation.
7. The size and power properties of the test are not affected by the panels' having an unbalanced structure with or without gaps in the individual series.
8. The test may have more power against the MA alternative than against the AR alternative in small samples.

(*Continued on next page*)

Table 1: Fixed effects and homoskedastic idiosyncratic errors

Model	Corr.	Balanced				No Gaps	With Gaps
		N=500 T=5	N=500 T=10	N=1000 T=5	N=1000 T=10		
AR	0	.05 [.041, .061]	.056 [.046, .066]	.050 [.04, .06]	.052 [.043, .063]	.057 [.047, .068]	.060 [.05, .071]
AR	.1	.713 [.693, .733]	.986 [.98, .991]	.948 [.937, .957]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.997, 1]
AR	.2	.999 [.996, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]
AR	.3	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]
MA	.1	.774 [.755, .792]	.996 [.991, .998]	.968 [.959, .975]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.997, 1]
MA	.2	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]
MA	.3	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]

Table 2: Fixed effects and conditionally heteroskedastic idiosyncratic errors

Model	Corr.	Balanced				No Gaps	With Gaps
		N=500 T=5	N=500 T=10	N=1000 T=5	N=1000 T=10		
AR	0	.066 [.056, .078]	.06 [.05.071]	.054 [.044, .064]	.048 [.039, .058]	.048 [.039, .058]	.049 [.04.059]
AR	.1	.56 [.538, .582]	.949 [.938, .958]	.811 [.793, .827]	.999 [.996, 1]	.994 [.990, .997]	.989 [.983, .993]
AR	.2	.972 [.963, .978]	1.00 [.998, 1]	1.00 [.997, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]
AR	.3	1.00 [.997, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]
MA	.1	.64 [.619, .661]	.972 [.964, .979]	.889 [.874, .902]	1.00 [.998, 1]	.999 [.996, 1]	.996 [.992, .998]
MA	.2	.996 [.992, .998]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]
MA	.3	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]

Table 3: Random effects and homoskedastic idiosyncratic errors

Model	Corr.	Balanced				No Gaps	With Gaps
		N=500 T=5	N=500 T=10	N=1000 T=5	N=1000 T=10		
AR	0	.055 [.045, .066]	.053 [.044.064]	.058 [.048, .069]	.058 [.048, .069]	.057 [.047, .068]	.052 [.042, .062]
AR	.1	.71 [.69, .73]	.985 [.978, .989]	.94 [.928, .95]	1.00 [.998, 1]	1.00 [.997, 1]	1.00 [.998, 1]
AR	.2	.996 [.992, .998]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]
AR	.3	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]
MA	.1	.79 [.771, .807]	.998 [.994, .999]	.978 [.97, .984]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]
MA	.2	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]
MA	.3	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]

Table 4: Random effects and conditionally heteroskedastic idiosyncratic errors

Model	Corr.	Balanced				No Gaps	With Gaps
		N=500 T=5	N=500 T=10	N=1000 T=5	N=1000 T=10		
AR	0	.056 [.046, .066]	.047 [.038.057]	.052 [.042, .062]	.049 [.04, .059]	.054 [.044, .064]	.055 [.045.065]
AR	.1	.617 [.595, .638]	.98 [.972, .985]	.885 [.87, .898]	1.00 [.998, 1]	.998 [.994, .999]	.996 [.992, .998]
AR	.2	.989 [.983, .993]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]
AR	.3	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]
MA	.1	.704 [.683, .724]	.99 [.985, .994]	.938 [.926, .948]	1.00 [.998, 1]	1.00 [.997, 1]	1.00 [.998, 1]
MA	.2	.999 [.996, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]
MA	.3	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]	1.00 [.998, 1]

Let us now address each of these observations. Some size distortion, although not too large, does appear with smaller sample sizes in the case of fixed effects and conditionally heteroskedastic idiosyncratic errors. Given this caveat, observation 1 is plausible. Observations 2 and 3 emerge quite clearly from results and need no qualifications.

In the presence of conditionally homoskedastic errors, in only one case out of 42 the empirical rejection frequency for the random-effects case lies outside the 95% confidence interval of the empirical rejection frequency for the corresponding fixed-effects empirical rejection frequency and vice versa. This is interpreted as evidence in support of observation 4.

In contrast, in the conditionally heteroskedastic case, the empirical rejection frequency in the fixed-effects case lies outside the 95% confidence interval of the corresponding random-effects empirical rejection frequency and vice versa in 10 out of 42 cases. In all the power simulations, the empirical rejection frequency in the case of fixed effects and conditionally heteroskedastic errors is less than or equal to the corresponding frequency in the case of random effects and conditionally heteroskedastic errors. This indicates that the test has less power in the case of conditionally heteroskedastic errors and fixed effects than it does in the case of conditionally heteroskedastic errors and random effects. However, the loss of power is small and only matters for smaller sample sizes at low levels of serial correlation, giving rise to observation 5.

In all the power simulations, the empirical rejection frequencies for the fixed-effects conditionally heteroskedastic case are less than or equal to their counterparts in the fixed-effects homoskedastic case. Likewise, in all the power simulations, the empirical rejection frequencies for the random-effects conditionally heteroskedastic case are less than or equal to their counterparts in the random-effects homoskedastic case. However, the difference is small and only matters for smaller sample sizes with low levels of serial correlation. These results give rise to observation 6.

Observation 7 comes from the fact that the unbalanced results are very close to the $N = 1000$, $T = 10$ results. The slight loss of power can be attributed to the smaller sample size. There is no large loss of power or size distortion that would indicate a failure of the test with unbalanced data. Most notably, the presence of gaps in the individual series does not appear to have any effect on the results aside from what can be attributed to the loss of sample size.

That the test shows higher empirical rejection rates against the MA alternative, compared to the AR alternative with the same level of correlation in all four tables, leads to observation 8.

6 Concluding remarks

The simulation results presented in this paper have shown that Wooldridge's test for serial correlation in one-way linear panel-data models can have good size and power properties with reasonably sized samples. The test may need larger sample sizes to achieve the same power in the presence of conditional heteroskedasticity. As with all

Monte Carlo results, these results apply only to models with the properties assumed in the simulations. Confirming these results with alternative parameterizations is a topic for future research.

The Stata command `xtserial` and the Stata do-files used to produce the results documented here are available at `http://www.stata.com/users/ddrukker/xtserial`. Alternatively, typing

```
. findit xtserial
```

in Stata will make the command and the do-files available within Stata. Both the command and the do-files require Stata 8 or higher.

7 References

Baltagi, B. H. 2001. *Econometric Analysis of Panel Data*. 2d ed. New York: John Wiley & Sons.

Baltagi, B. H. and Q. Li. 1995. Testing AR(1) against MA(1) disturbances in an error component model. *Journal of Econometrics* 68: 133–151.

Baltagi, B. H. and P. X. Wu. 1999. Unequally spaced panel data regressions with AR(1) disturbances. *Econometric Theory* 15: 814–823.

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.