

# Cambridge Working Papers in Economics

Cambridge Working Papers in Economics: 1944

## XTSERIALPM: A PORTMANTEAU TEST FOR SERIAL CORRELATION IN A LINEAR PANEL MODEL

Koen Jochmans

Vincenzo Verardi

26 April 2019

We introduce the command `xtserialpm` to perform the portmanteau test developed in Jochmans (2019). The procedure tests for serial correlation in the errors of a linear panel model after estimation of the regression coefficients by the within-group estimator. The test is different from the test of Inoue and Solon (2006) that is performed by `xttstest` (Wursten 2018) in that it allows for heteroskedasticity. In simulations documented below, `xtserialpm` is found to provide a much more powerful test than `xttstest`. `xtserialpm` can deal with unbalanced panel data.

# xtserialpm: A portmanteau test for serial correlation in a linear panel model

Koen Jochmans  
University of Cambridge  
Cambridge, United Kingdom  
kj345@cam.ac.uk

Vincenzo Verardi  
Université de Namur  
Namur, Belgium  
vverardi@unamur.be

**Abstract.** We introduce the command `xtserialpm` to perform the portmanteau test developed in [Jochmans \(2019\)](#). The procedure tests for serial correlation in the errors of a linear panel model after estimation of the regression coefficients by the within-group estimator. The test is different from the test of [Inoue and Solon \(2006\)](#) that is performed by `xtistest` ([Wursten 2018](#)) in that it allows for heteroskedasticity. In simulations documented below, `xtserialpm` is found to provide a much more powerful test than `xtistest`. `xtserialpm` can deal with unbalanced panel data.

**Keywords:** `xtserialpm`, heteroskedasticity, fixed-effect model, portmanteau test, serial correlation, short panel data, unbalanced panel.

**Date:** April 26, 2019

## 1 Introduction

Consider panel data on an outcome  $y_{g,i}$  and a set of covariates  $\mathbf{x}_{g,i}$ , where  $g = 1, \dots, G$  and  $i = 1, \dots, N$ . The data are independent across groups  $g$  but may be dependent within groups. The workhorse specification to analyse such data is the regression model

$$y_{g,i} = \mathbf{x}_{g,i}^\top \boldsymbol{\beta} + \eta_{g,i}, \quad \eta_{g,i} = \alpha_i + \varepsilon_{g,i},$$

where  $\alpha_i$  is a latent individual effect and  $\varepsilon_{g,i}$  is an idiosyncratic disturbance whose mean is normalized to zero. These disturbances are taken to be mean independent of the regressors and the individual effects but are otherwise allowed to be (conditionally) heteroskedastic and correlated within each group  $g$ . Although we do not make it explicit in the notation, everything to follow applies to settings where the panel is unbalanced provided, of course, that missingness is at random.

The command `xtserialpm` performs a portmanteau test for the (multivariate) null

$$\mathbf{H}_0 : \mathbb{E}(\varepsilon_{g,i} \varepsilon_{g,i'}) = 0 \text{ for all } i \neq i',$$

against the alternative

$\mathbf{H}_1 : \mathbb{E}(\varepsilon_{g,i}\varepsilon_{g,i'}) \neq 0$  for some  $i \neq i'$ .

This is done via the test developed in [Jochmans \(2019\)](#) after estimation of  $\beta$  by the within-group estimator.<sup>1</sup>

Several other tests for serial correlation in linear panel models with group effects are available. `xtserial` ([Drukker 2003](#)) implements the test for first-order autocorrelation suggested in [Wooldridge \(2002, pp. 282–283\)](#) and `xthrttest` ([Wursten 2018](#)) implements a heteroskedasticity-robust version of this test (see [Born and Breitung 2016](#)). `xtqptest` ([Wursten 2018](#)) performs a generalization of the Wooldridge test for correlation up to a given order. The portmanteau test of [Inoue and Solon \(2006\)](#) is available in `xtistest` ([Wursten 2018](#)) but this assumes homoskedasticity and does not use all information available under the null (see [Jochmans 2018](#)). This has implications for power, as we illustrate in simulations below. See [Wursten \(2018\)](#) for numerical evaluations of these tests.

## 2 The test statistic

The approach followed in [Jochmans \(2019\)](#) is based on the observation that a test of the above hypothesis is equivalent to a test of the null that the *difference* between all pairwise within-group correlations are zero. Noting that

$$\mathbb{E}(\eta_{g,i'}\Delta\eta_{g,i}) = \mathbb{E}(\eta_{g,i'}\varepsilon_{g,i}) - \mathbb{E}(\eta_{g,i'}\varepsilon_{g,i-1}) = \mathbb{E}(\varepsilon_{g,i'}\varepsilon_{g,i}) - \mathbb{E}(\varepsilon_{g,i'}\varepsilon_{g,i-1}),$$

for all  $i$  and  $i'$ , where we let  $\Delta$  denote the first-differencing operator, we can then test that

$\mathbf{H}_0 : \mathbb{E}(\eta_{g,i'}\Delta\eta_{g,i}) = 0$  for all  $i$  and each  $i' \leq i - 2$  and  $i' = i + 1$ ,

against the alternative

$\mathbf{H}_1 : \mathbb{E}(\eta_{g,i'}\Delta\eta_{g,i}) \neq 0$  for some  $i$  and  $i' \leq i - 2$  or  $i' = i + 1$ .

This null involves  $q := (N + 1)(N - 2)/2$  restrictions that can be written compactly as

$$\mathbb{E}(\mathbf{H}_g^\top \Delta\eta_g) = \mathbf{0},$$

---

1. The test in [Jochmans \(2019\)](#) is applicable more generally but we focus on the static regression model and the within-group estimator here.

where we have introduced the  $(N-1) \times q$  matrix

$$\mathbf{H}_g := \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & \eta_{g,3} & 0 & \cdots & 0 \\ \eta_{g,1} & 0 & 0 & 0 & & 0 & & 0 & 0 & \eta_{g,4} & & \vdots \\ 0 & \eta_{g,1} & \eta_{g,2} & 0 & & 0 & & 0 & \vdots & & \ddots & 0 \\ \vdots & & & \ddots & & & & \vdots & 0 & 0 & & \eta_{g,N} \\ 0 & 0 & 0 & 0 & \cdots & \eta_{g,1} & \cdots & \eta_{g,N-2} & 0 & 0 & \cdots & 0 \end{pmatrix},$$

and the  $(N-1) \times 1$  vector  $\Delta\boldsymbol{\eta}_g := (\Delta\eta_{g,2}, \dots, \Delta\eta_{g,N})^\top$ . The left block of the matrix  $\mathbf{H}_g$  is reminiscent of the instrument matrix for GMM estimator of dynamic panel models (see, e.g., [Arellano 2003](#), pp. 88–89). The right block does not appear there as it would not provide valid moment conditions in that context. The null can be tested using a minimum-distance statistic in a sample version of the moment restrictions as soon as three observations per group are available. In the case of only two observations per group the null can always be rationalized due to the presence of the group effects. Note that the dimension of the null grows with  $N$ . As such, the approach is designed for short panels, where  $q$  is small compared to  $G$ .

To make the test operational an estimator of  $\boldsymbol{\beta}$  is needed. `xtserialpm` uses the within-group least-squares estimator,

$$\mathbf{b} := \left( \sum_{g=1}^G \mathbf{X}_g^\top \mathbf{M} \mathbf{X}_g \right)^{-1} \sum_{g=1}^G \mathbf{X}_g^\top \mathbf{M} \mathbf{y}_g,$$

where we have collected all observations for a given group in  $\Delta\mathbf{y}_g := (\Delta y_{g,2}, \dots, \Delta y_{g,N})^\top$  and  $\mathbf{X}_g := (\mathbf{x}_{g,1}, \dots, \mathbf{x}_{g,N})^\top$ , and  $\mathbf{M}$  denotes the usual  $N \times N$  projection matrix that transforms observations into deviations from within-group means. Given  $\mathbf{b}$ , the residuals

$$e_{g,i} := y_{g,i} - \mathbf{x}_{g,i}^\top \mathbf{b}$$

can be used as estimators of the  $\eta_{g,i}$ .

We then define

$$\mathbf{v}_g := \mathbf{E}_g^\top \Delta \mathbf{e}_g - \left( \sum_{h=1}^G \mathbf{E}_h^\top \Delta \mathbf{X}_h \right) \left( \sum_{h=1}^G \mathbf{X}_h^\top \mathbf{M} \mathbf{X}_h \right)^{-1} \mathbf{X}_g^\top \mathbf{M} \mathbf{e}_g,$$

where  $\mathbf{E}_g$  denotes the plug-in estimator of  $\mathbf{H}_g$  using the residuals  $e_{g,i}$  instead of  $\eta_{g,i}$  and we have introduced the notation  $\mathbf{e}_g := (e_{g,1}, \dots, e_{g,N})^\top$ ,  $\Delta \mathbf{e}_g := (\Delta e_{g,2}, \dots, \Delta e_{g,N})^\top$ , and  $\Delta \mathbf{X}_g := (\Delta \mathbf{x}_{g,2}, \dots, \Delta \mathbf{x}_{g,N})^\top$ . The test statistic for our null can then be written as the quadratic form

$$s := \mathbf{v}^\top \mathbf{V}^{-1} \mathbf{v},$$

where  $\mathbf{v} := \sum_{g=1}^G \mathbf{v}_g$  and  $\mathbf{V}$  is an estimator of the variance of  $\mathbf{v}$ . `xtserialpm` uses the uncentered estimator

$$\mathbf{V} := \sum_{g=1}^G \mathbf{v}_g \mathbf{v}_g^\top$$

as default. Use of a centered variance estimator instead is available as an option.

Under the null,

$$s \xrightarrow{d} \chi_q^2,$$

as  $G \rightarrow \infty$ . While

$$\sum_{g=1}^G \mathbf{v}_g = \sum_{g=1}^G \mathbf{E}_g^\top \Delta \mathbf{e}_g,$$

the test statistic needs to use the sample variance of  $\mathbf{v}_g$  and not of  $\mathbf{E}_g^\top \Delta \mathbf{e}_g$  for this distributional approximation to hold. The second right-hand side term in  $\mathbf{v}_g$  is needed to account for the fact that the statistic is computed from residuals rather than the (latent) errors. `xtserialpm` can be run directly on the outcomes rather than residuals from a within-group regression. In this case, no correction for within-group estimation is needed.

### 3 Stata command `xtserialpm`

`xtserialpm` is a stand-alone routine that can run without first running `xtreg`. The data must be `tsset` prior to executing `xtserialpm`. Unbalanced panel data is allowed.

The command has the following syntax:

```
xtserialpm depvar[indepvars][if][in], [center]
```

The option `center` returns the test statistic computed with centered covariance matrix as discussed above.

Running the command produces a table with the value of the test statistic and the associated  $p$ -value. The layout of the table mimics the layout of the table produced by `xtserial`.

The following output is saved to `r`:

`r(stat)` returns the value of the test statistic;

`r(df)` returns the degrees of freedom of its limit distribution;

`r(p)` returns the  $p$ -value of the test.

Help is available by typing `help xtserialpm`.

## 4 Example

We use the data from the illustration in [Drukker \(2003\)](#). The following extract loads and manipulates the data.

```
. use http://www.stata-press.com/data/r8/nlswork.dta
(National Longitudinal Survey. Young Women 14-26 years of age in 1968)
. xtset idcode year
      panel variable:  idcode (unbalanced)
      time variable:  year, 68 to 88, but with gaps
                  delta: 1 unit
. drop if year>70
(24,241 observations deleted)
. gen age2 = age^2
(9 missing values generated)
. gen tenure2 = tenure^2
(147 missing values generated)
```

The portmanteau test is computed as

```
. xtserialpm ln_wage age* ttl_exp tenure* south
```

and delivers the output

```
Jochmans portmanteau test for within-group correlation in panel data.
H0: no within-group correlation
Chi-sq( 2) =      25.658
Prob > Chi-sq =      0.0000
```

The result provides strong evidence for the presence of serial correlation in the errors.

To compute the test statistic using a centered covariance matrix estimator use the `center` option as

```
. xtserialpm ln_wage age* ttl_exp tenure* south, center
```

The output is

```
Jochmans portmanteau test for within-group correlation in panel data.
H0: no within-group correlation
Chi-sq( 2) =      26.180
Prob > Chi-sq =      0.0000
```

The test statistic is slightly larger and our initial conclusion unaltered.

To perform the test of [Inoue and Solon \(2006\)](#) in this example we first generate residuals

from the within group regression by typing

```
. quietly xtreg ln_wage age* ttl_exp tenure* south, fe
. predict u, residuals
```

The test is then performed on these residuals

```
. xttestest u, lags(all)
```

By default, `xttestest` checks for serial correlation up to second order only. Here, `xttestest` is invoked with the lag option set to `all`, so that the command yields the proper portmanteau test of [Inoue and Solon \(2006\)](#).

The output of the test is

```
Inoue and Solo (2006) LM-test on variables u
Panelvar: idcode
Timevar: year
```

Variable	IS-stat	p-value	N	maxT	balance?
u	159.44	0.000	2206	3	gaps

```
Notes: Under H0, LM ~ chi2((T-1)(T-2)/2)
H0: No auto-correlation of any order.
Ha: Auto-correlation of some order.
```

The same conclusion regarding our null is reached.

## 5 Simulations

We provide a size and power comparison between `xtserialpm` and `xttestest`. For numerical comparisons between `xttestest` and the other routines available in Stata, see [Wursten \(2018\)](#).

Outcomes were generated with fixed effects drawn from a standard normal and with two regressors—the first standard normal and the second zero/one according to the toss of a fair coin—each with a coefficient set to unity. We report results for sample sizes  $G = 100$  and  $N = 3$  and consider four different specifications for the errors. In Specification **(A1)** the errors follow a stationary moving-average process of order one,

$$\varepsilon_{g,i} = \epsilon_{g,i} + \theta \epsilon_{g,i-1},$$

with independent standard-normal innovations. In Specification **(B1)** the errors follow

a stationary autoregressive process of order one,

$$\varepsilon_{g,i} = \rho \varepsilon_{g,i-1} + \epsilon_{g,i},$$

again with standard-normal innovations. Specifications **(A2)** and **(B2)** differ from, respectively, **(A1)** and **(B1)** in that the initial error is fixed as opposed to being drawn from its steady-state distribution. More precisely, we set  $\epsilon_{g,0} = 0$  in **(A2)** and  $\varepsilon_{g,1} = 0$  in **(B2)**. The non-stationarity introduces heteroskedasticity only under the alternative in **(A2)** and under both null and alternative in **(B2)**.

Simulation results are provided in the form of power curves. Power is plotted against the (stationary) first-order autocorrelation coefficient of the errors. In the autoregressive specification this is simply  $\rho \in (-1, 1)$ , the autoregressive coefficient. For the moving average process the first-order autocorrelation, say  $\rho$ , is a function of  $\theta$ . We plot power against  $\rho \in (-\frac{1}{2}, \frac{1}{2})$ , which are all values for which the covariance matrix of the errors has full rank. This is done by setting

$$\theta = \frac{1 + \sqrt{1 - 4\rho^2}}{2\rho}$$

when  $\rho \neq 0$  and  $\theta = 0$  if  $\rho = 0$ . Figure 1 provides the power functions of `xtserialpm` (solid line) and `xtistest` (dashed line) along with a horizontal line that indicates the nominal size of the test (dashed-dotted). The nominal size of both tests was set to .05.

The results for Specifications **(A1)** and **(B1)** in the upper two plots of Figure 1 show that both tests are approximately size correct and unbiased. They also reveal a rather dramatic difference in power between the two tests, with `xtserialpm` providing an (almost) uniformly more powerful test than `xtistest`, and this by a large margin. Both tests have some difficulty rejecting the null when the errors process is close to a random walk.

The results for Specifications **(A2)** and **(B2)** in the bottom of Figure 1 reveal the impact of non-stationarity on the performance of both tests. In Specification **(A2)** both tests are size correct. `xtserialpm` is again the more powerful of the two tests, especially against negative patterns of autocorrelation. Close to the null `xtistest` is slightly more powerful. This can be understood by recalling that this test interprets heteroskedasticity as evidence against the null so that it boosts power. In Specification **(B2)** only `xtserialpm` is size correct. `xtistest` delivers a biased test that is severely size distorted. This, again, is a consequence of heteroskedasticity in the errors which,



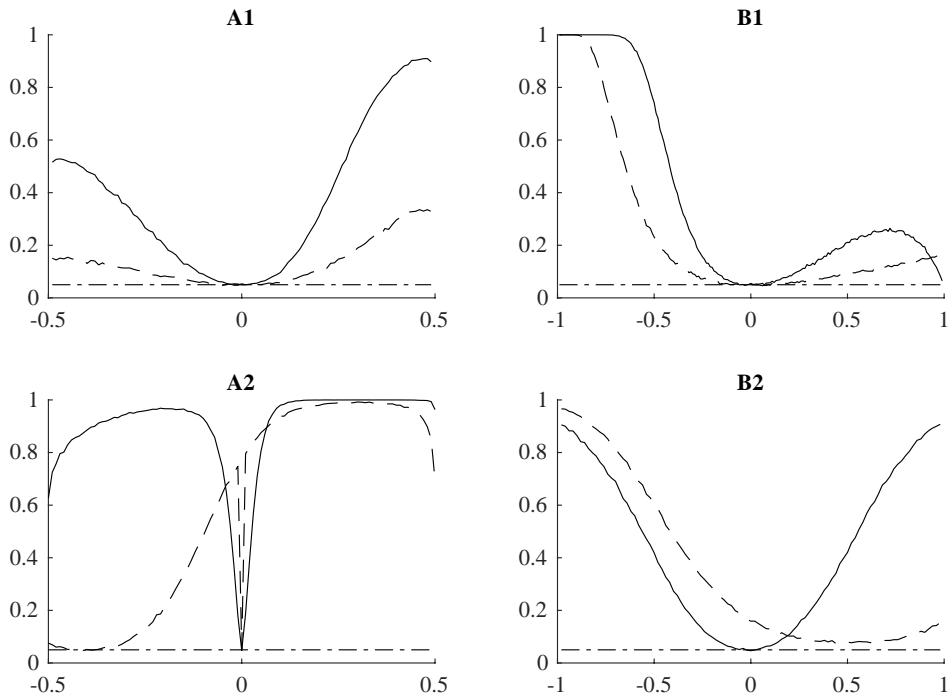


Figure 1: Simulated power curves

contrary to in Specification (**A2**), is now also present under the null. This makes the size properties of the test break down.

Taken together, these results show that `xtserialpm` provides the more reliable test. Size is controlled in all designs and any violation of the null tends to be spotted with high probability.

## 6 Conclusion

We have introduced the command `xtserialpm`. It can be used to test for arbitrary patterns of serial correlation in the errors of a fixed-effect regression model. Contrary to the existing `xtistest`, it is robust to heteroskedasticity. A simulation study further shows `xtserialpm` to provide a considerably more powerful test.

## 7 References

- Arellano, M. 2003. *Panel Data Econometrics*. Oxford University Press.
- Born, B., and J. Breitung. 2016. Testing for serial correlation in fixed-effects panel data models. *Econometric Reviews* 35: 1290–1316.
- Drukker, D. M. 2003. Testing for serial correlation in linear panel-data models. *Stata Journal* 3: 168–177.
- Inoue, A., and G. Solon. 2006. A portmanteau test for serially correlated errors in fixed effects models. *Econometric Theory* 22: 835–851.
- Jochmans, K. 2018. A portmanteau test for correlation in short panels. Cambridge Working Paper in Economics 18/86.
- . 2019. Testing for correlation in error-component models. Cambridge Working Paper in Economics 19/10.
- Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press.
- Wursten, J. 2018. Testing for serial correlation in fixed-effects panel models. *Stata Journal* 18: 76–100.

### Acknowledgments

Koen Jochmans gratefully acknowledges financial support from the European Research Council through Starting Grant n° 715787. Vincenzo Verardi gratefully acknowledges financial support from the FNRS. Comments and suggestions by Helmut Farbmacher on the test implemented here were much appreciated.

### About the authors

Koen Jochmans is a Reader in the Faculty of Economics at the University of Cambridge. Vincenzo Verardi is Associate Researcher of the FNRS.