

Five ways to detect correlation in panels

Jesse Wursten¹

¹`jesse.wursten@kuleuven.be`
Faculty of Economics and Business
KU Leuven

23rd London Stata User Group Meeting, September 2017

Introduction

Get to know your data (and regressions)

- Sometimes difficult to get a grip on larger panels
- 5 new commands to get to know your data (and your regressions)
 - ▶ `xtqptest`, `xthrttest` and `xtistest` test for correlation over time (serial correlation)
 - ▶ `pwcorr` and `xtcdf` test for correlation across panel units (cross sectional dependence)
- Bonus: might indicate you don't need cluster-robust standard errors (useful if you don't have 20+ clusters)

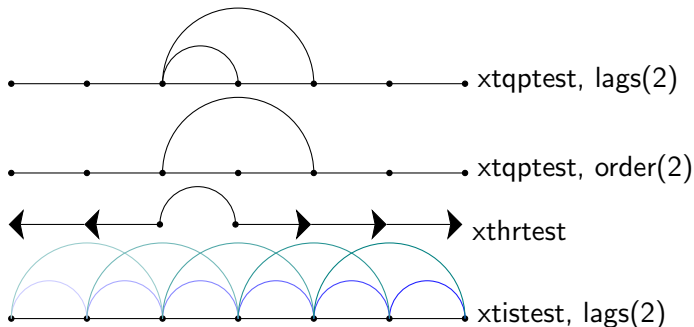
Serial Correlation

Is your data correlated over time?

- To keep things real, imagine you have a panel of calories consumption for 3 individuals (N) over 365 days (T) [sysuse xtline1.dta]
- Is calorie consumption in each day a random draw, or is it correlated over time?
- Does my fixed/random effects model for calorie consumption produce a relatively decent fit?
- Three new commands which improve on current industry standard (i.e. xtserial & abar)
 - ▶ More flexible: not limited to respectively 1st order serial correlation and GMM postestimation
 - ▶ More robust: better power and size in various scenarios

Serial Correlation - Overview

Four pictures say more than a thousand words



Serial Correlation - Q(p) test

A true jack of all trades

- Syntax: `xtqptest [varlist], lags(p)`
- Tests for serial correlation **up to** order p
- Best size/power results in Monte Carlo

Bias-corrected Born and Breitung (2016) Q(p)-test on variables calories

Panelvar: person

Timevar: day

p (lags): 2

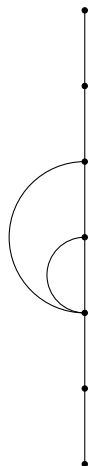
Variable	Q(p)-stat	p-value	N	maxT	balance?
calories	15.84	0.000	3	365	balanced

Notes: Under H_0 , $Q(p) \sim \chi^2(p)$

H_0 : No serial correlation up to order p .

H_a : Some serial correlation up to order p .

- Test indicates there might be some serial correlation up to the 2nd order



Serial Correlation - LM(k) test

Focus on a specific order

- Syntax: `xtqptest [varlist], order(k)`
- Tests for serial correlation **of** order k
- Sometimes more informative than the $Q(p)$ test

Bias-corrected Born and Breitung (2016) LM(k)-test on variables calories

Panelvar: person

Timevar: day

k (order): 2

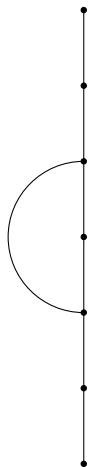
Variable	LM(k)-stat	p-value	N	maxT	balance?
calories	1.63	0.103	3	365	balanced

Notes: Under H_0 , $LM(k) \sim N(0,1)$

H_0 : No serial correlation of order k .

H_a : Some serial correlation of order k .

- Test indicates data might be free of 2nd order serial correlation



Serial Correlation - HR test

When there's the occasional storm

- Syntax: `xthrtest [varlist]`
- Tests for first order serial correlation
- Specialised in situations where the variance changes over time (e.g. stock markets)
- Boils down to regressing forwards demeaned values on lagged backwards demeaned values

Heteroskedasticity-robust Born and Breitung (2016) HR-test on calories

Panelvar: person

Timevar: day

Variable	HR-stat	p-value	N	maxT	balance?
calories	1.65	0.099	3	365	balanced

Notes: Under H_0 , $HR \sim N(0,1)$

H_0 : No first-order serial correlation.

H_a : Some first order serial correlation.

- Test indicates data might be free of 1st order serial correlation



Serial Correlation - IS test

In cases of severe amnesia

- Syntax: `xttstest [varlist], lags(p)`
- Tests for serial correlation **up to** order p
- Accepts any kind of unbalanced data (including gaps)

Inoue and Solo (2006) LM-test on variables calories

Panelvar: person

Timevar: day

p (lags): 2

Variable	IS-stat	p-value	N	maxT	balance?
calories	3.00	1.000	3	365	balanced

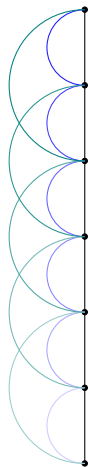
calories: N(3) is smaller than the dimension of H0(727), results unreliable. Consider `xtqptest`.

Notes: Under H_0 , $LM \sim \chi^2(p \cdot T - p(p+1)/2)$

H_0 : No auto-correlation of any order.

H_a : Auto-correlation up to order 2.

- Test only works when $N > p \cdot T$



Cross sectional dependence

Is your data correlated across panel units?

- Remember our panel of three individuals and their eating habits
- Does their calorie intake spike and drop together? (e.g. Sunday Roast)
- Did my fixed/random effects model properly control for unobserved similarities between the individuals (which might otherwise bias the results)?
- Two new commands which improve performance of existing code (i.e. `pwcorr` & `xtcd/xtcd2`)
 - ▶ More flexible: can test multiple variables, which do not need to be mean-zero
 - ▶ More efficient: faster than existing commands

Cross sectional dependence - pwcorr

The f stands for fast

- Syntax: `pwcorr` varname, reshape
- Calculates correlations between panel units
- More convenient and faster than first reshaping and then using `pwcorr`
- Syntax: `pwcorr` varlist
- Calculates correlations between variables
- Faster than `pwcorr` if varlist is long

```
Variable(s): calories
Panel var: person

corrMatrix[3,3]
      1      2      3
1      1      0      0
2  .72879721  1      0
3  .77735323  .90375738  1
```

```
Variable(s): calories day

corrMatrix[2,2]
      calories      day
calories      1      0
day      .152202      1
```

Cross sectional dependence - xtcd

xtdvdf didn't have the same ring to it

- Syntax: xtcd varlist
- CD-test boils down to verifying whether sum of correlations between panel units is equal to zero

```
xtcd test on variables calories
```

```
Panelvar: person
```

```
Timevar: day
```

Variable	CD-test	p-value	average joint T	mean ρ	mean abs(ρ)
calories	26.582	0.000	365.00	0.80	0.80

Notes: Under the null hypothesis of cross-section independence, $CD \sim N(0,1)$
P-values close to zero indicate data are correlated across panel groups.

- Test strongly indicates calorie intake is correlated across individuals
- This is not the first command to perform the CD-test, but ...
 - ▶ xtcsd can only be used as postestimation command
 - ▶ xtcd is slow in larger datasets and reports the wrong number of joint observations
 - ▶ xtcd2 assumes mean-zero variables (residuals) and only takes a single variable at the time

Conclusion

This slide is redundant, yet somehow essential

- Introduced 3 commands to test for correlation over time: `xtqptest`, `xthrttest` and `xtistest`
- ... and two to test for correlation between panel units: `pwcorr` and `xtcdf`
- They are more convenient/flexible/efficient than existing commands
- More info can be found in the Econometrics papers
 - ▶ `xtqptest`, `xthrttest`: Born and Breitung (2016)
 - ▶ `xtistest`: Inoue and Solon (2006)
 - ▶ `xtcdf`: Pesaran (2004)
- Any questions?

- Born, Benjamin, and Jörg Breitung.** 2016. "Testing for Serial Correlation in Fixed-Effects Panel Data Models." Econometric Reviews, 35(7): 1290–1316.
- Inoue, Atsushi, and Gary Solon.** 2006. "A Portmanteau Test for Serially Correlated Errors in Fixed Effects Models." Econometric Theory, 22(5): 835–851.
- Pesaran, M. Hashem.** 2004. "General Diagnostic Tests for Cross Section Dependence in Panels." CESifo Group Munich CESifo Working Paper Series 1229.

Bonus slide: send to smartphone

sendtoslack command in Stata

```
. sysuse auto, clear  
(1978 Automobile Data)
```

```
. reg price mpg
```

Source	SS	df	MS	Number of obs	=	74
Model	139449474	1	139449474	F(1, 72)	=	20.26
Residual	495615923	72	6883554.48	Prob > F	=	0.0000
				R-squared	=	0.2196
				Adj R-squared	=	0.2087
				Root MSE	=	2623.7
Total	635065396	73	8699525.97			

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
mpg	-238.8943	53.07669	-4.50	0.000	-344.7008	-133.0879
_cons	11253.06	1170.813	9.61	0.000	8919.088	13587.03

```
. mat results = r(table)
```

```
. mat b_mpg = results["b", "mpg"]
```

```
. local b_mpg = round(b_mpg[1,1], 0.001)
```

```
. mat p_mpg = results["pvalue", "mpg"]
```

```
. local p_mpg = round(p_mpg[1, 1], 0.001)
```

```
. sendtoslack, message(MPG: `b_mpg' (p: `p_mpg')) url(https://hooks.slack.com/services/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)  
Message sent: MPG: -238.894 (p: 0)
```

Bonus slide: send to smartphone

Slack app on smartphone

1:31 PM 0.02K/s 48%

s StataMessage

@danferno (you)

statamessage APP 1:17 PM
test

statamessage APP 1:26 PM
This program is the best!

Today

statamessage APP 12:41 PM
Stata has done something you wanted to know about!

Oh no, error 198 occurred, you might want to fix this!

Oh no, error 198 occurred, you might want to fix this!

NEW

statamessage APP 1:31 PM
MPG: -238.894 (p: 0)