

Some Life Rings

Jan Ditzen

Free University of Bozen-Bolzano, Bozen, Italy

www.jan.ditzen.net, jan.ditzen@unibz.it

<https://janditzen.github.io/>

September 25, 2025

Motivation I

Not so much econometrics...

- Empirical work often relies on small programs which are overlooked.
- This presentation puts the spotlight on my personal favourites which are collected in a package called `xttools`
 - 1 `xtgetpca`
 - 2 `xtplot2`
 - 3 `xtcorri`
- Work in progress, but get in touch if you want to have a working version.

Motivation II

Not so much econometrics...

- Stata comes with some very handy tools to analyse and process panel data.
- However, some functions from `ts` are missing.
- `xttools` consists of three programs:
 - ▶ `xtgetpca` obtain principal components from panel data.
 - ▶ `xtcorri` calculate unit specific correlations or covariances.
 - ▶ `xtplot2` identify panel structure for variables

xtgetpca

PCs in panel data

- pca extracts principal components (PC) from a variable list.
- Not possible with repeated samples (panel data).
- xtgetpca internally reshapes the data into wide format, extracts the PCs and then adds them to the using dataset.
- Allows standardisation with respect to overall, unit and time dimension.
- Can predict score, fit, residual and q from pca.
- Wide dataset can be copied into frame.
- Supports unbalanced panels.
- Mata and Python implementations for speed improvements.

xtgetpca

PCs in panel data

Syntax:

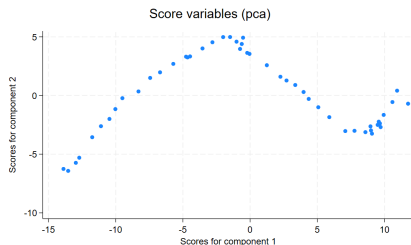
```
xtgetpca [varlist] , num(real) [name(string) standardize(string)  
frame(string) predict(string) covariance correlations output mata  
python]
```

- num() defines number of PCs
- name() prefix for new variable, default PCA_
- stand() standardizes data
- frame() copies wide data to frame
- predict() adds prediction to data, default is score
- output display output from pca
- mata obtain PCs using mata
- python obtain PCs using Python (only balanced panels)

xtgetpca

Example

```
. use "xtdcce2_sample_dataset.dta" , clear
. timer on 1
. xtgetpca log_rgdp, frame(test) num(2)
Dataset used for cacluation of PCs copied to frame test
Variables PCA_1 PCA_2 containing score(s) added to dataset.
. timer off 1
. return list
scalars:
      r(rho) = .7894428507822338
macros:
      r(predict) : "score"
      r(frame)   : "test"
. frame test: scoreplot
. timer on 2
. qui xtgetpca log_rgdp, frame(test) num(2) mata
. timer off 2
. timer on 3
. qui xtgetpca log_rgdp, frame(test) num(2) python
. timer off 3
. timer list
1:    0.33 /      1 =    0.3330
2:    0.06 /      1 =    0.0600
3:    0.12 /      1 =    0.1220
```



xtcorri |

Unit specific correlations

- When working with panel data, often correlation between variables on unit level are of interest, i.e. $\text{corr}(X_i, Y_i)$
- `by id, sort: corr variables` does the job, but contains too much information and output not clear.
- `xtcorri var1 varlist` calculates the correlation on a unit level between *var1* and variables specified in *varlist*.
- Also calculates overall correlation and correlation between cross-section averages.

xtcorri II

Unit specific correlations

- Syntax:

```
xtcorri var1 [varlist] [if], [ivar(varname) cov fmt(string)  
plotoptions(string) combine(string) nocsa ycsa ]
```

- `ivar()` specifies unit name
- `cov` calculate covariance rather than correlation
- `fmt` set format
- `plot[(options)]` plot scatterplot, pass options
- `combine(options)` pass options to graph combine
- `nocsa` do not calculate cross-section averages
- `ycsa` convert *var1* to cross-section averages

xtcorri

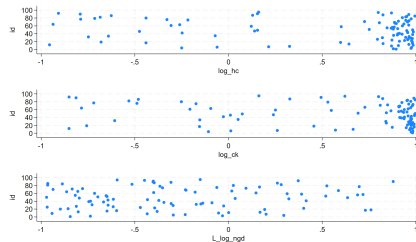
Example

```
. xtcorri log_rgdpo log_hc log_ck L.log_ngd, ivar(country) plot combine(rows(3))
```

Correlations with log_rgdpo on unit level

	log_hc	log_ck	L.log_ngd
Country_1	.8948912	.9824009	-.8446966
Country_2	.8798358	.979756	-.7372838
Country_3	.9743094	.9754843	-.0320244

...			
Country_93	.8699171	.8010562	-.4484579
Total	.7993354	.6211605	-.5942171
CSA	.1608569	.1620696	-.15088



xtplot2

Advanced plots for panel data

- First stage of any empirical project should be data investigation.
- Identifying outlier and missing values in large datasets difficult.
- xtplot2 is inspired by panelview, but quicker to handle.
- It is essentially a wrapper for Ben Jann's heatplot.
- Syntax:

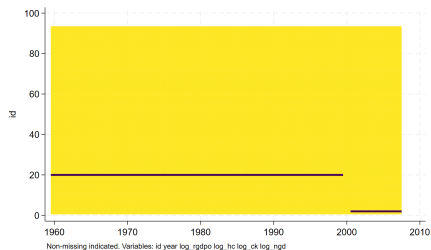
```
xtplot2 [varlist] [if], [values combineopt(string)
ivar(varlist) title seperate ]
```

- values plots values rather than indicating missing values.
- combine() passes options to graph combine.
- ivar() specifies unit name.
- title uses variable label as title
- seperate plot each variable separate.

xtplot2

Example

xtplot2



xtplot2 log_rgdp, values

