

Reshaping the World Development Indicators (WDI) for Panel Data and Seemingly Unrelated Regression Modeling

Summer North American Stata Users' Group Meeting 2008

By P. Wilner Jeanty

Dept. of Agricultural, Environmental, and Development Economics
The Ohio State University

July 24-25, 2008

Outline

- Background
- Delivery forms of WDI
- Previous work on reshaping WDI from wide to long form
- The new Stata program: WDIRESHAPE
- Data preparation
- Examples
- Summary
- Conclusions

Background

- The WDI compilation is a rich and widely used dataset published annually by the World Bank.
- 2008 WDI online version was published in April 2008.
- 2008 WDI CD-ROM version will be available.
- With an online subscription, users can download statistical data for over 800
- development indicators and time series data from 1960 to 2007 for 227 economies.
- Data contains social, economic, financial, natural resources, and environmental indicators.
- Examples of research questions that can be answered using WDI:
 - What are the magnitudes and patterns of poverty in the developing world?
 - How has poverty evolved over the past decade?
 - Why do some countries grow faster than others?
 - Why are some countries richer than others?
 - Does foreign direct investment affect income inequality?
- WDI software very flexible and easy to use

Delivery forms of WDI

- Depending on whether one chooses to download the data with series or countries in row and time in column, or time in row and series or countries in column, or any of such combinations, several forms can result.
- None of the structures obtained can be used immediately for statistical analysis. Data management is needed and can be frustrating.
- In addition, when downloading the data, users are not prompted to choose variable names of their liking for the series.
- However, downloaded data come with series descriptors, which is very good.
- In an efficient way, WDI users would want to have a form ready for immediate statistical analysis, be able to name the series with names of their choosing, and set the series descriptors as Stata variable labels in the final dataset.
- WDIRESHAPE addresses these issues.

Previous work on reshaping WDI from wide to long form

- Baum and Cox (2007) explicitly show how WDI users can apply the reshape command successively to obtain an XT modeling structure .
- Kossinets (2006) provides Stata code to reshape a WDI dataset one variable at a time for panel data analysis.
- On the web (including Statalist), lots of pieces of Stata code can be found to reshape WDI to long form.
- Now WDI users can download the data in long form for XT modeling. Thus, there is no need for the above. Still, when using this form, serious data management is needed before XT analysis can, in fact, be done.
- None of the above have addressed all the issues taken up by WDIRESHAPE

The WDIRESHAPE command

- WDIRESHAPE, a wrapper of the Stata's RESHAPE command, takes a WDI dataset downloaded with countries or series in row and time in column and reshapes it into a format suitable for panel data , seemingly unrelated regression, or change analysis.
- Users provide names of their choosing for the series, and in the reshaped dataset the WDI series descriptors are automatically attached to the variable names supplied.
- Users can choose to reshape the whole dataset at once (provided that there is enough memory to do so), one variable at a time, one year at a time, 5 years at a time, or 10 years at a time.
- For more details, after installation, type in: help wdireshape

Data preparation

- No data preparation is needed for datasets obtained from the online version of WDI. In downloaded datasets, the years are prepended with two letters, “YR”.
- The csv file generated by the WDI software (if csv was the format chosen by the user) can be imported directly into Stata using the `INSHEET` command.
- However, the WDI missing value symbols, the double dots (`..`), must be dealt with. At the user’s request, `WDIRESHAPE` will take care of them.
- Users of 2005 WDI, CD-ROM version, for example, may need to prepend the years with one or two letters before importing the data into Stata. They can choose to have blanks or double dots (`..`) to represent missing values when querying the WDI software.

Syntax of WDIRESHAPE

- `wdireshape varlist, prepend(string)
ctyname(varname) sername(varname)
[other options]`

Options

- prepend(string) indicates the prepending letter(s)
- ctyname(varname) sets the variable holding the names of the countries
- sername(varname) sets the variable holding the names of the series
- ctycode(varname) indicates the name of the variable containing the countries' codes
- sercode(varname) indicates the name of the variable holding the series' codes
- byper(#) causes the dataset to be reshaped by a #-year sub-period (#=1, 5, and 10)
- startyr(#) indicates the start year of the time-series
- endyr(#) indicates the end year of the time-series
- byvar causes the dataset to be reshaped one variable at a time
- sur causes the dataset to be reshaped in wide form for SUREG-type analysis
- cross causes the dataset to be reshaped in wide form for cross-sectional or change analysis
- nstring(#) sets the number of string variables in the dataset before series begin in order to have the double dots removed

Examples

• 1) Reshape everything at once for XT modeling

- `wdireshape fdingdp fdincur fdiout gdp trade, prep(x) ctyn(country_name) sern(series_name) serc(series_code) ///
ctyc(country_code)`

To estimate a fixed effects model, I type in the following:

- `xtreg gdp fdingdp fdi fdiout trade, fe`

• 2) Reshape 5 years at a time for SUREG analysis

- `wdireshape agland tractsk fertilha gdpcnst gdpcur gdppg irrigpct croplnd popdens popg ruraldens trade urbp, ///
prep(yr) ctyn(countryname) sern(seriesname) start(1961) end(2006) byp(5) sur ns(4)`

To conduct a SUREG analysis for five Latin American countries c11, c18, c66, c67, and c86; I type in the following:

- `foreach num of numlist 11 18 66 67 85 {
 local eqn "" eqn' (gdppgc`num' L.gdppgc`num' tractskc`num' fertilhac`num' croplndc`num' /// popgc`num') "
}`
- `sureg `eqn', dfk corr`

• 3) Reshape one variable at a time for change or cross-sectional analysis

- `wdireshape agland tractsk fertilha gdpcnst gdpcur gdppg irrigpct croplnd popdens popg ruraldens trade urbp, ///
prep(yr) ctyn(countryname) sern(seriesname) byv cros ns(4)`

Now suppose I want to calculate, for each country, percentage change in population density from 1990 to 2000, for example, I just type in the following:

- `gen popdenssch=((popdens2000-popdens1990)/popdens1990)*100`

In summary

- WDIRESHAPE eliminates/reduces data management tasks when WDI users want to :
 - 1) Conduct a panel data analysis (default)
 - Run WDIRESHAPE without specifying the SUR and CROS options
 - 2) Analyze a time series of averages across countries
 - First run WDIRESHAPE, and then run COLLAPSE by the variable containing the years
 - 3) Analyze a series of averages across years (pure cross-sections)
 - First run WDIRESHAPE, and then run COLLAPSE by the variable containing the names of the countries
 - 4) Analyze a series of p-year averages (with $p=2, 3, \dots, 10$)
 - First run WDIRESHAPE, and then run [PAVERAGE](#) if installed
 - 5) Conduct a seeming unrelated regression analysis
 - Run WDIRESHAPE with SUR option
 - 6) Perform a change or cross-sectional analysis
 - Run WDIRESHAPE with CROS option
 - 7) Operate on and manipulate WDI as a panel dataset of countries
 - Run WDIRESHAPE, and apply Baum's (2006, chap. 3) suggestions as to how to operate on and manipulate panel datasets in Stata.

Syntax and options for PAVERAGE

- Syntax
 - `paverage varlist, p(#) indiv(varname) yr(varname)`
- Options
 - `p(#)` indicate the number of year averages need to be calculated. # ranges from 2 to 10. For example, specifying `p(5)` will create a 5-year average dataset
 - `indiv(varname)` specifies the name of the variable holding the countries, individuals, or firms.
 - `yr(varname)` specifies the name of the variable containing the years
 - After installation, type in: `help paverage`

Websites

- To know more about WDI:
 - <http://go.worldbank.org/UoFSM7AQ40>
- To download a small WDI dataset to try WDIRESHAPE:
 - <http://ddp-ext.worldbank.org/ext/DDPQQ/member.do?method=getMembers&userid=1&queryId=135>

Conclusions

- WDIRESHAPE equips WDI users with efficient ways to reshape WDI and obtain a dataset suitable for immediate statistical analysis.
- Data management time and effort are reduced significantly when WDI users want to perform several types of analyses
- Users only need to provide, in a *varlist*, names of their choosing for the series and choose a reshaping method and the form into which the dataset should be reshaped.
- In the reshaped dataset, the WDI series descriptors are automatically attached to the user-supplied variable names as labels regardless of the reshaping methods or forms requested by the user.
- However, there are issues related to using panel data methods with cross country data. For instance, see Maddala (1999).

References

- Baum, C. F. 2006. *An Introduction to Modern Econometrics Using Stata*. College Station, TX: Stata Press.
- Baum, C. F., and N. J. Cox. 2007. Stata tip 45: Getting those Data into Shape. *The Stata Journal* 7: 268-271.
- Kossinets, G. 2006.
<http://www.columbia.edu/acis/eds//statpark/stata/tips/reshape-manyvar>.
- Maddala, G.S. 1999. “On the Use of Panel Data Methods with Cross Country. Data,” *Annales d’Economie et de Statistique*, 55-56, 429-448.
<http://www.pse.ens.fr/adres/anciens/n5556/vol5556-18.pdf>

Acknowledgments

- This work was partly funded by a grant from the Mershon Center at the Ohio State University (OSU).
- Support from the Swank Program headed by Dr. Mark Partridge is also gratefully acknowledged.



THANKS !!!