Using the Standardized World Income Inequality Database

Frederick Solt
Assistant Professor of Political Science
University of Iowa
frederick-solt@uiowa.edu

The Standardized World Income Inequality Database (SWIID) uses a custom missing-data multiple-imputation algorithm to standardize observations collected from the United Nations University's World Income Inequality Database version 2.0c, the OECD Income Distribution Database, the Socio-Economic Database for Latin America and the Caribbean generated by CEDLAS and the World Bank, Eurostat, the World Bank's PovcalNet, the UN Economic Commission for Latin America and the Caribbean, the World Top Incomes Database, the University of Texas Inequality Project, national statistical offices around the world, and many other sources. Luxembourg Income Study data serves as the standard.

As described in Solt (2009), the SWIID maximizes the comparability of available income inequality data for the broadest possible sample of countries and years. But incomparability remains, and it is sometimes substantial. This remaining incomparability is reflected in the standard errors of the SWIID estimates, making it absolutely crucial to take this uncertainty into account when making comparisons across countries or over time (Solt 2009, 238; Solt 2014, 21). Using previous versions of the SWIID, however, incorporating the standard errors into an analysis required considerable effort. It is now very straightforward.

In version 5.0 of the SWIID, the inequality estimates and their associated uncertainty are represented by 100 separate imputations of the complete series: for any given observation, the differences across these imputations capture the uncertainty in the estimate. The SWIIDv5_0.zip includes files that

are pre-formatted for use with the tools developed for analyzing such multiply imputed data in Stata (SWIIDv5_0.dta) and in R (SWIIDv5_0.RData).

Each file includes the following four variables:

- `gini_net`: Estimate of Gini index of inequality in equivalized (square root scale) household disposable (post-tax, post-transfer) income, using Luxembourg Income Study data as the standard.
- `gini_market`: Estimate of Gini index of inequality in equivalized (square root scale) household market (pre-tax, pre-transfer) income, using Luxembourg Income Study data as the standard.
- `abs_red`: Estimated absolute redistribution, the number of Gini-index points market-income inequality is reduced due to taxes and transfers: the difference between the `gini_market` and `gini_net`.
- `rel_red`: Estimated relative redistribution, the percentage reduction in market-income inequality due to taxes and transfers: the difference between the `gini_market` and `gini_net`, divided by `gini_market`, multiplied by 100.

1 Using Stata

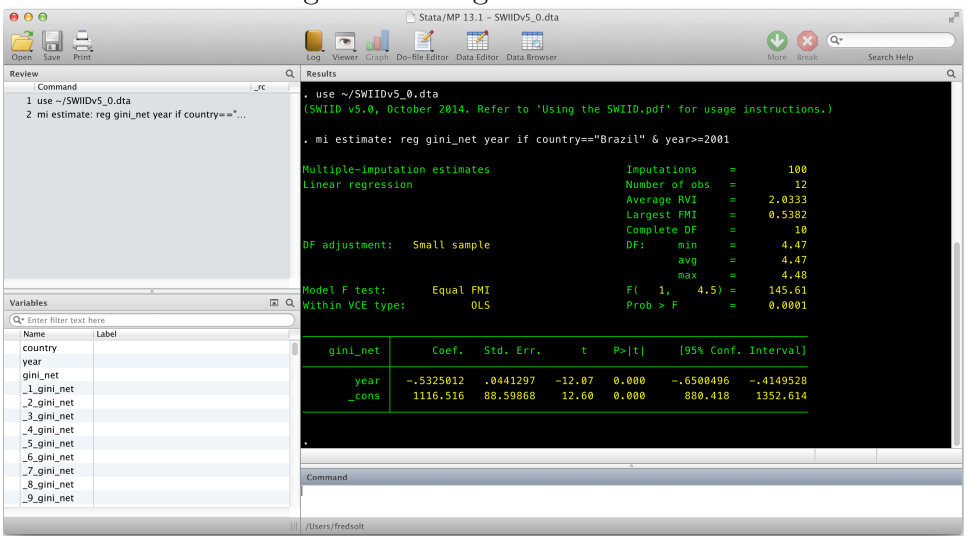
To use the SWIID estimates in an analysis in Stata, simply merge any other variables into the SWIID dataset (the SWIID should be the 'master' file in the merge, the other data should be the 'using' file), and then prefix commands with `mi estimate` and a colon. For example, if we were interested in whether income inequality in Brazil had declined since the introduction of national conditional cash transfers in 2001, we might wish to regress the SWIID estimate of the Gini index for net income for that country on time since that year. To find out, after loading the data, we need only precede our usual command

```
reg gini_net year if country=="Brazil" & year>=2001
```

with `mi estimate:` as shown in Figure 1.

The results indicate that, according to the SWIID estimates, income inequality has been dropping by an average of about a half (plus or minus about a tenth) of a Gini-index point per year in Brazil since 2001 and that

Figure 1: Using the SWIID in Stata



this pattern of decline is statistically significant. Subsetting and merging in other data require no special attention, but note that when generating new variables from the SWIID estimates, the `mi passive:` prefix should be used.

2 Using R

The SWIID estimates in the .RData file are formatted for ease of use with the R package `mitools`. The following sections describe how to subset the data, merge in additional variables, and do analyses.

2.1 Subsetting

Subsetting the SWIID requires a bit of care. Again suppose that we were interested in whether income inequality in Brazil had declined since 2001, and we decide to regress the SWIID estimate of the Gini index for net income for that country on time since that year. Ordinarily, we would probably use subscripting to restrict our analysis to this subset of the data. However, the SWIID estimates are formatted as a list of data frames making subsetting the data in advance a better approach. We do this on each of the 100 data frames in the list using `lapply`:

```
load("SWIIDv5_0.RData")
swiid_brazil <- lapply(swiid, function(x)
    x[x$country=="Brazil" & x$year>=2001,, drop=F])
```

The result is a new list of 100 data frames, each containing only the SWIID data for Brazil since 2001.

2.2 Merging

Merging additional data into the SWIID also needs to be done carefully. Suppose we wanted to include GDP per capita in our dataset on Brazil, drawing on information the Penn World Tables. After loading the PWT dataset, constructing the per capita figure, and extracting it and the country and year identifiers, we again use `lapply` to merge it into every data frame in the `swiid_brazil` list:

```
pwt <- read.dta("pwt80.dta")
pwt$rgdpepc <- pwt$rgdpe/pwt$pop
pwt1 <- pwt[c("country", "year", "rgdpepc")]


swiid_brazil1 <- lapply(swiid_brazil, function(x) merge(x, pwt1))
```

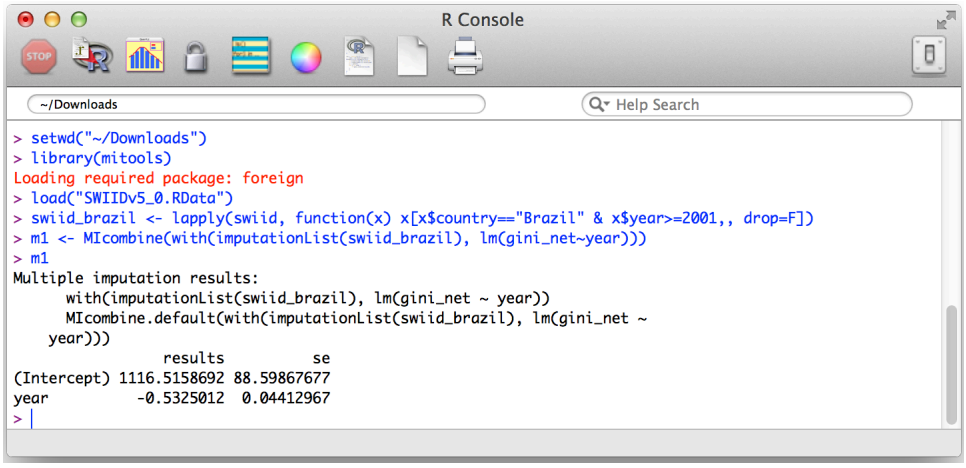
2.3 Analyzing

Once the desired subset has been extracted and any additional variables merged in, the `imputationList` and `MIcombine` commands from the `mitools` package allow us to conduct our analysis while taking the uncertainty in the SWIID estimates into account:

```
library(mitools)
m1 <- MIcombine(with(imputationList(swiid_brazil), lm(gini_net~year)))
```

Figure 2 shows the results. They indicate that, according to the SWIID estimates, income inequality has been dropping by an average of about a half (plus or minus about an tenth) of a Gini-index point per year in Brazil since 2001.

Figure 2: Using the SWIID in R



3 Citing the SWIID

Please cite to the SWIID by referring to its article of record and including the version number and date of release:

Solt, Frederick. 2014. "The Standardized World Income Inequality Database." Working paper. SWIID Version 5.0, October 2014.

References

Solt, Frederick. 2009. "Standardizing the World Income Inequality Database." *Social Science Quarterly* 90(2):231–242.

Solt, Frederick. 2014. "The Standardized World Income Inequality Database." Working paper. SWIID Version 5.0, October 2014.