

Data and data manipulations

Alwyn Young's data are posted at: <http://gsbwww.uchicago.edu/fac/alwyn.young/research/> (last accessed 1 Jan. 08).

I use AY's data as well as additional data.

1. Provincial specialization (output convergence)

a. China

The raw data are in three worksheets of the file
DRIE-output-pure-data-prov-88-02-2Jan08.xls

b. U.S.

The data can be obtained from <http://www.bea.gov/beanational/gsp> (last accessed 22 May 04).

The data manipulations/calculations are described in the appendix on U.S. data on cross-state specialization and price dispersion.

2. Provincial price dispersion

a. China

I am using AY's data, with some manipulations.

All retail price data, identical to AY's corresponding file, except for the removal, here, of products 121 and 122 (which he did not include in his calculations):

Retail_prices_annual_07woProd121and122.txt

The columns are, with coding for the first six columns by AY:

product

city

year (from 1986 through 1993)

price

count (number of cities for which retail prices for this product in this period are available)

state (dummy that assumes the value 1 if the price is the retail price of state-owned commerce rather than of the "whole society" [*quan shehui*])

dummy that assumes the value 0 throughout (included so that I would have the same number of columns as in the case of the monthly data)

Retail prices of 41 core products (for details see the appendix on price data):

Retail_prices_annual_07_41prod.txt

The columns are the same as above.

Monthly data (for details see the appendix on price data):

IndPricesEarly90s.txt

IndPricesLate90s.txt

IndPricesMid90s.txt

The columns are, with coding for the first six columns by AY:

product

city

month (from -38 through 72)

price

type (unclear, always takes the value 2)

count (number of cities for which prices for this product in this period are available)

dummy that assumes the value 1 in the first period when the reporting date switches from the 5th to the 15th of each month, and in the second period when the reporting date switches from the 15th to the 25th of each month; it assumes the value zero throughout the third period

Data manipulations: I took AY's data into .xls files, examined/ manipulated/ corrected data, then saved data into .txt files, retrieved these in matlab to calculate the various measures of dispersion (and to drop product-periods with less than 5 observations, to remove outliers, etc.), used stata transfer to transfer the matlab output of measures of dispersion and other information into .dta files, which I opened in stata where I then ran the regressions. I copied the regression output into .xls files where I created the charts.

Also see my explanations of the data in the appendix on price data.

b. U.S.

David Parsley and Shang-Jin Wei's (1996) data are available at <http://mba.vanderbilt.edu/david.parsley/Research.htm> (last accessed 2 Jan. 08)

I use the data on 26 non-perishable goods (file "Prices1.dat").

3. Variance decomposition

a. China

The raw data are in

China-variance-covariance-calculations-raw-data-3Jan08.xls

b. U.S.

The data can be obtained from

Output data (same as used above, in first section): <http://www.bea.gov/bea/regional/gsp/>
(accessed 22 May 2004)

Employment: <http://www.bea.doc.gov/bea/regional/spi/default.cfm> (accessed 16 Dec. 2005)

For details on some manipulations of the employment data see the corresponding figure in the paper.

4. Regression analysis

a. China

Data: Regression-data-8Jan08.xls

List of variables used in the individual regressions: Variables-used-in-regr-8Jan08.doc

b. U.S.

Data and their sources are explained in the appendix on regression analysis.

Total farm employment: <http://www.bea.doc.gov/bea/regional/spi/default.cfm> (accessed on 16 Dec. 2005 for 1977-2001 data by SIC industry)

Agricultural censuses: http://151.121.3.33:8080/Census/Create_Census-US.jsp (accessed 22 Dec. 2005)