Problem Set 1:
Consumption in the US

Nicola Fuchs-Schündeln

December 7, 2009

This problem set uses data from the CEX (datacex.dta on the course webpage). Many thanks to Dirk Krüger, who put these data together and made them available to me. Please hand in your program (send it by email), plus a short write-up of results, and graphs. This is a very nice data set to play around with and do some interesting preliminary analyses (e.g. you can get a feeling for the evolution of consumption inequality over time, analyze whether business cycle trends are visible in the consumption data, compare consumption and income volatities etc.). If you decide to do that, please hand in your additional results to me as well. In this problem set, I give some hints on useful commands to use in Stata, but of course there exist different possible ways to program things.

The data set contains the following information from the CEX from 1986 to 2000:

1. hhnr: Household identification number.
2. intnr: Interview number. Each household is interviewed five times, but the first interview is useless for our purposes (and hence not included in the data set). Thus, this variable takes values from 2 to 5, indicating whether the row corresponds to the second, the third, the forth or the fifth interview.
3. intmonth: Interview month. Indicates the calender month in which the interview took place.
4. intquart: Interview quarter. Indicates the quarter of the year in which the interview took place.
5. intyear: Interview year. Indicates the calendar year in which the interview took place.
6. hhmembers: Number of household members, including children.
7. age: Age of the household head, i.e. the person who did answer the questionnaire.

8. pretaxinc: Household income before taxes in the 12 months prior to the interview, deflated by the CPI.

9. aftertaxinc: Household income after taxes in the 12 months prior to the interview, deflated by the CPI.

10. foodexp: Food consumption of the household in the last three months prior to the interview, deflated by the price deflator for food.

11. nondurexp: Nondurable consumption expenditures in the last three months prior to the interview, where each component is deflated by its corresponding price deflator. It includes strictly nondurables expenditures as defined below, plus expenditures for apparel, health, education and reading.

12. strictnondurexp: Strictly nondurable consumption expenditures in the last three months prior to the interview, where each component is deflated by its corresponding price deflator. It includes expenditures for food, alcoholic beverages, tobacco, utilities, personal care, household operations, public transportation, gasoline and motor oil, and miscellaneous expenditures.

13. consumption: Nondurables plus imputed service flows from consumer durables. The precise definition and discussion of the imputation procedure is contained in the appendix of Krueger and Perri (2003).

14. totalexp: Total consumption expenditures in the last three months prior to the interview, deflated by the CPI.

1 Summary statistics and plots

1.1 Do the data make sense?

You should always "play around" with a new data set first, to see whether the data make sense. Write up shortly where you see potential problems with the data. Here are some suggestions:

- Are there any negative observations where they do not make sense?
- Should some variables always have larger values than others, and is this the case?
- Are there any outliers that simply seem unrealistic?

1.2 Consumption profile over the life cycle

Generate a figure that displays the shape of consumption over the life cycle. Following are three suggestions with increasing sophistication.

1. Assume that there are no cohort and year effects at all. Thus, simply take the average consumption for a given age, and plot consumption against age. *Hint: To take averages over age, I use the command "collapse" in Stata.*

2. Now test for the existence of cohort effects (e.g. it is reasonable to assume that due to macroeconomic growth younger cohorts are better off than older cohorts). Generate a variable for birth cohort, using year and age. Run a regression of consumption on cohort dummies and a cubic function in age: is there a trend in the cohort dummies? *Hint: you can use the command "xi" in Stata to do the regression on cohort dummies.*

3. Depict a cohort-age consumption profile. Pick 4 cohorts (e.g. born in 1930, 1940, 1950, and 1960), and depict their respective profile over the life cycle, all in one graph. Do the cohort effects become evident? *Hint: You can preserve the main data set, keep only the cohort you like, save the age-averages for this specific cohort, and later on merge the data for all 4 cohorts.*

1.3 Family profile over the life cycle

Do the same thing you did for consumption for the number of household members. Would we expect a cohort effect here? Do we see a cohort effect?

2 Perfect risk sharing test

In the presence of perfect risk sharing (e.g. through Arrow-Debreu securities), individual consumption growth should be independent of individual income growth (whether this income growth is predictable or not). Moreover, individual consumption growth should be perfectly correlated across individuals, since it is only influenced by aggregate shocks. We can test the perfect risk sharing hypothesis by running the following regression:

$$\Delta \ln \left(c_t^i\right) = \alpha_0 \Delta \ln \left(c_t^a\right) + \alpha_1 \Delta \ln \left(y_t^i\right) + \varepsilon_t^i$$

where i indicates an individual, a indicates the average across individuals, and t indicates time. $\ln \left(c_t^a\right)$ is defined as $\ln \left(c_t^a\right) = \frac{1}{N} \sum_i \ln c_t^i$. Perfect risk sharing implies $\alpha_0 = 1$ and $\alpha_1 = 0$. Test for perfect risk sharing in the data set.

1. For each household, generate variables for the quarterly change in log consumption and income.

2. Calculate the average of the logarithm of consumption at any point in time, and run the above regression.