

Heterogeneous Agent Models

Fall 2020

Project 2

Alexander Ludwig

Remember to hand in the solution electronically to
Irina Popova (Popova@econ.uni-frankfurt.de)
with me in cc.

In this problem set I ask you to work with **CEX data and to redo the analysis of Mace (1991)**. In order to do so, you need the statistical software **STATA**. I encourage you to extend Mace' analysis. But if you do so, you should always provide justifications for the regressions you run and for the graphs you plot. Notice that this is a project not an exercise. Hence, there is no correct answer. I would like you to hand in the **computer code** that is commented (so that I can understand it) and a **short summary of the main results, using tables, graphs or other tools and a brief interpretation**. You should present your work in such a way that it is easy to understand. The data can be found here:

<https://www.dropbox.com/s/6o3t2n0may24ifd/cexdata.zip?dl=0>.

The data are given from 1986 to 2000. The data set is organized as follows: each row is an observation for a household in a given quarter. Hence, since all households are interviewed at most 4 times before being rotated out of the sample, a given household contributes up to 4 rows to the data set. Each row is composed of 14 columns, corresponding to 14 variables being collected for the household in the particular quarter. The variables are, in the order in which they appear in the data set:

- **id**: Household identification number. This is a number with a maximum of six digits that identifies a household. The number stays with the household for the entire time it is in the sample. So if one wants to find household consumption in different quarters for the same household, one has to search for rows with the same entries in the first column, indicating the same household.
- **inumber**: 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.

- **month**: Interview month. Indicates the calendar month in which the interview took place.
- **quarter**: Interview quarter. Indicates the quarter of the year in which the interview took place.
- **year**: Interview year. Indicates the calendar year in which the interview took place.
- **hhsz**: Household size. Number of household members, including children.
- **age**: Age of the household head, i.e. the person who did answer the questionnaire.
- **grinc**: Household income before taxes in the 12 months prior to the interview, deflated by the CPI.
- **netinc**: Household income after taxes in the 12 months prior to the interview, deflated by the CPI.
- **food**: Food consumption of the household in the last three months prior to the interview, deflated by the price deflator for food.
- **ndcons1**: 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.
- **ndcons2**: 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.
- **ndconsserv**: Nondurables plus imputed service flows from consumer durables.
- **totcons**: Total consumption expenditures in the last three months prior to the interview, deflated by the CPI.

Thus, the total number of columns in each data set is 14, the total number of rows is 276,185. Note that I already selected out incomplete income respondents, as suggested by Nelson (1994), so that the data only contains complete income reporters.

- Load the data. Since the file is huge you have to increase memory in Stata. Do so by typing `set mem 100m`. Provide summary statistics (mean, variance, possibly other measures) for the different variables of interest (in particular income and consumption) by survey quarter. Can you detect the effects of the business cycle in these data? (you may want to look at some aggregate data first to get a sense about the macroeconomic conditions during the sample period).
- Test the perfect risk sharing hypothesis with this data set. Your analysis may be guided by Mace (1991). A useful source of information is also Cochrane (1991) and Nelson (1994)'s comment on the paper by Mace. Remember from the Mace paper that only the second and the fifth interview contain useful income information; this may guide your choice of which of the consumption data (there are (at most) four quarters of observations for each household) to use. Every output table should come with the regression equation to which it is the output to.

References

- Cochrane, J. (1991). A Simple Test of Consumption Insurance. *Journal of Political Economy* 99, 957–976.
- Mace, B. (1991). Full Insurance in the Presence of Aggregate Uncertainty. *Journal of Political Economy* 99, 928–956.
- Nelson, J. A. (1994). On Testing for Full Insurance Using {Consumer Expenditure Survey Data}. *Journal of Political Economy* 102, 384–394.