Backward Imputation of Financial Household Wealth

Raun van Ooijen, Rob Alessie

Discussion by

Hans Bloemen
(Vrije Universiteit Amsterdam, TI, Netspar)

1. Summary

- Imputation of financial household wealth
- Applied to IPO: specific properties
  - IPO Income: 2001-2010
    - Tax records
    - Taxable income in Box 3: wealth as a censored variable
  - IPO Wealth: 2005-2010
    - Information provided by banks
- Imputation: “corners”:
  - Box 3 taxable
  - Differences amounts reported by banks and to tax authorities

2. Remarks

- Motivation: what is the purpose?
  - Get a picture of the distribution of wealth in the years 2001-2004?
  - Development in wealth holdings for those years?
  - Apply to policy evaluation methods?
- Presentation of assumptions along the way

2. Remarks

- Random effects probit model for corners (section 3):

$$y_{it}^* = x_{it}'\beta + c_i + u_{it}, \quad i = 1, \ldots, N, t = 1, \ldots, T_i.$$

$$u_{it} | x_i, c_i \sim NID(0, 1).$$

$$y_{it} = \begin{cases} 1 & \text{if } y_{it}^* = 0 \\ 0 & \text{if } y_{it}^* \neq 0. \end{cases}$$

- "No serial correlation errors"
- Strict exogeneity assumed or required? Usually yes.
- Table F: No! But then: no consistent parameter estimates (section 4):
- Does that matter: no, only interested in projection!
- But why then using random effects from the start?

2. Remarks

F Probit modelling of dum_sav_t

Table 23: Right hand side variables used in probit regressions

Name of variable	Description of variable	Source
dum_sav_{t+1}	value of dum_sav in year $t+1$	IPO Wealth
$dum_{t=\alpha}$	time dummies $\alpha = \{2006, \ldots, 2009\}$	
$dum_{bankteg_t=0}$	dummy for $bankteg_t = 0$	IPO Income
$dum_{bankteg_{t+1}=0}$	dummy for $bankteg_{t+1} = 0$	IPO Income
$dum_{balsav_{t+1}=0}$	dummy for $balsav_{t+1} = 0$	IPO Wealth
$dum_{interest_t=0}$	dummy for $interest_t = 0$	IPO Income
$I_{interest_t>0} * \ln(interest_t)$	natural logarithm of interest income from checking and savings accounts in year t	IPO Income
$I_{bankteg_t>0} * \ln(bankteg_t)$	natural logarithm of $bankteg_t$ in year t	IPO Income
$I_{assets_t>0} * \ln(assets_t)$	natural logarithm of the average value of assets in year t	IPO Income
$I_{risky\_assets_t>0} * \ln(risky\_assets_t)$	natural logarithm of the value of stocks and bonds in year t	IPO Income
$I_{balsav_{t+1}>0} * \ln(balsav_{t+1})$	natural logarithm of $balsav_{t+1}$	IPO Wealth
$I_{balsav_{t+2}>0} * \ln(balsav_{t+2})$	natural logarithm of $balsav_{t+2}$	IPO Wealth
$I_{contrsav_t>0} * \ln(contrsav_t)$	natural logarithm of contractual savings ("spaarloon") in year t	IPO Income
cons	constant term	

2. Remarks

- Step 2: levels, fixed effects model is used:

$$s_{it} = z_{it}'\theta + \alpha_i + v_{it}, \quad i = 1, \ldots, N, t = 1, \ldots, T_i.$$

The model allows for correlation between α_i and $z_i = (z_{i1}', \ldots, z_{iT_i}')'$,

- Fixed effects estimation
- So step 1: we do not care, in step 2, we do(?)
- FE: only Within Group variation. Could Between Group variation add to predictive value?
- By the way FE not consistent due to dynamic nature (see GMM literature, Arellano Bond, etc)

2. Remarks

- Two part model: where does the sharp cut-off come from? Two separate samples
- In a simulation, we would not expect that.
- Cut-off may introduce additional error
- Is two part approach in the aggregate better than one part?

2. Remarks

- General: tables F and G deserve a much more prominent place in the paper
- They show which variables are used for the prediction: missing in section 3, which shows textbook formulations of models
- Underlying motivation? Does economic theory matters? Life cycles? Measurement error literature? IV?

2. Remarks

- Out of sample prediction of the fixed effect?
- GLS step with autocorrelated errors?
- Statistical properties of the imputed variables? (bias, consistency, MSE).

2. Remarks

- Year 2005: easy to compare (both IPO income and wealth)
- Years 2001-2004: we see similarities and differences.

- But: Spearman's ρ of 0.9579!