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

### Table 23: Right hand side variables used in probit regressions

<table>
<thead>
<tr>
<th>Name of variable</th>
<th>Description of variable</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>dum_sav_t+1</td>
<td>value of dum_sav in year (t+1)</td>
<td>IPO Wealth</td>
</tr>
<tr>
<td>dum_t=α</td>
<td>time dummies (α = {2006, \ldots, 2009})</td>
<td>IPO Income</td>
</tr>
<tr>
<td>dum_bankteg=0</td>
<td>dummy for bankteg = 0</td>
<td>IPO Income</td>
</tr>
<tr>
<td>dum_bankteg+1=0</td>
<td>dummy for bankteg(_t+1) = 0</td>
<td>IPO Income</td>
</tr>
<tr>
<td>dum_balsav_t+1=0</td>
<td>dummy for balsav(_t+1) = 0</td>
<td>IPO Wealth</td>
</tr>
<tr>
<td>dum_interest_t=0</td>
<td>dummy for interest(_t) = 0</td>
<td>IPO Income</td>
</tr>
<tr>
<td>(_{interest_t}&gt;0 \times \ln(\text{interest}_t))</td>
<td>natural logarithm of interest income from checking and savings accounts in year (t)</td>
<td>IPO Income</td>
</tr>
<tr>
<td>(_{bankteg_t}&gt;0 \times \ln(\text{bankteg}_t))</td>
<td>natural logarithm of bankteg(_t) in year (t)</td>
<td>IPO Income</td>
</tr>
<tr>
<td>(_{assets_t}&gt;0 \times \ln(\text{assets}_t))</td>
<td>natural logarithm of the average value of assets in year (t)</td>
<td>IPO Income</td>
</tr>
<tr>
<td>(_{risky_assets_t}&gt;0 \times \ln(\text{risky_assets}_t))</td>
<td>natural logarithm of the value of stocks and bonds in year (t)</td>
<td>IPO Income</td>
</tr>
<tr>
<td>(<em>{balsav_t+1}&gt;0 \times \ln(\text{balsav}</em>{t+1}))</td>
<td>natural logarithm of balsav(_{t+1})</td>
<td>IPO Wealth</td>
</tr>
<tr>
<td>(<em>{balsav_t+2}&gt;0 \times \ln(\text{balsav}</em>{t+2}))</td>
<td>natural logarithm of balsav(_{t+2})</td>
<td>IPO Wealth</td>
</tr>
<tr>
<td>(_{contrsav_t}&gt;0 \times \ln(\text{contrsav}_t))</td>
<td>natural logarithm of contractual savings (&quot;spaarloon&quot;) in year (t)</td>
<td>IPO Income</td>
</tr>
<tr>
<td>cons</td>
<td>constant term</td>
<td>IPO Income</td>
</tr>
</tbody>
</table>
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 \( \alpha_i \) 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

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 $\rho$ of 0.9579!