Nowcasting the household income distribution

February 18th 2016
Fabrice Murtin
OECD Statistics Directorate
“Big picture”

- STD initiative on timeliness: enhance Secretariats’ capacities to make more timely statements in a range of fields
- Drive: 2012 Internal Audit on Statistical Activities
- The initiative includes three pilots:
  - Trade in value added (TiVA)
  - Income Distribution (2 projects)
  - Subjective Well-Being (through Big data)
Background and motivations

- Since 1998, regular OECD data collection on income distribution and poverty (OECD IDD) based on national sources and comparable definitions.

- Strong internal and external demand for IDD (e.g. COPE & its reports, How’s Life?, Inclusive Growth, Economic Surveys, G20, etc.)

- However, despite annual collection, information is not timely: average lag is 2-3 years...

- This limits the possibility to use distributive information in macro-economic assessments where timeliness is key requirement (e.g. Economic Outlook, Going for Growth)
Background and motivations

- Project seeks to *nowcast* household income by decile (in year $T$) in as many OECD countries as possible based on contemporaneous information through reduced-form econometrics.

- Once methodology has been thoroughly tested, estimates could be released regularly by the OECD in various forms (NAD household dashboard, MDLS, *How’s Life?*, *ad hoc* statistical briefs, G20 documents).

- In the very short term: working paper and feedback from experts in a variety of fora.
Nowcasting: basic principles

- **Objective**: Construct a predictive model that can be evaluated by out-of-sample (OOS) performance
- **Parsimony**: a complicated model increases in-sample fit (R2) but may decrease OOS
- **Credibility**: meaningful coefficients
- **Specificity**: the model must be decile-specific and possibly country-specific
The dependent variable

- Average equivalised household disposable income per decile from IDD
- We consider two income series per country (waves 6 and 7)
- Linear interpolations used to cover gaps up to 3 years for countries lacking annual surveys and for earlier (pre-2000s) periods
- We also considered and tested a model to predict individual income’s components for each decile (i.e. wage, self-employment income, capital income, transfers received, taxes paid) but model performed less well than model for total income
Explanatory variables

- We created a group of 23 predictors, drawn from national accounts and other official sources, that are timely and available for most countries

- Examples: GDP, unemployment rate, mean net household disposable income (SNA), self-employment rate, wage rate, hours worked per worker, long-term interest rates, house prices, property income, share prices, current transfer received by households, taxes on business and on different kinds of households etc...
The predictive model

- For each decile we aim to predict the growth rate of real household disposable income (defl. CPI):
  \[ \Delta_{t,t-1} \log y = \Delta_{t,t-1} x \beta_1 + \Delta_{t-k,t-k-1} \log y \beta_2 + \varepsilon \]
  through a single panel regression

- **Variable selection**: we tested LASSO/LARS to identify the ‘best’ predictors, then proceeded with manual adjustments (significance and interpretability of all coefficients)

- **Performance**: we evaluated 1 year-ahead out-of-sample performance against observed growth rates and a naive ‘random walk’ model (forecasted growth=last observation)
### Estimated model

- All coefficients have the same sign across all income deciles (and all variables are ‘correctly’ signed)
- But not all variables are relevant for all deciles

<table>
<thead>
<tr>
<th>Control for lags</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>D6</th>
<th>D7</th>
<th>D8</th>
<th>D9</th>
<th>D10</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Change in/growth of:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment rate</td>
</tr>
<tr>
<td>GDP</td>
</tr>
<tr>
<td>Wage rate</td>
</tr>
<tr>
<td>Average tax wedge</td>
</tr>
<tr>
<td>Disbursement of government</td>
</tr>
<tr>
<td>Current transfers</td>
</tr>
<tr>
<td>Net current household receipts</td>
</tr>
<tr>
<td>Self-employment rate</td>
</tr>
<tr>
<td>Share price</td>
</tr>
<tr>
<td>Disposable income</td>
</tr>
</tbody>
</table>
Estimated model

- Average OOS correlation (across all deciles) is 0.72
- Tails are harder to capture with a linear model but non-linear models are quite unstable
The estimated model easily beats the random walk. But D1 and D10 are the hardest to predict.
The model ‘predicts’ some countries much better than others. When focusing on the more ‘stable’ countries (13 out of 22), prediction error falls by ~50% (of growth rate) in absolute terms.
Prediction errors are quite different across countries: remaining heterogeneity
Other issue: change in inequality is not well captured
Conclusions

- Nowcasting real changes in household income for various deciles is difficult because:
  - A complicated model is unstable
  - A simplistic model is inaccurate

- More research is needed to:
  - Better model the tails of the distribution
  - Better capture ‘regime changes’ (large deviations)
  - Better account for country heterogeneity

- On country heterogeneity:
  - other methodologies (microsimulations) may outperform regression-models but are difficult to implement in a consistent way and are much more demanding in terms of information
  - predicting the distribution from NA totals
Thank you!