Median Labor Income in Chile Revised: Insights from Distributional National Accounts

Manuel Taboada ¹ José De Gregorio ²

¹Central Bank of Chile

²Universidad de Chile

¹Our views do not necessarily represent the Central Bank of Chile or its Board.
# Table of Contents

1. Motivation

2. Data and Methods

3. Results and Final Comments

4. Appendix
Motivation

The Problem

- What is the median income in Chile? Survey data used to represent national income has significant inconsistencies.
- Total survey income (CASEN) accounts for just half of GNI.

Table 1: National Income per worker and survey median income, 2017

<table>
<thead>
<tr>
<th>National Income per worker</th>
<th>Median income (surveys)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,350,000 CLP</td>
<td>Income concept, inequality, <strong>bias</strong>? 400,000 CLP</td>
</tr>
</tbody>
</table>

Source: Own preparation based on INE and CBCh. National Income excludes indirect taxes and depreciation.
Use national accounts, household surveys, and administrative records to provide consistent distributional series for 2006-2020, emphasizing labor income.

Estimated median gross income for dependent workers is $600,000 (32% gap) and $570,000 (39% gap) for all workers in 2017.

About a quarter of the gap is attributed to the “missing rich” in the survey.

Median dependent income grew to $670,000 in 2020 (12%) but for 12% less workers.

Corrected mean-median income ratio (Gini) is 26% (17%) larger than in the raw survey of 2017, and falls only 6% (3%) between 2006 and 2017 compared with a larger decline of 12% (11%) in the original data.
Motivation
Combining Sources

- The CASEN Survey is the best available source for low and middle income but it is biased, specially at the top.
- Tax records are better for top income but not reliable for people under the taxable threshold ($\approx p70$).
- Real inequality is higher than portrayed in CASEN, but average and median income as well.
- Our methodology corrects the known limitations of different data sources and combine them coherently for the first time using Chilean data.
The CASEN survey adjusted five income components to national accounts. The validity of those adjustments was questioned as it dubiously reduced poverty and, increased inequality (Bravo and Valderrama, 2011)

Lustig (2019) analyzes the *missing rich* in household surveys, its magnitude, causes and correction methods

Alvaredo et al, 2020 develop a methodology to generate consistent and comparable inequality measurement using National Account concepts

De Rosa, Flores and Morgan (2022) use DINA to measure inequality in Latin America, consequently providing less detail for Chile
Methodology

Step 1 and 2

- **Step 1:** With tabulated tax data generate a synthetic distribution for net taxable income using generalized Pareto interpolation (Blanchet, Fournier and Piketty, 2017).

- **Step 2:** Construct the same income definition from CASEN, adding labor income (dependent and independent), pensions and capital (declared and imputed)

- Add a fraction, increasing in income, to impute capital income starting at the 80th percentile
Step 3: *missing-rich* Correction (BFM)

- Using the common income definition, the survey representativeness is corrected with tax data following Blanchet, Flores and Morgan (2018).
- Survey weights are calibrated to match the density of tax data at the top.
- $95^{th}$ percentile as baseline merging point, in line with Ruiz and Woloszko (2016).
- The assumption is that bias over the $95^{th}$ percentile is greater than in the rest of the curve.
Step 4 and 5: Tax imputation and NA adjustment

- **Step 4:** Detailed tax and social security imputations, simulating evasion with survey questions

- **Step 5:** Survey aggregates are scaled to match National Account totals including partial evasion of social security

- The adjustment factor for the independent (around 1.5) is more reasonable than 3.8, obtained by ECLAC

- Also estimate results omitting **step 3** to quantify the effect of the *missing rich*
## Results
### Adjustment Factors

Table 2: National Accounts adjustment factors, with and without BFM correction

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wages</td>
<td>1.26</td>
<td>1.43</td>
<td>1.43</td>
<td>1.43</td>
<td>1.44</td>
<td>1.41</td>
<td>1.38</td>
</tr>
<tr>
<td>Wages BFM</td>
<td>1.22</td>
<td>1.34</td>
<td>1.38</td>
<td>1.34</td>
<td>1.30</td>
<td>1.29</td>
<td>1.21</td>
</tr>
<tr>
<td>Self-employed income</td>
<td>1.46</td>
<td>1.35</td>
<td>1.42</td>
<td>1.85</td>
<td>1.82</td>
<td>1.63</td>
<td>1.37</td>
</tr>
<tr>
<td>Self-employed income BFM</td>
<td>1.49</td>
<td>1.33</td>
<td>1.57</td>
<td>1.76</td>
<td>1.64</td>
<td>1.46</td>
<td>1.18</td>
</tr>
</tbody>
</table>

Source: Own elaboration based on data from CASEN, the Central Bank of Chile, and SII.

- Gini coef. (2017) starts at **0.49** for net taxable income in CASEN, then **0.53** with imputed capital, **0.56** with *missing-rich* and **0.6** for the NA corrected gross income.
Real median wage grew by 4.5% annually until 2017, which is 2% more than both the income of the 90th percentile and GDP per capita.

The increase in 2020 masks a significant and heterogeneous decrease in employment of 12%.
Corrected Gini is 17% larger in 2017 and, decreased only 3% since 2006 (v. 11%). Partial reversal due to COVID-19.
### Table 3: Sensibility Analysis for Gross Income, 2017

<table>
<thead>
<tr>
<th></th>
<th>Dependents p50</th>
<th>Dependents p90</th>
<th>Independents p50</th>
<th>Independents p90</th>
<th>All Workers p50</th>
<th>All Workers p90</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CASEN¹</strong></td>
<td>450,345</td>
<td>1,302,293</td>
<td>300,000</td>
<td>1,007,393</td>
<td>408,384</td>
<td>1,250,000</td>
</tr>
<tr>
<td><strong>p = 0.9</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k = 0.5</td>
<td>554,309</td>
<td>2,115,156</td>
<td>391,217</td>
<td>1,956,086</td>
<td>545,222</td>
<td>2,115,156</td>
</tr>
<tr>
<td>k = 0.9</td>
<td>576,258</td>
<td>1,998,850</td>
<td>424,454</td>
<td>1,899,253</td>
<td>565,938</td>
<td>2,013,136</td>
</tr>
<tr>
<td>k = 1.2</td>
<td>593,877</td>
<td>1,986,959</td>
<td>454,087</td>
<td>1,793,945</td>
<td>591,412</td>
<td>1,986,959</td>
</tr>
<tr>
<td><strong>p = 0.95</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k = 0.5</td>
<td>572,958</td>
<td>2,067,182</td>
<td>403,846</td>
<td>1,807,754</td>
<td>538,699</td>
<td>2,070,588</td>
</tr>
<tr>
<td>k = 0.9</td>
<td><strong>598,330</strong></td>
<td><strong>2,004,139</strong></td>
<td><strong>436,940</strong></td>
<td><strong>1,747,760</strong></td>
<td><strong>568,414</strong></td>
<td><strong>2,004,139</strong></td>
</tr>
<tr>
<td>k = 1.2</td>
<td>622,183</td>
<td>1,934,057</td>
<td>469,971</td>
<td>1,753,167</td>
<td>583,296</td>
<td>1,925,604</td>
</tr>
<tr>
<td><strong>p = 0.99</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k = 0.5</td>
<td>609,872</td>
<td>1,830,973</td>
<td>442,080</td>
<td>1,649,148</td>
<td>568,853</td>
<td>1,801,112</td>
</tr>
<tr>
<td>k = 0.9</td>
<td>620,223</td>
<td>1,827,207</td>
<td>466,094</td>
<td>1,737,037</td>
<td>580,633</td>
<td>1,808,245</td>
</tr>
<tr>
<td>k = 1.2</td>
<td>626,638</td>
<td>1,835,134</td>
<td>486,573</td>
<td>1,724,638</td>
<td>587,760</td>
<td>1,815,095</td>
</tr>
<tr>
<td><strong>Without BFM²</strong></td>
<td>635,098</td>
<td>1,836,554</td>
<td>490,155</td>
<td>1,645,928</td>
<td>596,513</td>
<td>1,828,436</td>
</tr>
</tbody>
</table>

Source: Own preparation based on CASEN, SII and CBCh.

- Median income is monotonically increasing in $p$ and $k$, while the 90th percentile is decreasing but not regularly
Median labor income is strongly underestimated. In 2017, median gross dependent income is corrected from $450,000 to $635,000 (+41%) or to $600,000 (+32%) incorporating the *missing rich*.

Even in simulations with extreme levels of inequality (Gini = 0.65 share 1% = 27%) median wage is $500,000 (+23%)

The deficiencies that prevent income surveys capturing higher income are well recognized, but it is also necessary to incorporate the conclusions of this study regarding middle income earners into public discussion and policy design
Current Research at CBCh

- This research served as a base for complete Distributional National Accounts at the Central Bank (in process)
- Working to incorporate accurate measurement of inequality and distribution into the macro analysis
- Unusually rich administrative data on wages, taxes, pensions, profits, debt, etc. enables to correct them with CASEN (and others) instead of the reverse
- Collaboration with the Statistics Division in order to generate the missing National Accounts needed (Net National Income, retained profits, mixed income, etc.)
Current Research
New methods

- Match 1:1 administrative records with CASEN at the individual level using Optimal Transport (Blanchet et al. 2022) bringing demographic characteristics and informal income into tax data
- Modified OT preserves the formal structure of the HH, spouses and children, from the civil registry and adds “statistical clones” to achieve the real HH size distribution from the survey
- Also, precisely allocate undistributed profits using firm-to-firm and firm-to-individual ownership reaching final (indirect) owners
- Use the Leontief Inverse Matrix to allocate profits from the universe of firms ($\approx 200k$) to the population

THANK YOU!
Appendix
Step 1 Partial Results

Figure 3: Generalized Pareto Coefficients \( b(p) \) from Net Taxable Income

- \( b(p) \) corresponds to the ratio between the average income of the incomes that are above the \( p \) percentile and the income corresponding to that percentile.

- Net income results in smaller coefficients and larger \( p \) for \( b'(p) > 0 \)
Step 3
Partial Results

Figure 4: Original and BFM-Corrected Lorenz Curve for Taxable Income in CASEN 2017

- Re-weighting at the top moves the whole curve to the right
- The effect on the median and low incomes is smaller
A key parameter is the proportion of the distributed income of corporations (D42) attributed to cuasicorporations (D421). ECLAC used a proportion of 90.7% resulting in an adjustment factor of 3.79.

Table 4: Independent Income Adjustment Factors for Different Fractions of D.42

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>1.15</td>
<td>1.00</td>
<td>1.01</td>
<td>1.33</td>
<td>1.30</td>
<td>1.19</td>
<td>1.08</td>
</tr>
<tr>
<td>20%</td>
<td>1.46</td>
<td>1.34</td>
<td>1.42</td>
<td>1.85</td>
<td>1.82</td>
<td>1.63</td>
<td>1.36</td>
</tr>
<tr>
<td>40%</td>
<td>1.77</td>
<td>1.69</td>
<td>1.83</td>
<td>2.38</td>
<td>2.34</td>
<td>2.08</td>
<td>1.65</td>
</tr>
<tr>
<td>60%</td>
<td>2.08</td>
<td>2.03</td>
<td>2.24</td>
<td>2.91</td>
<td>2.86</td>
<td>2.52</td>
<td>1.94</td>
</tr>
<tr>
<td>80%</td>
<td>2.39</td>
<td>2.38</td>
<td>2.65</td>
<td>3.44</td>
<td>3.38</td>
<td>2.97</td>
<td>2.22</td>
</tr>
<tr>
<td>100%</td>
<td>2.70</td>
<td>2.72</td>
<td>3.06</td>
<td>3.96</td>
<td>3.90</td>
<td>3.42</td>
<td>2.51</td>
</tr>
</tbody>
</table>