# Data Appendix

This note provides sources and explanations for the data used in the main study, as well as detailed instructions for replicating the results. All the data used are publicly available. In this way it is possible to replicate the results from scratch.

It is noted that there are small differences, up to 1%, in the data used and the currently available source. This is due to corrections and updates in the data by the institution that prepares and delivers them.

**National Accounts and other records**

From the central bank, the integrated economic table available at the date of publication at <https://si3.bcentral.cl/estadisticas/Principal1/enlaces/Informes/AnuariosCCNN/anuario_CCNN_2019.html> is used.

From there we obtain: Compensation of employees (D.1), Gross operating surplus/Gross mixed income (B.2b-B.3b), Distributed corporate income (D.42) and Interest (D.41).

<https://si3.bcentral.cl/estadisticas/Principal1/enlaces/Informes/AnuariosCCNN/CCNN-anuarios.html>

Gross operating surplus for the housing services sector was derived from the 12x12 tables and the 111x176 tables in 2017 given the change in the criteria for cat

**OECD**

OECD tax database social security contributions in CLP for each year are used. The series used are 2100: Social Security Contributions of Employees and 2200: Social Security Contributions of Employers.

Both correspond to the 7% health discount, in the first case for dependent workers and in the second for self-employed workers.

It is assumed that the rest of the contributions (AFP, unemployment insurance and others) are made proportionally to the 7% for health, according to the corresponding rate. The average discount rate is assumed for rates that vary, such as the AFP service fee.

<https://stats.oecd.org/Index.aspx?DataSetCode=REV>

**SII data**

This section describes how to obtain the synthetic distribution by fractiles necessary to correct the CASEN survey for the missing-rich.

First, the raw tabulated data, publicly available at <http://www.sii.cl/sobre_el_sii/estadisticas_de_personas_naturales.html> , is downloaded. The required file is "Number of taxpayers and amounts of global complementary and single second category tax 2005-2019". The file is renamed "Net income tabulated SII 2006-2017". On this data, the net income of each bracket is constructed. The sheet "Consolidated 2005-2018" presents the values required as input for the "gpinter" adjustment. The same data are summarized in the sheets named according to each year. These sheets are copied to the "gpinter 2006-2017" file that will be used to generate the synthetic distribution.

Then, on the page https://wid.world/gpinter/ the methodology proposed by Blanchet, Fournier and Piketty (2017) is implemented following the instructions given on the website. The "gpinter results WID" file is obtained as a result. Each sheet should be separated into a separate file to be used with STATA as described in the next section.

**CASEN**

The bases obtained in the previous step are named "gpinter [year]". They are used in the correction proposed by Blanchet, Flores and Morgan (2018) implemented with STATA's bfmcorr command.

Each do-file named "Adjustment income [year].do" applies the described methodology and saves the results in the Excel file "resultados\_casen\_stata".

To download the databases of the CASEN surveys used, access http://observatorio.ministeriodesarrollosocial.gob.cl/casen/basededatos\_historico.php by retrieving the following files:

- "Database [year]", for 2006, 2009 and 2011.

- "Base de datos [año] complementaria de ingresos y subsidios ORIGINALES, desagregados" for 2006, 2009 and 2011.

They are combined using the file "[year]merge.do".

The 2013, 2015 and 2017 bases are available at http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/basedatos.php by retrieving the files:

- "Casen database [year], for 2013, 2015 and 2017.

With the merged bases for 2006, 2009 and 2011 and the full 2013, 2015 and 2017 bases, the final income correction is applied using the file "Income adjustment [year].do".

The monthly values of the comparable aggregates with which the adjustment is made, must be copied manually to the do-file "Income adjustment [year].do" from the Excel "Aggregates for adjustment and calibration". It is also possible to edit the parameters "p\_bfm" and "k\_i".

It is also possible to modify the tax rates and/or brackets to simulate the incidence of different tax structures.