Median labor income in Chile revised: Insights from Distributional National Accounts*

Ingreso mediano en Chile revisado: Un análisis con Cuentas Nacionales Distributivas

José de gregorio **
Manuel taboada ***

Abstract

A commonly used figure to highlight inequality in Chile is the median income of the Chilean socioeconomic household survey (known by its acronym in Spanish, CASEN). According to this survey, in 2017 the median monthly income per worker was CLP (Chilean pesos) 400,718 pesos, which compares to average income per worker from National Accounts of CLP 1,350,000 in the same year. For this difference to be correct, the implied Gini coefficient would be 0.7, which much above the Gini implied by the same survey. However, surveys, such as CASEN, often underreport income, particularly for middle- and high-income earners, leading to an underestimation of the median income. This study compares various data sources, including national accounts, household surveys, and administrative records, to create a more accurate picture of income distribution and median income. The corrected data shows higher median incomes and greater inequality than previously reported. On average, the underestimation of gross wages in the Chilean national household survey as compared to national accounts is 40%, significantly larger than other countries. About a quarter of this gap is attributed to the "missing rich" in the survey. For 2017, this equates to an estimated median gross income for dependent labor of CLP 600,000 and CLP 570,000 for all workers. The 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.

Key words: *Income inequality; median income; national accounts; income surveys; Distributional National Accounts.*

JEL Classification: *D31, D33, E01.*

---

* We are grateful to Rómulo Chumacero, Jeanne Lafortune, Andrea Repetto, Jaime Ruiz-Tagle, Rodrigo Valdés, and seminar participants at the Central Bank of Chile and SECHI for their valuable comments. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Central Bank of Chile or its Board.
** Universidad de Chile. Email jdegregorio@fen.uchile.cl
*** Central Bank of Chile. E-mail: mtaboada@bcentral.cl

Received: August, 2023 Accepted: August, 2024

Resumen

Una cifra comúnmente utilizada para resaltar la desigualdad en Chile es el ingreso mediano según la Encuesta de Caracterización Socioeconómica Nacional (CASEN). Según esta encuesta, en 2017 el ingreso mediano mensual por trabajador era de 400.718 pesos, en comparación con el ingreso promedio por trabajador de las cuentas nacionales, que era de 1.350.000 pesos en el mismo año. Para que esta diferencia sea correcta, el coeficiente de Gini implícito sería 0.7, mucho más alto que el Gini que sugiere la misma encuesta. Sin embargo, las encuestas como CASEN a menudo subestiman los ingresos, particularmente para los hogares de ingresos medios y altos, lo que conduce a una subestimación del ingreso mediano. Este estudio compara varias fuentes de datos, incluidas las cuentas nacionales, encuestas de hogares y registros administrativos, para crear una imagen más precisa de la distribución del ingreso y del ingreso mediano. Los datos corregidos muestran ingresos medianos más altos y una mayor desigualdad de la reportada previamente. En promedio, la subestimación de los salarios brutos en la encuesta nacional de hogares de Chile en comparación con las cuentas nacionales es del 40%, significativamente mayor que en otros países. Aproximadamente una cuarta parte de esta brecha se atribuye a los missing-rich en la encuesta. Para 2017, esto equivale a un ingreso bruto mediano estimado de 600.000 pesos para el trabajo dependiente y 570,000 pesos para todos los trabajadores. La proporción corregida de ingreso promedio a ingreso mediano (Gini) es un 26% (17%) mayor que en la encuesta original de 2017, y solo disminuye un 6% (3%) entre 2006 y 2017 en comparación con una disminución mayor del 12% (11%) en los datos originales.

Palabras clave: *Desigualdad de ingresos; ingreso mediano; cuentas nacionales; encuestas de ingreso; Cuentas Nacionales Distributivas.*

Clasificación JEL: *D31, D33, E01.*

> *"While we often must focus on aggregates for macroeconomic policy, it is impossible to think coherently about national well-being while ignoring inequality and poverty, neither of which is visible in aggregate data. Indeed, and except in exceptional cases, macroeconomic aggregates themselves depend on distribution."*
>
> *Angus Deaton. Nobel Prize Lecture, 2016*

1. Introduction

In Chile, income distribution features prominently in academic research, public policy, and the media. In addition to a few recognized inequality indicators, one statistic routinely cited in discussions about the well-being of the population is the median income of workers or households. The median captures a dimension on inequality when compared to the mean, but it also reflects aggregate income. However, survey data used to represent national income has significant inconsistencies that are seldom acknowledged but are revealed by the huge gap between labor income reported in national accounts and labor income effectively reported surveys. For example, in 2017, the Chilean socioeconomic household survey (known by its acronym in Spanish, CASEN)[1] reported a net average monthly employee compensation of 570,000 Chilean pesos (CLP), whereas the gross figure derived from the national accounts for the same year was CLP 955,000.[2] Even when taking taxes and social security into account, this represents a difference of over 40%. The disparity between these two statistics, and the fact that the survey captures only half of national income, raises doubts about the median income of workers, an almost omnipresent statistic, estimated at CLP 400,000 in the CASEN and other surveys (INE, 2019). Policymaking needs to be grounded in accurate data. For example, the distribution of labor income is important in the determination of minimum wages and social transfers. But, at the same time, the limitations of the data must be made explicit to avoid misinterpretation. Furthermore, despite the inconsistencies that arise in the calculation of inequality and income levels, it is widely recognized that Chile's average income has increased significantly since the return of democracy in 1990 and that poverty and inequality have been reduced (UNDP, 2017).

In this context, any conclusions about the substantial improvements in the standard of living and inequality levels in Chile since 1990 require a reappraisal given the various shortcomings of the data sources on which they are based. Notwithstanding numerous efforts to measure poverty and income share of the super-rich adequately, the measurement of inequality and income levels in the middle of the distribution are subject to two noteworthy limitations. The first limitation is the large gap between figures presented in national accounts, which are focused on macroeconomic aggregates, and those presented in poverty and inequality studies, which use data from surveys or administrative records, and which often present figures that are not consistent with

---

1 CASEN is the most important income and socioeconomic characterization survey available for Chile.

2 As a reference in 2017 the exchange rate was 650 pesos (CLP) per dollar. After the social unrest and later the pandemic the CLP weakened to 760 CLP per dollar in 2021.

macro-aggregates. The second limitation is that socioeconomic surveys tend to underestimate the income of the richest households to a greater extent than the income of middle- and lower-income earners, unlike tax records. Although underestimation is higher for top-income, middle-income earners also underreport in surveys for fear of additional taxation or mere bad accounting. Preliminary evidence suggests these deficiencies are especially acute in Chile.

In order to address these measurement issues, this paper compares and combines different data sources—national accounts, household surveys and administrative records (tax and social security)—to generate a consistent series of income distribution in Chile, focusing on the middle of the distribution. In this way, the known deficiencies of each data source can be corrected, and the advantages of each source can be exploited to achieve a coherent and unified empirical framework that allows the definitions and phenomena captured in micro and macroeconomic data to be effectively reconciled. In our central scenario, using 2017 CASEN corrected survey data, the result is a median gross income for all active workers of CLP 570,000, CLP 600,000 for dependent workers, and CLP 440,000 for independent (self-employed) workers. This equates to an increase of at least 40% compared to the figures reported in the original 2017 CASEN survey data. As a consequence of the corrections, inequality rises from a Gini of 0.52 and a mean-median-ratio of 1.65, to 0.6 and 2.09, respectively. These income gaps are very unusual in developed countries, especially for wages where Zwijnenburg et al. (2017) estimates an average gap around only 10% for a subset of OECD countries.

This paper proposes three innovations with respect to the existing income distribution literature. First, and for the first time using Chilean data, national accounts, household surveys, administrative records (tax and social security) are combined to analyze the complete distribution of income, and not only the income share of the super-rich as in Fairfield and Jorratt (2016)[3]. Second, this is the first study focused on the middle of the income distribution range that incorporates other data sources in addition to income surveys in order to examine the aforementioned problems with the microeconomic data. Third, the research covers the period between 2006 and 2020 in order to have medium-term trends, although 2020 should be treated with caution considering the uncertainty introduced by the social unrest in 2019 and the COVID-19 shocks. The results show the average gap between statistics presented in income surveys and comparable figures in national accounts is up to 45% for dependent labor, with larger gaps for capital and mixed income. Careful tax and social security simulations provide more plausible results than those presented in previous

[3] At the time writing, De Rosa et al. (2020) have a work in preparation that uses a similar methodology but the analysis is at a Latin-American level, providing less detail on Chile specifically.

studies. Our methodology serves as a basis for other developing countries with similar data quality and availability issues.

Controversy surrounding the magnitude of inequality in a country is usually generated because of conceptual and methodological differences between studies that give rise to contradictory results. There are numerous examples. Chile's Gini coefficient, as reported by the World Bank, suggests that inequality decreased between 1990 and 2015, falling from 0.57 to 0.45. And between 1990 and 2013, the CASEN survey shows that the real income of households in the 10th percentile increased by 4 times but that of the 90th percentile only increased by 2.8 times (Larrañaga and Rodríguez, 2014).[4] Some studies, however, such as Atria et al. (2018), contradict the evidence from income surveys and conclude that inequality increased after 2000. And in Fairfield and Jorratt (2016), the publication with access to the most detailed information, it is estimated that income concentration remained constant between 2003 and 2012. The literature also provides estimates of the income share of the richest 1% in a strikingly wide range, between 8.7% and 33%. The lower bound is calculated from raw total net household income from CASEN 2015, and the upper bound is obtained from López et al. (2013) and includes attributed capital gains. These discrepancies are caused in part by differences between the sample and the definitions of income used. They are, however, also a consequence of the difficulties associated with obtaining accurate information and designing reliable methodologies.

Given the inconsistency in results when data from national accounts, household surveys, and administrative records are used separately, it is very difficult to consistently estimate how economic growth has been distributed to the population in Chile. This inconsistency between sources is universal to a certain extent but mainly affects countries with lower income or statistical development. Numerous international efforts have been made to standardize income measurement in national accounts and household surveys separately or using multiple sources. Notable examples are the 2008 revision of the System of National Accounts (United Nations, 2008), the Canberra Group Handbook on Household Income (UNECE, 2011), and the Distributional National Accounts (DINA) concepts and methods developed in Alvaredo et al. (2020). All three will be used throughout this work.

Ultimately, our research is motivated by the need to improve the measurement of well-being in Chile. Indeed, there is a growing consensus among economists and policymakers to do so from a broader perspective that considers vertical, horizontal, and opportunities inequality, as well as subjective well-being, economic security, sustainability, trust, and social capital, in addition to the

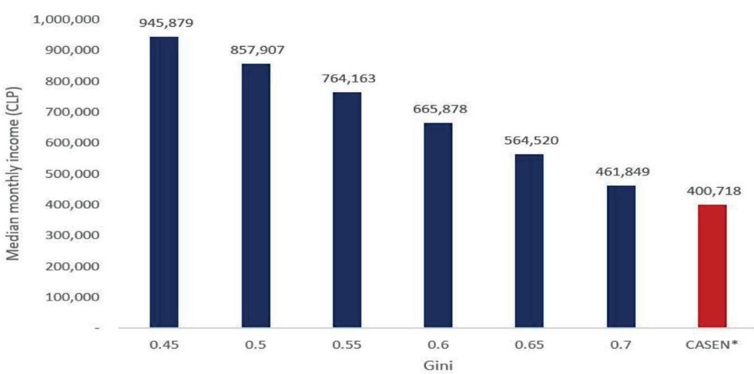
4 These statistics correspond to the measurement of income as defined by the CASEN survey and are adjusted for inflation but not to national accounts.

usual focus on income, consumption, and wealth (Fitoussi et al., 2018). These authors highlight DINA as one of the most significant recent advances in the measurement of well-being. Although restricted to the monetary dimension, DINA generate income distributions that are comparable between countries and alleviate the known problems of the different sources they incorporate. The results obtained using DINA should always be interpreted in a context that considers other aspects of well-being.

To gain understanding of the relationship between aggregate or average income and inequality we perform a simple statistical exercise following Pinkovskiy and Sala-i-Martin (2009). Using a standard distribution (lognormal or Weibull), a Gini coefficient greater than 0.7 would be needed to generate the average income reported in national accounts (CLP 1,350,000) and the median income reported by CASEN (CLP 400,000) (Figure 1). The methodology used to obtain this result is detailed in the Appendix. The extreme levels of inequality implied by a Gini of 0.7 do not seem consistent with empirical evidence, even considering reasonable uncertainty. The highest Gini coefficient calculated for Chile in the literature is 0.63 (López et al., 2016). It is therefore sensible to conclude preliminarily that the CASEN survey significantly underestimates income in the middle of the distribution.

FIGURE 1

PARAMETRIC SIMULATION OF MEDIAN AUTONOMOUS GROSS INCOME, 2017



Note: Own elaboration based on data from the 2017 CASEN survey and the Central Bank of Chile. The median income for each Gini coefficient level corresponds to own calculations based on Gross National Income (GNI), capital depreciation, and indirect taxes described in the Appendix. The x-axis shows the Gini coefficients necessary for each median income.
* The last column corresponds to the income from capital and from gross dependent and independent labor as reported by CASEN, indicating that the Gini coefficient implied by this mean-median combination is greater than 0.7.

It is important to recognize that there is still no consensus on whether aggregates obtained through national accounts are more accurate than those obtained through consumption surveys or income surveys. And advocates of each highlight valid conceptual problems with their nonpreferred source (Pinkovskiy and Sala-i-Martin, 2016). That being said, for the exercise displayed in Figure 1 it is assumed that national accounts aggregates are more accurate. Since some components of national accounts household income, such as measures of capital income and self-employment, are residual, the precision that can be expected from them is lower other items measured directly. For the construction of other components, such as wages, multiple sources are used, contrasted, complemented, and updated in the same coherent conceptual framework. For each type of economic activity, depending on the characteristics of the activity, specialized surveys, accounting and administrative records, sworn statements, and sector indicators are used, harmonized, and adjusted to concepts, definitions, classifications, and methodologies established in the SNA 2008 (Central Bank of Chile, 2017). Surveys, on the other hand, are based on self-reporting and have multiple known problems related to precision, representativeness, and systematic individual biases (Moore et al., 2000; Hurst et al., 2014).

National accounts also capture substantially more income than surveys for independent (self-employed) and informal labor income. Income from independent work presents particular accounting and statistical challenges that make its measurement in surveys, national accounts, and tax records difficult. These difficulties are related to the imperfect accounting of income and expenses, voluntary or involuntary omission when declaring income, and even the blurred boundary between informality and illegality (Hussmanns, 2004). National accounts in Chile explicitly estimates the informal sector using imputation methods based on employment and small entrepreneurship surveys.

Notwithstanding the advantages described, macroeconomic aggregates provided by national accounts only present the total of each variable without reporting its distribution. Hence, it is not possible to obtain population characterizations beyond averages. One population statistic that is particularly useful for summarizing information on variables associated with well-being is the median. From a statistical perspective, the median is not sensitive to extreme values, so it is less likely to be affected by errors in the sample. Moreover, it is more than adequate to measure economic well-being because, in general, it is positively affected by growth and negatively by inequality (Birdsall and Meyer, 2015). The median is therefore frequently used as a reference in public discussion because it is commonly understood and recognized as a good summary statistic.

In order to generate better approximations to the well-being of populations, various efforts have been made to achieve conceptual and methodological compatibility between surveys and national accounts. For example, until 2011, the CASEN survey adjusted five components of income to their theoretical counterparts in national accounts. The adjusted components were wages and salaries, independent income, social security benefits, property income, and imputed rent. These adjustments were discontinued because, among other reasons, they reduced the poverty rate by more than 1.5 percentage points (pp) for some years, and there was not sufficient certainty about their methodological validity (Campos et al., 2013).

This study primarily uses the DINA methodology developed in Alvaredo et al. (2020). The objective of the DINA project is to create a systematic conceptual framework to obtain homogeneous and internationally comparable income series in the same way that the SNA (United Nations, 2008) guidelines allow the precise comparison of GDP levels and other macroeconomic aggregates between countries. Other projects, such as the Luxembourg Income Study Database (LIS), provide harmonized and systematized distributional data and explain conceptual differences; however, the data is not standardized so the comparison between countries is less straightforward.

At the time of writing, the only study that follows DINA guidelines for the full distribution in Latin America is Morgan (2017), who characterizes inequality in Brazil. Our investigation serves as the basis for improving the measurement of capital income and at- tributing the components of GDP that are not directly accrued by households, following the complete DINA methodology. We do not distribute some income components to individuals to avoid relying excessively on arbitrary assumptions and jeopardizing accuracy.

The rest of the paper is organized into seven parts. Section 2 reviews the literature on the use of national accounts and taxes in generating microeconomic statistics. Sections 3 and 4 provide the conceptual framework necessary to understand the relationships between the different definitions of income that will be used. Section 5 presents a selection of sample statistics on income. The gap between income measured by the CASEN survey and the GNI is also described and quantified. Section 6 explains the methodology used to correct survey biases, and Section 7 presents the main results, with an emphasis on the median income for the different categories of workers. Finally, Section 8 concludes and establishes some areas for future research.

2. Literature Review

The harmonization and integration of national income measurement has been going on for decades and continues to develop. The first international effort to coordinate and standardize the measurement of aggregate income was the SNA, first published in 1953. Before this, some household surveys were undertaken, and partial national accounts and administrative tax records were collected, but none of these data provided an adequate level of international comparability.

Aiming to improve the measurement of income in Latin America, two pioneering papers by Oscar Altimir (Altimir (1986, 1987)) proposed an explicit methodology of adjustment for five specific income components captured in surveys to their theoretical counterparts in national accounts. Following the publication of these papers, detailed discussions on the different ways to perform these adjustments have been undertaken by Székely et al. (2000), Ravallion (2000), Deaton and Dreze (2002) and others, who describe the advantages, limitations, and alternative methods of harmonization. It is still an active research topic and as yet there is no consensus on the best way, or indeed whether it is desirable, to carry out these adjustments. Moreover, the informational requirements to perform the adjustments in an optimal manner are costly; thus, for researchers the decision is often between adjusting inadequately or not adjusting at all.

Many of issues with income surveys stem from the fact that they are susceptible to multiple errors and biases, such as survey design errors, sampling errors, questionnaire errors, or data treatment errors. Additionally, one fundamental problem is the nonresponse and underreporting bias associated with socioeconomic characteristics. Interviewees—particularly independent (self-employed) workers—often do not record the accounting information necessary to answer appropriately. In addition, misrepresentation or omission of information, whether done voluntarily or involuntarily, is a significant problem (Moore et al., 2000). These biases are usually greater and more relevant for the upper part of the income distribution range (Ruiz and Woloszko, 2016; Lustig et al., 2020).

Despite multiple methodological problems and inaccuracies associated with the use of national accounts averages with survey distributions, until the 2010s the majority of inequality and poverty studies simply used GDP per capita as a proxy for the average income of individual households.[5] More recently, the consensus has been not to adjust to national accounts when estimating world poverty. Lakner and Milanovic (2016) and Ravallion (2012), for example, use different methods to harmonize definitions and correct biases. However, the statistical techniques and methodological corrections used to estimate incomes, with or without adjustment to national accounts, are generally simple due to the large volume of data and sources used. Although measuring income on a global scale, trying to estimate world poverty, for instance, poses challenges for a more detailed treatment of the information. When investigating a

5 The most relevant studies that use this method are Bourguignon (2011), Pinkovskiy and Sala-i-Martin (2009), Dowrick and Akmal (2005), and Chotikapanich et al. (1997)

specific country, it is feasible to reconcile the different definitions, magnitudes and methodologies with greater precision, which favors the decision to adjust to national accounts. The CASEN survey, for example, adjusted to national accounts from 1990 to 2011, but this decision was repeatedly questioned, most notably by Bravo and Valderrama (2011).

The Bravo and Valderrama paper, which reconstructed CASEN databases from 1990 to 2006 without adjustments, showed that the effect of adjusting to national accounts reduced poverty by 0.6 pp and increased the Gini coefficient of total household income by 3.4 pp. And the same year the paper was published, the Central Bank of Chile shifted the base year of the SNA from 1993 to 2008. A change that revealed the arbitrariness of the adjustment.[6] Between 2000 and 2011, a spliced series was used for the adjustment to national accounts that combined the growth rates of the new accounts with the base of the year 2000. Over time, spliced accounts came to represent only 69% of current year accounts. With the change in the reference series from SNA 1993 to SNA 2008, the adjustment factors increased to considerably higher values: from 1.09 for dependent income and 2.05 for independent income (using spliced accounts) to 1.31 for dependent income and 3.79 for the independent income (using updated accounts). The 20% increase in dependent income (wages and salaries) and the more than 100% increase in independent income (self-employed) could only be a consequence of an external change in methodology, making it impossible to establish continuity between the inequality and poverty estimates from previous surveys. It was therefore decided that adjustment to national accounts for the CASEN survey would cease, only the correction for nonresponse was retained.

The hegemony of national accounts in international average income comparisons suggests they are implicitly considered more accurate than surveys at the aggregate level. However, adjusting survey data to national accounts is not recommended when there is no adequate information on the equivalent aggregate to be used (Eurostat, 2018). Similarly, even if the total magnitude of the difference and compatibility of the specific component is known, adjustments are not recommended if it is suspected that the biases of the survey are not uniform throughout the distribution (Ravallion, 2016; Milanovic, 2012). Assigning the income difference proportionally, for example, can artificially inflate the income of the lowest deciles; thus, incorrectly reducing the proportion of low-income earners in the distribution.

In accordance with one of its main objectives, CASEN characterizes poverty in Chile as well as the evolution of poverty over time, and as discussed, it no

6 The main changes between SNA 1993 and SNA 2008 affected the measurement of financial services, insurance (except life insurance), and production for own use. These changes increased GDP for OECD countries by an average 3.8%.

longer adjusts any income components to national accounts. When the focus is on the middle and upper sections of the distribution, however, international comparability is more important, and it is necessary to implement a methodology that addresses the known biases that affect household surveys, as well as the methodological differences between them.

Székely and Hilgert (1999) review the methodological and conceptual differences between income surveys undertaken in Latin America. They found that wages in formal employment are underreported by up to 57% when compared to data from national accounts, and show important heterogeneities. In a more recent study, Del Castillo (2015) found an average underestimation of wages by household surveys in Mexico of up to 47% compared to this component in national accounts data. Castillo proposes to correct difference between survey-reported wages and national accounts by assigning the share of the gap to three employment groups in a new apportionment, favoring concentration in higher income categories. Specifically, and in a justified but ultimately arbitrary manner, the allocation is 80% to officials, managers, and bosses; 15% to professionals and technicians; and 5% to the least qualified workers. To close the gap with national accounts, each observation is multiplied by a constant according to its category. With these corrections, it is estimated that the underreporting of median income is 47%. Another approximation to the underreporting problem is provided by Lustig et al. (2020) who analyzes the biases that most affect the high-income, missing rich in household surveys, as well as the magnitude and these biases, and potential correction methods. Lustig distinguishes between correction approaches within the survey and those that combine survey data with external sources, generally tax, in a similar way to the correction proposed in this paper. Atkinson (2017) recognizes, however, that bias found in household surveys may not only be limited to high income earners but may also extend to low-income earners where the bias could be even more intense.

Currently, the most prominent methodology for measuring income and inequality is DINA, developed by Alvaredo et al. (2020). Although they recognize that the concepts of national accounts are not the most appropriate to measure economic well-being, DINA uses the SNA 2008 guidelines to maximize the comparability and consistency of income measures between countries. Thus, underestimating the income of countries with the highest spending on public goods is avoided, for example, because this component is not usually captured in surveys. Likewise, we recognize that globally speaking higher incomes are underrepresented in surveys while tax records are less prone to this problem and, national accounts are able to better capture aggregates.

3. Income Definitions

As discussed, some of the disparities in estimates for average income and inequality are due to variation in the definitions of income. Aside from conceptual differences, there are differences in data sources, collection timing, and estimation techniques. In general, no major effort has been made to ensure that the definitions of income or the size of the different components of income from microdata are compatible with national accounts. This section provides the rationale for our proposed methodology that combines the three different data sources used in the paper.

Taxable income varies according to the specific laws of each country. Income surveys also vary from country to country according to institutional needs or simply by lack of proper standardization. Several income and tax questions are often arbitrarily included or excluded from surveys (UNECE, 2011). Considering this and to facilitate understanding, we present the definitions and relationships between the main income categories according to surveys, national accounts, and administrative records, in the Chilean setting.

3.1 National Accounts

Gross Domestic Product (GDP) is generally used to measure income or production in a country. Yet, even if non-monetary dimensions are not considered, GDP is only an approximation of the well-being of individuals and households. A significant proportion of the components that make up GDP are not consumable or do not allow the accumulation of wealth. The most significant examples of nonconsumable components are capital depreciation and net factor payments abroad. Other components, such as indirect taxes and profits not distributed by companies, constitute gross income for the household sector; yet, it is complex, and even arbitrary, to assign a distribution to these components at the individual level. For many of the components, the data necessary to obtain distributive welfare measures consistent with national accounts are not available for Chile.[7]

Ideally, DINA assign a fraction of the consumable components of gross national income to each individual; in practice, however, the information requirements to achieve this are expensive and often unavailable so simplified versions are developed. Indeed, there are a wide range of sophisticated versions of DINA used to allocate these components, but the simplest are usually selected as there are no adequate justifications to the contrary. For example, by

[7] There are no regularly published estimates for capital depreciation, withdrawals from income of quasi-corporations, operating surplus of the homeownership sector and mixed income, among others. This issue will be discussed in Section 4.

assigning a constant fraction of government expenditure to all individuals and other proportional to income or the retained earnings proportionally to those distributed. The focus of this study is on the middle of the income distribution range, therefore we concentrate in labor income. This is because capital earnings are only a relevant source for the richest households (the first four quintiles of autonomous income in CASEN receive only CLP 3,000 in average monthly capital income). And, in addition, the interconnected structure of companies and individuals, together with tax evasion,[8] impose drawbacks that are difficult to remedy.

In order to use national accounts concepts to measure the income of individuals, it is necessary to understand the relationship between the household sector and the other sectors that make up the national economy. According to the generation of income account (SNA 2008), GNI from the perspective of income is equal to the compensation of employees, gross mixed income, gross operating surplus, and net indirect taxes on production (value added tax, import duties, subsidies, among others) that are generated by the economy. Then, the allocation of primary income account records how production is distributed to households and other sectors (private nongovernmental organizations (NGOs), nonfinancial companies, financial companies, government, and the rest of the world, not included in the GNI). In the same account, property income, receivable and payable, is distributed generating only a resource reallocation resulting in the balance of primary income. In the secondary distribution of income account, the income tax, social contributions, and benefits and transfers are incorporated. Social transfers in kind are recorded in the income redistribution account in kind. And, finally, in the income use account, individual consumption, collective consumption and savings of each sector are reported. In this investigation, we will use components from the three accounts to generate our estimates.

Within the income generation account, compensation to employees is defined as the total compensation (wages and salaries), in cash and in kind, paid by an employer to an employee in exchange for the work performed by the latter during the accounting year. It includes bonuses and housing allowances, as well as income tax and social security contributions that the employer makes on behalf of the employee. Some goods and services that employees receive but are obliged to use at work are excluded. For each productive sector, the information on wages is obtained from administrative records and statistics on employment and labor costs. Specifically, the annual income tax return, the sworn statement on the salaries paid by employers, and the tax income and expenses reports from the General Treasury of the Republic are used. These data

---

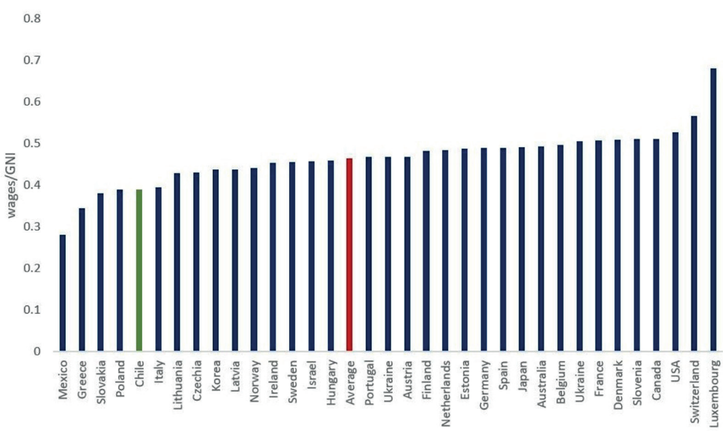
[8] According to Fairfield and Jorratt (2016) the evasion of the complementary global tax would reach more than 45%.

are complemented by employment statistics that from National Employment Survey (known by its acronym in Spanish, ENE) and the Labor Cost Index (Spanish acronym, ICMO), both provided by the National Institute of Statistics (Spanish acronym, INE) (Central Bank of Chile, 2017). In this way, multiple sources are integrated, reducing the probability that errors will persist.

Figure 2 shows the share of wages in GNI according to national accounts for OECD countries. Chile has the fifth lowest share out of the 35 countries, at 39%. This equates to 8 pp less than the group average and is indicative that the share of labor with respect to capital in the product is low. The indicator shown in Figure 2, however, does not correspond to the labor share in production. To estimate it, the depreciation and indirect taxes net of subsidies must be subtracted from GNI leaving the net national income (NNI) at factor cost. NNI is equivalent to the compensation of employees (attributable to work) plus the net operating surplus (attributable to capital), and the net mixed income (attributable to capital and labor). Considering this, the share of wages in GNI is expected to be higher in countries with greater economic development because of the lower rate of informality and concomitant lower mixed income. Still, available estimates of the labor share in Chile are low. Guerriero (2019) reports comparable labor share values between countries using different methodologies, estimating that the participation in Chile is 0.55 whereas the median for the world is 0.7.[9]

FIGURE 2

WAGE SHARE OF GNI, OECD AVERAGE, 2006–2017



Note: Own elaboration using data from the UN statistical Database.

9 This results correspond to the LS4 calculation in Guerriero (2019) that considers an imputation of the same share of the rest of the economy for the mixed income, equivalent to the ratio of wages and GNI minus mixed income.

Regarding capital income for activities with market production, the levels of gross operating surplus are obtained as a residual between production, intermediate consumption, salaries, taxes and subsidies. Estimates of this variable gathered in the production accounts are complemented with information from the annual income tax return. In Chile, this surplus is not recorded separately from mixed income for each sector. This implies that when quantifying mixed income and gross surplus, errors are more likely to be made because they are not measured directly.

3.2 Household Surveys

The CASEN Survey reflects international standards for income statistics established in the Handbook on Household Income Statistics (UNECE, 2011). In the Handbook, household income is defined as all income, monetary and in kind, that households receive at annual, or more frequent, intervals, excluding irregular and one-time payments (such as inheritances and earnings from gambling). This income must be available for present consumption and must not reduce net wealth or deteriorate equity positions.

In line with these standards, CASEN covers income from labor, capital, self-provision of household goods, and various transfer payments. Labor income is therefore considered to be all earnings obtained by people in their occupation from wages and salaries (monetary and in kind), earnings from independent work, and the self-provision of goods produced by the household. Autonomous income is defined as labor income plus interests, dividends, withdrawals from corporations, pensions, assistance funds, and current transfers between individuals. Cash income incorporates state subsidies. Finally, total income includes the imputed rental value of the home. Capital and labor income, both dependent and independent, is always recorded after deducting income tax, health contributions, social security, fees.

CASEN provides very detailed information on labor income. To capture income from dependent work the survey includes up to 10 questions about additional income (overtime, bonuses, tips, etc.), as well as up to 13 questions about income in kind (housing, transportation, food, etc.). For independent income, 6 comprehensive questions are included. Given this extensive survey detail, it is not reasonable to assume that the underestimation of the aggregate level of wages is because the questionnaire does not capture some income items obtained by individuals. The only notable exception is the omission of questions regarding third (and subsequent additional) dependent job. In any case, the aggregate magnitude of income third or higher job is expected to be small.

The COVID-19 pandemic posed significant challenges for all statistical systems and, particularly for CASEN 2020, resulting in major methodological changes. In response to the health restrictions, the 2020 version was conducted in a mixed sequential mode with three phases: face-to-face pre-contact, telephone application of the questionnaire and face-to-face recovery. This differences in the application mode imply possible changes in various aspects related to measurement, that makes difficult the comparison.

3.3 Tax Records

Tax income, according to Chilean income tax law, consists of all benefits, profits, and increases in equity that are received or accrued, whatever their nature, origin or denomination. It incorporates income from work and property, and also includes capital gains which do not constitute income in surveys or national accounts. Despite what is stipulated in general terms by law, in practice a few items are exempted. Among them, mortgage interest is deductible and income from the rental of DFL-2 homes[10] does not constitute taxable income. Additionally, legal contributions for health, social security, and voluntary social security savings are discounted. The composition of income, without considering the exempt items, does not affect the total tax payment because the Chilean tax system was fully integrated until 2016 and semi-integrated for some types of companies since 2017.[11]

3.4 Tax And Social Security Structure

The main direct tax in Chile is the income tax, which is broken down into three different levies: a flat rate on company profits, called the First Category Tax (known by its Spanish acronym, IPC); a tax on dependent labor withheld monthly and paid by the employer, called the Second Category Tax (Spanish acronym, ISC); and a general tax that is levied on all taxable income generated by natural persons, called the Complementary Global Tax (Spanish acronym, IGC). There may be small differences in the tax paid depending on the regime adopted and the source of the income. In the case of IGC or ISC, these differences are inflation-related because the ISC is paid month to month by the employer and the IGC is paid annually.[12] Likewise, as mentioned, there is also

[10] All houses below 140 square meters have special tax and real estate contributions exemptions.

[11] As of January 2017, companies that meet certain conditions, including public limited companies, are required to use the semi-integrated tax regime where the maximum marginal rate (for persons) is 44.5% instead of 35%.

[12] The ISC paid monthly serves as a tax credit for the IGC for payments at the end of the year in case the taxpayer has more than one source of income.

a difference between the attributed income regime and the semi-integrated regime when taxing income from capital combined with other sources.

In addition to direct taxes, dependent workers are subject to legal deductions charged to their employer: 10% for mandatory pension savings; 7% for public or private health insurance; 0.6% for pension insurance paid by the employer; close to 2.5% for insurance against work accidents, and disability or survival insurance, also charged to the employer; and a variable fee close to 1% paid as commission to private companies that administer the pension funds (known by their acronym in Spanish, AFPs). For simplicity, all the discounts will be called "social security", despite the fact that the OECD only considers 7% of health contributions as social security. It will be assumed that all dependent workers with a contract are subject to these contributions, even there are particular situations that exempt some taxpayers from contributing to some components, as well as employers who evade them.

4. Relationship Between Sources

In order to incorporate the different datasets into a coherent conceptual framework that takes advantage of the information provided by each source, it is necessary to generate harmonized income definitions and establish equivalences between the different sources. Income as measured by the CASEN survey will be used as a point of reference because the proposed adjustments will be applied to CASEN data.

All income variables in CASEN are explicitly or presumably recorded net of tax and deductions. To obtain the net taxable income, capital gains must be added to the autonomous income given in the survey, and compensation in kind and the lease of DFL-2 real estate must be subtracted.[13] Income is recorded net of social security contributions, which are tax exempt. Additionally, the CASEN survey does not record the effective payment of taxes; thus, to obtain gross income certain assumptions about tax evasion and compliance must be made. One last minor difficulty is that voluntary tax-deductible pension savings are also not reported, so it is assumed that this component is reported as part of net income.

Even if the missing variables discussed are ignored, to reconstruct GNI or other national accounts concepts based on the components captured in surveys requires certain assumptions to be made, and, occasionally, it is not possible to achieve complete compatibility. This is because some differences in scope, gaps in data collection, or quality problems remain, such as underreporting in the case of surveys and indirect or residual measurement in the case of national

[13] Information on the lease of DFL-2 real estate is not available in the survey.

accounts (UNECE, 2011). The degree of comparability between surveys and national accounts varies according to the particular component of income being compared. Wages, social security, and earned income taxes have high comparability; and income from self-employment, property, transfers, and imputed rent have medium or low comparability. High comparability means that the concepts can be homologated without adjustments. Low and medium comparability implies that there are elements measured in one source that are not in the other and vice versa. Elements that do not coincide between sources are not measured, so it is not possible to refine the concepts until they are completely comparable (Eurostat, 2018). Concept comparability may be correlated with reporting biases in surveys, but it is not the same, as two perfectly equivalent concepts can be measured incorrectly in one source.

Before turning to labor income, we propose a compatibility exercise between survey income and GNI.[14] By performing this exercise, it is possible to distinguish between differences attributed to omitted elements and differences attributable to the sum biases. Taking into account the above, GNI is equal to the sum of gross disposable income of each sector that is, primary income plus redistribution accounts. Thus, the elements of the household account that are not measured, as well as the income flows that are retained or finally spent by another sector, must be added to the accounts actually measured in CASEN. Income measured by CASEN is added to income tax and social security contributions and other income items not measured in the survey, such as financial intermediation services indirectly measured (FISIM)[15], and others[16], to obtain the gross disposable income of households.

4.1 Wages

Achieving conceptual compatibility in this income component requires only minor considerations. According to the SNA 2008, compensation to employees includes wages and salaries payable in cash or in kind and contributions to social security systems paid by the employer or employee, recorded gross of income tax or other discounts. The variables necessary to construct wages are directly accessible in the CASEN survey and in national accounts. Information regarding the tax and social security burden, however, is not available as it is not directly queried in the CASEN survey questionnaire.

[14] We are thankful for the comments from the ECLAC Department of Statistics on this matter.

[15] The gross value added for other services is calculated as the charge for the service minus intermediate inputs. For financial services it is recorded in the same way but adding property income received, less paid, to account for the opportunity cost of resources that are not explicitly charged.

[16] Specifically: Gross operating surplus of the housing sector (minus imputed rentals, gross saving of companies, income attributed to insurance policy holders, gross operating surplus of the government, consumption of NGOs and a price level correction.

4.2 Independent Income

The construction of independent income from national accounts concepts and its homologation to the concepts measured in CASEN is more complex than it is for wages. In contrast to wages, for example, no individual component in national accounts corresponds exactly to this independent income. Another difficulty is that, unlike the other accounts specified in the SNA, there are sizable disparities in the treatment and categorization of some elements that make up independent income. Finally, even if a single conceptual framework existed, it would not be feasible to individually measure some of the components necessary to construct independent income from the information available.

Recognizing potential irreconcilable differences between survey independent income and its national accounts counterpart, the convention to construct it is to use the sum of gross mixed or net income (codes B3b or B3n in SNA 2008) plus the income of quasi-corporations (code D.422 in SNA 2008) that are received by the household sector. Income from unincorporated corporations—that is, quasi-corporations—can potentially be found in two SNA categories: net mixed income and withdrawals from income of quasi-corporations. Knowing in which of the two categories the income of quasi-corporations is accounted for requires access to methodological notes for the country under investigation. When using both accounts together (code D.422 and B3b), however, it is not problematic to be unaware of the location of this component.

Due to the cost of measuring each subcategory in national accounts separately, in Chile less detail is provided for some items. No official estimates of depreciation are provided at the aggregate level; thus, these estimates are not available for mixed income either, only gross mixed income (B3b) is reported. Another important information deficiency is that gross mixed income is reported in conjunction with the gross operating surplus (B2b). In any case, information is provided for the gross surplus and mixed income of the household sector and for the gross surplus of the housing sector. Similarly, item D.422 is not accounted for separately from dividend income (code D.42 in SNA 2008).

Considering that dividend income is not separated from quasi-corporation income, to construct the second part of independent (self-employed) income—the income of quasi-corporations—it is necessary to assign a fraction of the dividends to arrive at the aggregate independent (self-employed) income. In an adjustment implemented until 2011, ECLAC as-signed 90.7% of total distributed income to independent (self-employed) workers for all years, without outlining the justification for this adjustment.[17] It can therefore be concluded

[17] The only reference to this adjustment is in an internal ECLAC Excel spreadsheet stating that it corresponded to a “historic coefficient”.

that this adjustment is the main reason why ECLAC reported moderate adjustment magnitudes for capital income and unreasonable values for independent (self-employed) income. See Table 7 and the discussion that accompanies it for clarification of this argument.

Despite the difficulties independent (self-employed) workers have in keeping an adequate accounting record for their income flows and costs, it is not plausible that they report less than a third of their net income. Moreover, there is abundant evidence that income from capital is heavily underreported in surveys. Thus, we decided for the central calculations, in a more credible but also arbitrary manner, that 20% of the income distributed from national accounts corresponds to quasi-corporations. In any case, we perform a sensitivity analysis.

Considering these limitations, a national accounts equivalent was prepared for independent (self-employed) income. It is not free from inaccuracies or arbitrariness. It is, however, an improvement over the old CASEN adjustment methodology developed by ECLAC. The mixed income of households is obtained by subtracting the gross surplus of the homeownership sector from the gross surplus and mixed income of the household sector. A fraction of the dividends paid by companies (D.42) is added to approximate the required item, D.422. As specified, 20% is selected to correspond to quasi-corporations in the central scenario, a conservative estimate to avoid overestimating the income of independent workers.[18] A higher value would give a quotient between the total of national accounts and the total of the CASEN survey greater than 2, which is not realistic. We recognize that the choice is therefore arbitrary and has a considerable impact on independent (self-employed) income. For instance, if the value is set at 90% the ratios between the aggregates are close to 4, as shown by the ratios calculated by ECLAC[19]. Table 7 shows the evolution of the calculated ratios, or adjustment factors, of independent income considering different fractions of D.42. Depending on the proportion of the component that is imputed, the adjustment factor can take values between 1 and 3.9.

An effort was made to construct an equivalent aggregate for independent income, but important inaccuracies persist. Accordingly, we present the results of dependent and independent employment income separately, and highlight that dependent income is much more precise and that independent income should only be used as a reference. It would be a significant improvement to integrate in a more clear, complete and disaggregated way the elements of national accounts necessary to adequately measure the independent sector. The investigation of the independent and informal sector from the micro- and macroeconomic perspective is a separate topic and is in its early stages in Chile.

18 This figure is 27% less than the same ratio using comparable US national accounts data in 2016.

19 See Literature Review

4.3 Capital Income

Capital income measured in the CASEN survey taken to national accounts concepts corresponds to interests received by households (D.41) net of FISIM and distributed income of the companies corresponding to dividends (D.421). Between both these items in national accounts, equivalent to the income of capital measured in the survey, a maximum of 15% of GNI is reached in the years analyzed.

The Central Bank of Chile does not regularly publish FISIM for household deposits, but it does for the Chilean economy as a whole. To approximate the proportion of FISIM that corresponds to household income, the average of the years for which information is available for FISIM income account prepared by ECLAC between 2008 and 2013 (29.8%) was used, approximated to 30%. Constructed FISIM makes up less than 4% of capital income, so attributing a different fraction to FISIM does not significantly alter the results.

Nonresidential property and machinery leases are not recorded as property income in national accounts. Because the gap between capital income in CASEN and in national accounts is higher than 300%, all related items in capital income from the CASEN survey are included. In the Online Appendix, all the series used, their sources, and the relationships between the variables used for the adjustment to national accounts are given.

5. Descriptive Statistics

This section describes the sources and provides relevant descriptive statistics from the CASEN survey and national accounts, and data reported by Chile's Internal Revenue Service (known by its acronym in Spanish, SII). Additionally, following the guidelines set out in Section 4, the gap between income captured in the CASEN survey and national accounts is quantitatively decomposed. We use 2017 as our references since it is the last year with data unstained by methodological adjustments and macroeconomic effects due to the pandemic.

In Table 1 statistics from the CASEN survey and the SII for income of dependent and independent workers and net taxable income are reported. Notably, the last two rows of the table show that the average net taxable income measured in the CASEN survey is only 6% lower than the same statistic as reported by the SII. The gap between the 90th percentile, however, increases to 29%. Similarly, the share of income of the richest 1% is almost 2% higher in absolute terms and 16% higher in relative terms according to data from SII, reflecting the missing rich phenomenon. Taking into account that the tabulated tax base includes 600,000 more individuals than CASEN[20], the gap between the richest in tax data and the richest in CASEN is greater if the comparison is made by number of individuals rather than percentiles or shares. On the other hand, median taxable income (P50) is 12% lower according to the SII data[21].

The information in Table 1 allows us to conclude that income captured by the upper end of the distribution is underreported in the CASEN survey. It also shows that, income measurement is apparently more accurate for the middle and lower income brackets in the survey. One reason for this is that the tax-exempt bracket starts from approximately the 75th percentile and below, and therefore the incentives to report or audit correctly are weaker below this percentile.

Table 1 also highlights some differences between labor income from dependent and independent workers. Dependent workers are a much more homogeneous category than independent (self-employed) workers, which is reflected by the fact that the income share of the top 1% dependent workers is just 9%, almost half that of independent (self-employed) workers. Another difference highlighted by the table is that although the 90th percentile of both categories is very similar, the median of dependent workers is 40% higher. This difference is not only caused by gaps in human capital but also by the considerable heterogeneity in the use and ownership of physical capital for independent workers. As a group, independent workers include low-skilled informal workers together with employers and entrepreneurs of different sizes that may have much higher incomes.

It is also useful to consider the distribution of income that is provided by official administrative records. Figure 3 shows the average and median salaries for November 2017, according to the records of Chile's private national pension system (known by its acronym in Spanish, AFP), the unemployment insurance (Spanish acronym, AFC) and the CASEN survey, together with the monthly average of national accounts. The 27% gap between the monthly average wage in national accounts and the two other data sources (AFP and AFC) is partially explained by nontaxable income and unrecorded year-end bonuses. The primary explanation, however, is by the evasion of state-mandated social security contributions. The difference between national accounts and AFP and AFC is approximately 3% smaller if untaxed income in kind (2%) and year-end

---

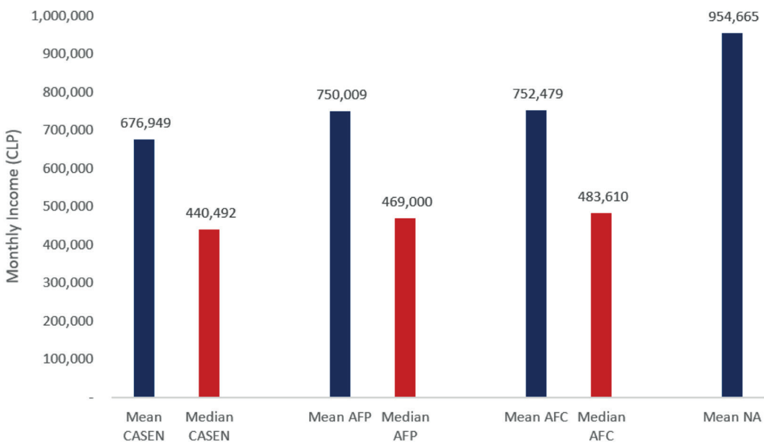
[20] The income difference between high earners in CASEN and the SII is even larger when comparing by ranking rather than fractions (e.g., the thousandth richest individual). The population of tax fillers is mechanically larger because it includes every individual who received taxed income in at least one month instead of just November.

[21] The median, or P50, for the SII data is calculated using the generalized Pareto interpolation method with data tabulated from P75, so it is less precise.

salary bonuses (< 1%)[22] are considered. Moreover, the CASEN survey records information on 6 million worker salaries, whereas there are only 5.3 million AFP records and 4.6 million AFC records. This suggests that approximately 700,000 workers do not pay the state-mandated retirement contributions and that another 1,400,000 do not pay compulsory unemployment insurance. The difference of 11% for the average and 10% for the median between the income of the CASEN and the AFPs and AFC is because the effect of underreporting in the survey exceeds that of evasion. A more thorough study of the phenomena is necessary to quantify the employer contribution evasion; given the data presented here, it is expected to be significant.

FIGURE 3

AVERAGE AND MEDIAN WAGE IN CASEN, ADMINISTRATIVE RECORDS, AND NATIONAL ACCOUNTS, 2017



Note: Own elaboration based on CASEN, Pensions Superintendence, and Central Bank data. The average of national accounts corresponds to item D.1 (wages) divided by the number of salary recipients in the CASEN survey. The median of the AFP and unemployment insurance (known by its acronym in Spanish, AFC) affiliates is calculated using a sample that corresponds to 4% of the universe, and the average is obtained directly from the statistics calculated by the Pensions Superintendence to avoid the bias caused by the data truncated in the taxable maximum.

Table 2 shows the total gap between GNI and CASEN (fourth row) and between its theoretical equivalent constructed from the survey (fifth row). This implies that the survey captures on average only half of GNI (fourth row) and three quarters of the components it effectively measures (fifth row). The data

22 For simplicity, this is calculated as the difference between the average salary in December and the average salary in November divided by 12. This variable is not included in CASEN.

shown corresponds to nominal, per capita, monthly income (the total amount was divided by 12 to obtain monthly data). The first row corresponds to income from work, property, self-provision of household goods, miscellaneous transfers, and imputed rent, always net of taxes and contributions, and measured in CASEN. The theoretical equivalent, shown in the second row, includes those items that, due to their definitions, are not measured in the survey but are included in the GNI from the income perspective. These are income tax (D.5), social contributions paid by households to the government and financial companies (D.61),[23] gross savings from financial and nonfinancial companies (B.8), taxes net of subsidies (D.2 minus D.3), property income attributed to holders of household insurance policies (D.44), NGO consumption, income from FISIM[24] of households, and the depreciation of government assets. Finally, a correction is applied for the differences in price level over the whole year in question compared to the month in which the survey is conducted. If there are no measurement errors or differences in definitions that cannot be reconciled, rows 2 and 3 should return the same values.

Considering results in Table 2, the average underreporting of the components measured in the survey ranges is between 16% and 36%, demonstrating a decreasing trend. This includes some components that are overestimated, such as the imputed rent of houses inhabited by their owners. The percentages in the last row should not be interpreted as only the bias of the quantities measured in the CASEN. Although national accounts are considered to be more accurate, both forms of measurement are subject to bias. This exercise is meant as a preliminary study to better understand the relationship between survey income and GNI; it is not, however, the core of this research.

6. Methodology

This section outlines our methodology, which is inspired by DINA (Alvaredo et al., 2020) but adapted to reflect labor income and median income more accurately. Since our focus is not on detailed treatment of the super-rich, inequality indicators are reported as complementary analysis. The methodology consists of five steps. First, using tabulated tax data, synthetic fractiles[25] of net taxable income are formed through generalized Pareto interpolation, as described in Blanchet et al. (2022b). Second, the net taxable income from

[23] This corresponds to contributions for pension savings and other discounts paid to AFPs.

[24] As explained in the previous section, this income comes from, for example, imputed interest income associated with servicing a bank account.

[25] This general term corresponds to the concepts of quintile, decile, percentile, etc. They are expressed in base 100 notation, as percentiles, but allow decimals.

the survey, including imputations for capital income in constructed. Then, the underreporting and nonresponse at the top of the CASEN survey distribution, or missing rich, is corrected with the synthetic data generated following Blanchet et al. (2022a). Fourth, gross tax revenues are calculated considering the tax-rate structure and evasion determinants of the CASEN survey to obtain a theoretical collection very similar to the effective one. Finally, the different income streams are proportionally adjusted until the magnitude measured in national accounts is reached.

In this paper, the CASEN survey is used as the base data source for the adjustments proposed. It is also possible to use tax data as a base for adjustments. In jurisdictions where administrative tax data is accessible and accurate, such as France and the US, it is natural to use these records as a starting point, even for the lower part of the income distribution and informal workers. However, in countries where tax bases are tabulated, evasion is relatively high, or the informal sector is significant, it is advisable to use income surveys as an initial reference (Alvaredo et al., 2020). The choice of base data source dictates whether the high-income segment is corrected or the middle and low income segments are corrected. The nature and availability of the data for Chile suggest that surveys should be used as a starting point.

As surveys are selected, the data correction methodology fundamentally depends on whether or not there is an interval where both survey data and tax data are considered accurate or reasonably free of reporting bias. If the survey results are reliable up to an income limit that is much lower than the lower limit at which the tax data are reliable, Piketty et al. (2019) recommend to use an adjustment that extrapolates the income distribution in the bracket where no source is known to be reliable. For example, when tax records provide information only for very high-income taxpayers, above the 99th percentile, as they do in the case, for example, of China. Alternatively, when there is an overlap in the income range in which both sources are accurate, it is recommended to use a methodology that is data driven in all percentiles of the final distribution.

In Chile, reasonably accurate tabulated tax information is published starting from the 75th income percentile. Below this percentile, evasion and underreporting of income to the SII means that there are more individuals above the taxable income bracket in the CASEN survey than in the administrative tax data (Candia, 2018). Thus, surveys are the most reliable data source up to a certain percentile and it is only more reliable to use tax data for the upper percentiles, despite the fact that tax data are available for lower incomes.

For these reasons, we decided to use the methodology where the final distribution is based on data, and not on estimated parameters, for all income levels. In what follows, we outline a five-step adjustment specially designed to fully harness available data.

Step 1. A synthetic full distribution of taxable income is generated. The SII publishes information on the average gross income, number of individuals, and the tax collection for each from the seven or eight income brackets[26]. Using the gross annual income and the tax paid for each tranche on average, the net income for each tranche is constructed. The limits of each net income interval are easily inferred from the tax structure.

From the tabulated data for net income, 127 synthetic income fractiles are constructed: 99 fractiles for the first 99 percentiles; 9 fractiles for each tenth of a percentile between 99 and 99.9; 9 fractiles for each hundredth of a percentile between 99.9 and 99.99; and 10 fractiles for the remaining thousandths of a percentile. By allowing the Pareto co- efficient to vary between each fractile, continuous, smooth, and realistic distributions are obtained for the upper part of the curve (Blanchet et al., 2022b). This interpolation exceeds the simple Pareto interpolation because "the shape and thickness of the tail" better fits the empirical data, especially for countries with high inequality (Blanchet et al., 2022b). This is demonstrated by the lower mean relative error compared to the other three commonly used methods of interpolation.[27] In general, the true density at the top of the distribution is greater than that estimated density from a simple Pareto interpolation.

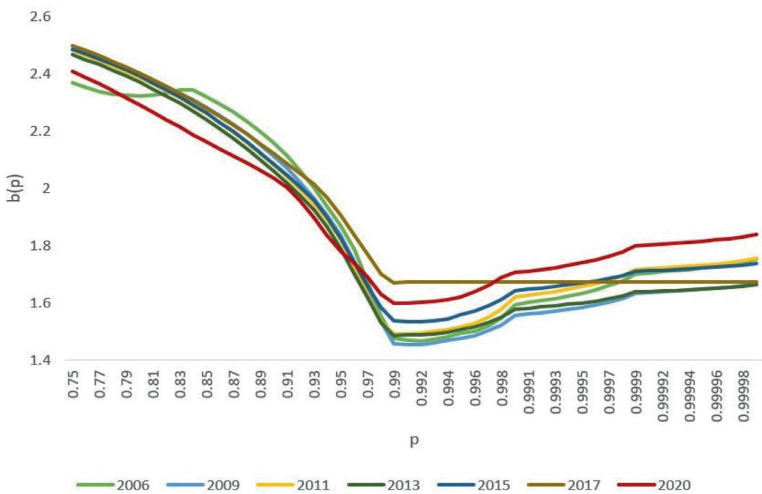
Figure 4 shows the generalized Pareto coefficients $b(p)$ obtained for each percentile. The data comes from SII tabulated data and covers CASEN survey years (i.e., every three years) from 2006 onward. These coefficients correspond to the ratio between the average income of the incomes that are above p percentile and the income corresponding to that percentile. To illustrate, if the 90th percentile receives CLP 10,000,000 and if from that point $b(p) = 2$ for all $p > 90$, the average income of the top 10% of individuals is CLP 20,000,000. Informally, $b(p)$ is a measure of the speed with which income grows at that point in the distribution; thus, the average coefficient, based on a percentile, can be interpreted as a local measure of inequality.

[26] The seven brackets correspond approximately to those separated by percentile 75 (tax exempt section), 90, 96, 98, 99, and 99.5. The eighth bracket (99.7 percentile) was removed by law in 2017.

[27] The code to implement this interpolation is publicly available from the World Inequality Database (https://wid.world).

FIGURE 4

GENERALIZED PARETO COEFFICIENTS (B(P)) FROM NET TAXABLE INCOME



Note: Own elaboration based on data from the SII. On the y-axis, 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.

Notably, the curve plotted in Figure 4 is not constant. This suggests that precision would have been lost if a simple Pareto interpolation from an arbitrary percentile had been used. The second feature to note is that unlike for gross income from tax data[28], the estimated Pareto coefficient decreases rapidly between the 75th and 99th percentiles and then grows moderately. This is consistent with the progressive effect of the Chilean tax system dominating the increasing concentration of income up to the 99th percentile. And from the 99th percentile even net income becomes more concentrated at the top of the distribution.

Step 2. Using the CASEN survey, we construct net taxable income, so it is comparable to the synthetic base generated in **step 1**.[29] To correct for the significant underreporting of capital income we propose a simple strategy, similar to the adjustment used by ECLAC until 2011. The ECLAC adjustment consisted of adding between 2.8% and 12.9% of autonomous income to the highest income quintile from 1990 to 2006 (Bravo and Valderrama, 2011). The problem, however, was the discontinuous nature of the imputation, adding zero

[28] For gross income, the Pareto coefficient b(p) is increasing in income. This demonstrates that the empirical distribution is more unequal than a standard Pareto distribution.

[29] The taxable income corresponds to the income from dependent and independent work, excluding payment in kind, plus dividends, shares, withdrawals, and property leasing. It is not possible to separate the tax-exempt DFL-2 house rental from taxable rentals.

to the 79th percentile and increasing income from the 80th percentile upwards by approximately 5%.

Based on this imputation strategy, we implement an alternative where the proportion of added capital is linearly increasing in income in such a way that the adjustment is continuous even at its starting point. The correction adds capital income proportional to total income, independently of the declared capital income. To the richest individual a fraction, k, of the ratio between the equivalent capital in national accounts and the total autonomous income measured in the survey is imputed, which equates to close to 25% for the years under study. The imputed capital then decreases linearly until it reaches zero at the 80th percentile and below. Accordingly, a continuous adjustment is generated that is consistent with the fact that capital participation tends to grow with income. Capital income is only added to individuals below the 80th percentile if they report it. After this correction, aggregate capital income is still considerably lower in the CASEN survey than it is in national accounts, and top income earnings higher in the tax base.

If this step is not completed, the missing rich correction will incorporate more high- wage individuals at the top of the distribution to the detriment of individuals with high capital incomes. Table 6, in the next section on results, and the accompanying discussion should clarify this argument.

Step 3. Following the methodology proposed in Blanchet et al. (2022a), the net taxable income is used as common variable to adjust for the underreporting and nonresponse of the richest individuals—that is, the missing rich. The tax data is assumed to provide a credible lower limit for the number of individuals who are above certain income level.[30] The 95th percentile is chosen as the confidence limit for the CASEN survey in the central scenario. For the 95th percentile and above it is assumed that the number of individuals represented by each observation is underestimated. As a result, new population weights consistent with the income from the tax records from the selected percentile are obtained.[31] The literature on the concentration of income and wealth of the super-rich suggests that the missing rich phenomenon in surveys is relevant only from the 90th percentile and up (Ruiz and Woloszko, 2016).

It should be noted that the missing rich problem is generated because of the *omission of individuals* who should be represented, as well as the underreporting of income *to a greater degree than in the rest of the distribution*. To correct the representation issue, it is assumed that individuals who did not answer the CASEN survey necessarily declared taxes, which is reasonable. It is not

[30] Of course, the nature of evasion implies that the real number is higher. Fairfield and Jorratt (2016) estimate an evasion level of 46% for the complementary global income tax.

[31] We decided not to use the multiplicative adjustment of income because the factor is greater than 1.5 for less than 5 observations, the average is 0.9999, and the standard deviation is less than 0.005. The adjustment was therefore deemed unnecessary.

assumed, as it would be very unrealistic, that the tax data source includes all taxable income. Tax underreporting is corrected, together with survey biases, in the proportional adjustment to national accounts outlined in **step 5**.[32]

One of the virtues of this calibration algorithm is that it allows the point from which the tax data replaces survey data in the final distribution to be found endogenously. Blanchet et al. (2022a) estimate that the merging point is close to the 80th percentile for the CASEN survey. Adjusting CASEN survey from the 80th percentile is the best choice only when using the CASEN and tax data as only sources, and not national accounts. And because in a later step of the algorithm income is adjusted for national accounts, the objective here is only to correct the additional biases not all biases. For this reason, an intermediate scenario, the 95th percentile, is selected as the merging point.

The methodology used here minimizes a combination of the distortions of the original sample and the deviations with respect to levels of the administrative tax record, maintaining the representativeness of the survey for selected variables. The proportion and number of contributors, dependent workers and independent workers are maintained because the survey is assumed to be representative in those variables.

Performing the missing rich correction adds observations to the top of the distribution. The population weights are rescaled to maintain the representativeness of the total number of individuals and the number of individuals in each income category. Rescaling the population is not relevant for calculating income shares, Gini coefficients, and other distributive analyzes.[33] It is important, however, when calculating the aggregate magnitudes necessary for the adjustment to national accounts. And, regarding this adjustment, although rescaling the population is not neutral in distributional terms, it is a correct approach in order not to generate a population larger than the effective one.[34]

To illustrate the effect of this step on the income distribution, Figure 5 shows the Lorenz curve of total net taxable income before and after correcting for representativeness. The replacement of observations in the upper part of the distribution shifts the entire Lorenz curve to the right, increasing the concentration in the top incomes In this way, the Gini coefficient of net taxable income increases from 0.49 to 0.53 when including the imputed capital and then to 0.56 when correcting for the missing rich.[35] The effect on the median income and on lower percentiles is much smaller or virtually nil.

---

[32] Tax evasion is not calculated separately because evasion is not a central research objective of the paper and several additional assumptions are required.

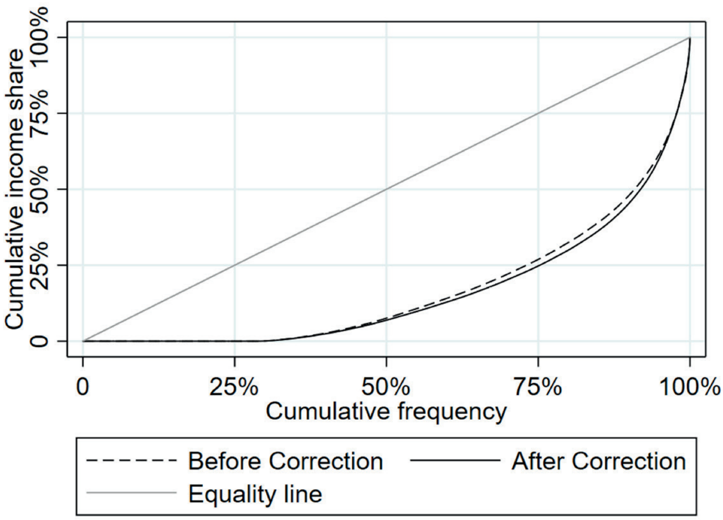
[33] It does not affect median income, Gini coefficients, or the shares of each fractile.

[34] The effect of rescaling is small because the methodology considers that the original population is similar to the corrected one.

[35] The reported Gini is slightly higher than 0.47 from the 2017 CASEN survey data for two reasons. One reason is that taxable income does not include subsidies and the other is that it uses the individual as the unit of observation rather than the household.

FIGURE 5

ORIGINAL AND BFM-CORRECTED LORENZ CURVE FOR TAXABLE INCOME IN CASEN 2017



Note: Own elaboration from CASEN and SII data. BFM is an abbreviation for the authors of Blanchet et al. (2022a).

In order to separate the effects of the adjustment to national accounts from the effects obtained by correcting for the missing rich, results will also be shown omitting this step (3). These effects, however, are not fully separable because the reassigned weightings mainly affect the upper part of the distribution and also shift the entire curve. In the results section, the values obtained when applying this step will be denoted by the abbreviation BFM in reference to Blanchet, Flores, and Morgan, the authors of the paper in which this methodology was developed.

Step 4. Having generated the net income and corrected for underreporting and nonresponse from households with higher income, the next step is to obtain the wages and other income gross of tax comparable with national accounts. For this, the inverse of the tax structure is simulated based on a procedure used by the Chilean Ministry of Social Development. We use variables reported in the CASEN survey to quantify tax and social security contributions evasion. Approximately 15% of dependent workers report they do not have an employment contract. For those, their net income equals their gross income.

The construction of the questions in the CASEN survey means that each variable reported is net of taxes and contributions. It is assumed that dependent employees with contract pay income tax in full. Independent, self-employed workers are supposed to pay tax only if they issue a tax invoice. And social security contributions are imputed only for independent workers who respond that they are registered to an AFP or another pension provider.

All income flows are annualized by multiplying reported values by 12, and then the inverse of the tax structure of the complementary global tax is applied. This requires the assumption that each income stream net of taxes is reported by applying the average rate to it and that the tax is only charged once on all components of income.

Step 5. In the final step, the aggregates calculated for each variable are compared to their counterparts in national accounts. The gap between each of these two aggregates corresponds to the extent of underreporting in the CASEN survey. Conventionally, it is assumed that after the BFM correction the total underreporting should be proportional to income. This implies that the correction consists simply of multiplying by the quotient between each pair of aggregates.

Adjusting to national accounts after adjusting for the missing rich has the advantage that underreporting to the SII is also corrected, which is especially relevant for capital income which is underestimated by around 400% on average.

Of that bias in the original data, approximately 200% is corrected with the capital charged in step 3, 40% is corrected when rebalancing the survey weights, and the rest is corrected in the final adjustment to national accounts. When multiplying by the quotient we assume that the remaining biases are proportional throughout the distribution (of each component); thus, acknowledging there is tax evasion for all income but particularly for capital income.

There are numerous alternative assumptions for the distribution of the gap with national accounts. Some authors perform this adjustment by maintaining the poverty rate of the original sample so as not to "statistically eliminate people living in poverty". One pragmatic reason not to perform the adjustment in this way is because it would be arbitrary, similarly to the capital income adjustment outlined previously, to generate a discontinuous adjustment around a particular point. But there is also a more compelling reason: strongly negative saving rates for the poorest households are calculated by considering consumption surveys and unadjusted income (INE, 2018).

Finally, to evaluate the quality of the simulation, the different aggregates that have been constructed are compared with the corresponding official social security records and the data from the SII.[36] The figures obtained for total deposits and total tax collection are very close to the effective ones. As a reference, the Gini coefficient obtained for 2017 is 0.6 considering gross income and corrections for imputed capital, the missing rich, and national accounts. A data appendix is provided with all the aggregate variables used, their sources, and how they were constructed. The code, also available, allows the adjustment

[36] The social security records were obtained from the OECD.

parameters to be changed and automatically updates the graphs and tables presented in an Excel file. It is therefore possible to verify the dependence of the results on the basis of the assumptions made.

7. Results

This section discusses the main results and provides a robustness analysis. First, the adjustment factors are analyzed in Table 3. Then, Figure 6, Figure 7, Figure 8, and Figure 9 show the evolution of the median and the evolution of the 90*th* percentile for independent and dependent workers between 2006 and 2020. Table 5 summarizes the main results for all active workers in 2017. A sensitivity analysis of the results is performed in Table 6, a varying the two relevant parameters. Finally, Figure 12 motivates a discussion about the implications of our methodology for inequality.

The results and conclusions are focused on the median income of each component, so measurements for high income and inequality are shown only as a reference and to facilitate an understanding of the effects of the methodology at relevant points of the distribution. For wages, the gap between national accounts and the CASEN survey is broken down into 30% corresponding to the proportional correction and 10% corresponding to the missing rich. For independent workers, 56% corresponds to the proportional correction and 7% to the missing rich. In our central scenario, this translates to a median gross income of CLP 600,000 for dependent workers and CLP 570,000 for all active workers. In the proposed sensitivity analysis, the median gross income for active workers is between CLP 540,000 and CLP 600,000. There is at least 30% underreporting in the original 2017 CASEN survey (including taxes and imputed contributions). The underreporting will be referred to as the adjustment factor minus one in order to establish a clear relationship between these two values throughout this section.

The adjustment factors shown in Table 3 correspond to the quotient between the comparable aggregate in national accounts and the total amount reported in the CASEN survey. To understand the effects of each step in the methodology, the ratios between aggregates are presented for before correcting for high income and after the BFM correction. A significant fraction of the gap between the CASEN survey and national accounts is caused by the very low probability of CASEN survey participation of the richest households, as well as their underreporting of income. Both phenomena are corrected by applying **step 3**. The rest of the gap is proportionally adjusted according to **step 5**. The total difference between the CASEN survey and national accounts is at least 26% for wages and up to 85% for independent workers (self-employed).

Notably, wages are underestimated by around 40% in the CASEN survey.

This number must be interpreted in conjunction with the tax and social security evasion reported in Table 4. For all years, except 2006, the simulated tax burden was slightly higher than the actual tax burden; thus, the underreporting could also be somewhat higher. If a higher tax burden is simulated, the gross income before correcting for national accounts grows. In any case, these results show that the CASEN survey bias is quite high.

For social security payments, we see the same happen: if a higher tax or a higher social security burden is simulated, the adjustment factor required is lower. This implies that gross adjustment factors, despite being the most reasonable estimates, are subject to a greater degree of error than gross income itself. The only effect of imputing a higher tax burden is greater inequality in gross income after the BFM correction and less inequality in corrected net income. As both adjustments go in the same direction for gross income, it is less problematic to have a small uncertainty about the exact separation.

The income adjustment factor for independent (self-employed) workers is even higher than wages, reaching more than 80% in some years with an average near 60%. These values are in line with the theoretical arguments associated with imperfect accounting that were set out in Section 4. The figures produced by this investigation, however, are more reasonable than those produced by ECLAC, which showed over 300% underreporting of income for some years. Fairfield and Jorratt (2016) find a ratio of 1.5 (equating to 50% underreporting of income) for 2009 tax data as compared to national accounts.

In order to obtain consistent estimates of gross income and employee compensation from the net variables reported in the CASEN survey, it is essential to appropriately allocate income tax and social security contributions. Accordingly, effective collection was compared with simulated collection for pension discounts and taxes. Table 4 shows that simulated complementary global tax collection was up to 20% higher than its effective collection. For social security evasion, in addition to those who declared not to contribute, those who contributed did not necessarily paid the full amount required every month. That is, they paid social security contributions for a few months income or only declared a part of actual income each month. To correct for partial payment of contributions, each individual payment was divided by the ratio of effective social security collection and the simulated one (after controlling for those who paid zero). Unlike social security contribution, tax payment was not adjusted so as not to underestimate the gross income of the upper percentiles. It should be noticed that the number reported in Table 4 for social security evasion does not attempt to measure actual evasion. The objective of this correction is to obtain simulations that match the real amount of total contributions, but the estimated magnitude of evasion is not strictly separable from wages and independent income mismeasurement.

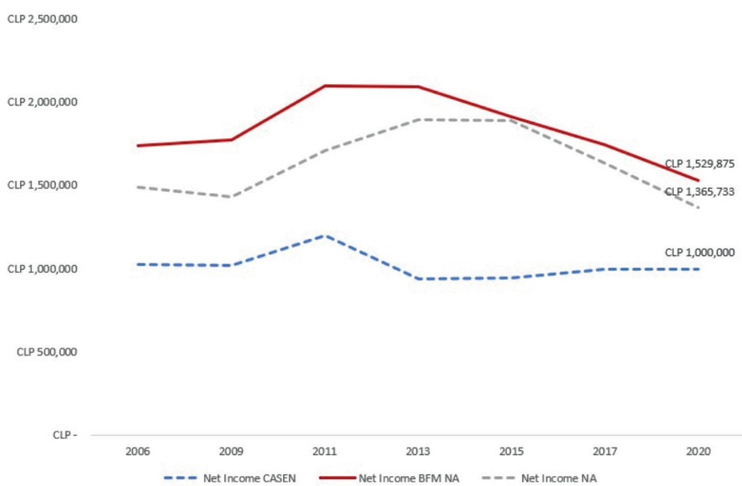
7.1 Main Results

A series of tables and graphs are now presented to show the estimated median income and the income of the 90*th* percentile for both dependent and independent workers. These statistics were chosen because they are common benchmarks for the middle- and upper-end of the distribution. The main occupation for each worker was used for categorization to ensure the categories are exclusive. It should be considered that each worker can also earn self-employed income despite being dependent on his main occupation and vice versa. Accordingly, another option is to report all workers who earn dependent income in one group and, all who earn independent income in another. The results do not change significantly whether the exclusive or the overlapping categorization are used. All results are in constant 2017 CLP for each year.

Figure 6 and Figure 7 show the evolution of three definitions of the net income of independent workers. Variables net of social security contributions are shown because of the high level of evasion. Gross variables are almost identical but not necessarily subject to additional error. The evolution of each definition of income is less regular than in the case of wage earners but the BFM correction has a similar effect. In 2017, the average gap was 63% but with the BFM correction it is distributed heterogeneously, reaching 75% for the 90*th* percentile.

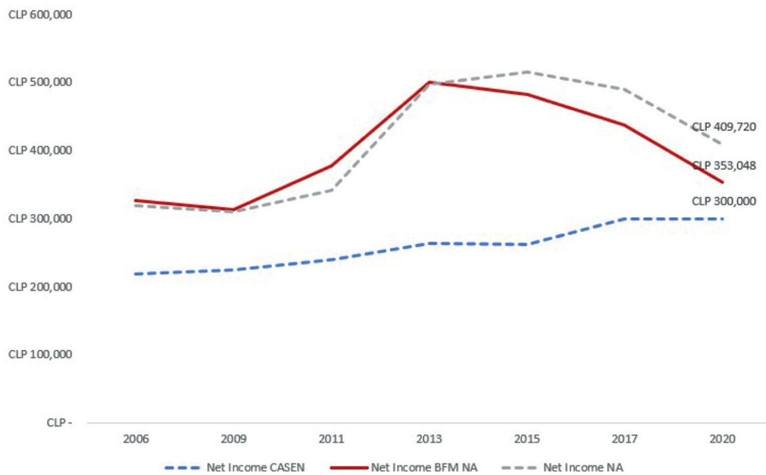
FIGURE 6

INCOME OF THE 90TH PERCENTILE OF INDEPENDENT WORKERS



Note: Own elaboration based on data from the CASEN survey, the Central Bank of Chile, and the SII. NA is national accounts. All figures are in 2017 CLP.

FIGURE 7
MEDIAN INCOME OF INDEPENDENT WORKERS



Note: Own elaboration based on data from the CASEN survey, the Central Bank of Chile, and the SII. All figures are in 2017 CLP.

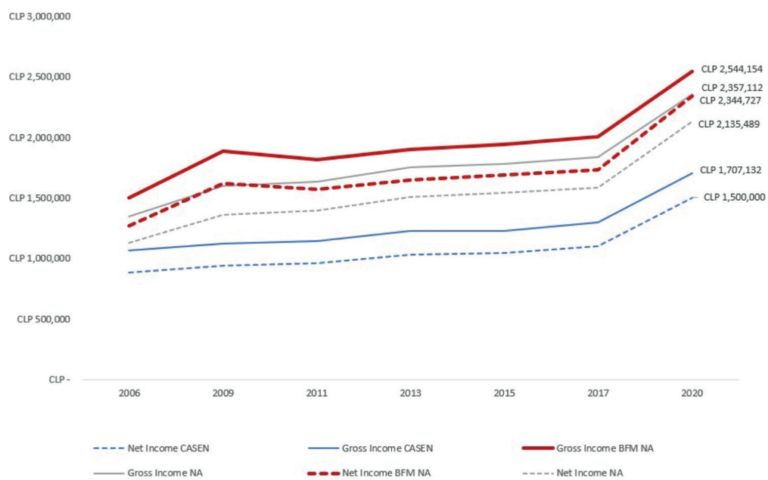
As mentioned, the income of independent workers (self-employed) is considerably less equal than dependent workers. This is because independent income is a heterogeneous category that not only includes informal workers but also professionals and unincorporated businesses with very diverse earnings. In addition, independent workers are more acutely affected by the economic cycle because they lack stabilizing mechanisms, such as an employment contract. After applying the corrections, the median income of independent workers decreased slightly between 2015 and 2017. This decrease is because the aggregate of equivalent national accounts grew less than the number of self-employed workers (7% versus 13%).

One hypothesis to explain the decrease in the average income between 2015 and 2017 is that the higher number of reported self-employed corresponds to people who started a low-paid activity. This idea is consistent with the fact that the income distribution share of the richest 1% of independents grew from 14.6% to 16.7% between these years. Income reported by independent workers is considered to be less precise due to the nature of their working lives and because the information used to make the correction is of lower quality. It is beyond the objective of this investigation to provide a better justification for these phenomena that affect estimates of independent worker income.

Figure 8 and Figure 9 show the main results for dependent workers. In this case it is possible to impute taxes and social security with greater precision, and the adjustment to national accounts is more reliable. The required adjustment for the 90*th* percentile is found to be greater than the required adjustment for the median when performing the BFM correction. For the 2017 CASEN survey, the adjustment was a total of 54% for gross income for the 90*th* percentile, and 33% for the median, which equates to a gross income of CLP 2,004,000 and CLP 598,000, respectively. For reference, the Gini coefficient obtained from autonomous income using the corrected series is 0.6, and the concentration of the richest 1% of total income is 19%. In the last subsection we will discuss the implications of our methodology to inequality estimations.

FIGURE 8

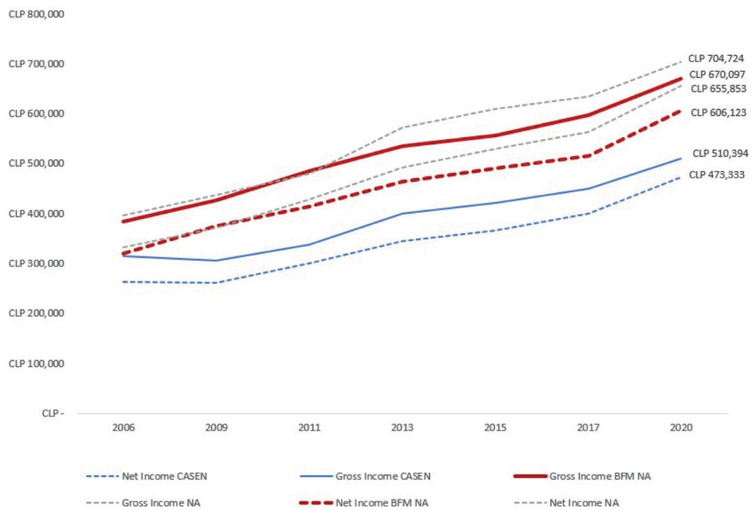
LABOR INCOME OF THE 90TH PERCENTILE OF DEPENDENT WORKERS



Note: Own elaboration based on data from the CASEN survey, the Central Bank of Chile, and the SII. All figures are in 2017 CLP.

FIGURE 9

MEDIAN LABOR INCOME OF DEPENDENT WORKERS

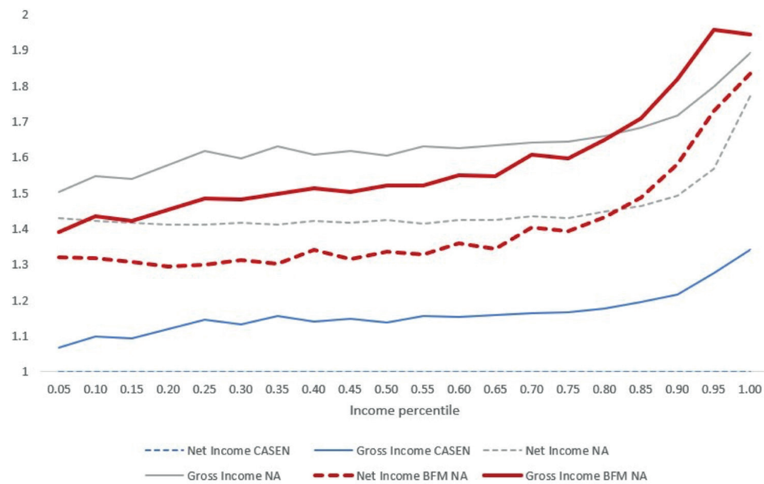


Note: Own elaboration based on data from the CASEN survey, the Central Bank of Chile, and the SII. All figures are in 2017 CLP.

In Figure 6 to Figure 9 it is also possible to see the effect of each adjustment. For example, in 2017, the median income of formal workers would be CLP 635,000 if the gap between the CASEN survey and national accounts were to be distributed proportionally. Considering the missing rich correction, gross median income would be close to CLP 600,000, which would equate to about CLP 515,000 net, discounting the partial payment of social security contributions. This net figure is 28% higher than the net income of CLP 400,000 reported in the 2017 CASEN survey, which is frequently quoted as the median income for workers in Chile.

Figure 10 shows the effect of each of the corrections at the percentile level for labor income of dependent workers. The original net income from the CASEN survey is normalized to one and the respective ratios are displayed. The ratios are increasing in the income percentile for two reasons. First, the progressivity of the income tax causes that the gap between net and gross variables is proportionally larger for higher incomes. Moreover, roughly 15% of workers report they do not have an employment contract, so their net income is equal to gross income. These informal workers are concentrated in the bottom of the distribution and have an average income around 50% smaller. Second, the missing rich re-weighting produces corrections that are larger than above the 80th percentiles and smaller below that percentile.

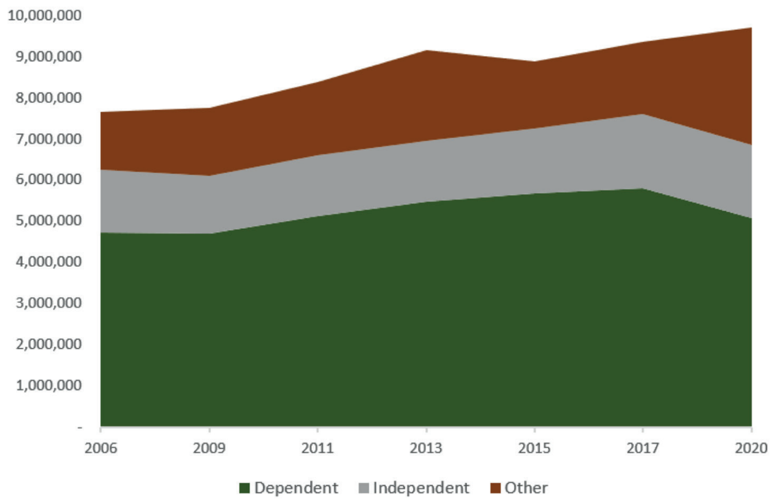
FIGURE 10
LABOR INCOME OF DEPENDENT WORKERS



Note: Own elaboration based on data from the CASEN survey, the Central Bank of Chile, and the SII. Each line represents the ratio of the corresponding definition of income for dependent workers with the original Net Income from CASEN.

It is interesting that between 2017 and 2020, dependent workers did not experience any fall in their average income in spite of the COVID-19 crisis. Moreover, the median wage evolved with a similar growth rate than in 2006 to 2017 and, the 90*th* percentile grew by nearly 30% in those three years. However, this significant increase has to be weighed against the decline of over 12% in the number of employed workers, as Figure 11 shows. Total real wages increased by barely 3% in those three years, but they were distributed unevenly to less workers. On the contrary, the number of independent workers decreased by just 1% but their median income corrected by national accounts fell by 9%. This is consistent with the hypothesis of counter-cyclical informal (proxied by independent) employment (Loayza and Rigolini, 2011).

FIGURE 11
NUMBER OF INCOME EARNERS BY TYPE



Note: Own elaboration based on data from the CASEN survey. Dependent and independent workers are defined using mutually exclusive categories. Other includes all people who receive some autonomous income such as interests, dividends and rentals.

The results for independent workers are presented separately to underline that the results are subject to greater uncertainty. It is, however, equally useful to describe workers in a single group. Different measures of workers' income are therefore shown in Table 5. Overall, the average net income of all workers is underreported by 53% compared to national accounts. Thus, the median gross income of workers, after correcting for the missing rich and national accounts, is 39% higher than in the original 2017 CASEN survey; it is CLP 570,000 or CLP 595,000 when the correction for high income earners is not applied.

7.2 Robustness

The four figures (Figure 6, Figure 7, Figure 8, Figure 9) and Table 5 illustrate what we consider the best estimates for the evolution of income of independent and dependent workers. However, results change when we change some key parameters, so we perform a sensitivity analysis of the assumptions for the year 2017. Specifically, it is difficult to achieve a high degree of certainty regarding capital income because the CASEN survey underestimates this component by up to 400% as compared to national accounts. Therefore, different combinations of the percentile from which capital income is corrected with tax data following BFM (p), and the magnitude of the imputed capital attributed to the highest income quin tile (k), were considered. The definitions of income presented in Table 6 are the same as in the previous figures and table.

On the basis of the proposed analysis, when looking at the extremes, the joint effect of the parameters *p* and *k* is up to 13% on wages and up to 24% on the income of independent workers. In the most conservative scenario, the gross median wage is CLP 555,000—that is, 23% higher than the gross median wage captured by the CASEN survey. At the highest point, the median wage reaches CLP 625,000 and is only 2% lower without the BFM adjustment. In Table 6 the central parameter combination, also used in the figures, is highlighted. This combination was chosen because it is considered to be a reasonable lower limit for the median income of each of the three groups in the table: dependent, independent, and all workers.

It follows that for the estimated median income of both independent and dependent workers there is a monotonic relationship with respect to the parameters *p* and *k*. The median is increasing with respect to *p* because the BFM correction increases the weight of fewer high-income observations, and this causes the gap and the consequent adjustment to national accounts to be greater. It is also increasing with respect to *k* because, when *k* is higher, the BFM-correction increases the weight of individuals with income from capital to the detriment of those with high salaries. Thus, before correcting for national accounts, there is a lower total wage bill and the adjustment factor is higher.

The effect of parameters *p* and *k* on the 90*th* percentile of both dependent and independent workers' income is somewhat less regular. The tendency is to be decreasing in *k* and *p*. The argument for *k* is analogue to the one presented for the median. When *k* is less, the corrected missing rich are found to have a greater share of labor income than if *k* is large, thus increasing the 90*th* percentile. With a higher *p*, simply a smaller number of individuals is adjusted from the upper part, so the effect is diluted for the 90*th* percentile.

As an additional robustness check, a higher inequality scenario than the one analyzed was also considered. The results for 2017 indicate that even with a Gini coefficient of 0.6 and top 1% income share of 19%, the result is a median gross income of CLP 570,000, which is 39% more than in the original 2017 CASEN survey. A yet more extreme inequality scenario, a Gini coefficient of 0.65 and a top 1% income share equal to 27%, results in a median gross income of CLP 500,000, which is still 23% higher than the 2017 CASEN survey. The Gini coefficients used in these two scenarios represent the highest inequality figures available in the literature for Chile, using the same definition of income. If capital gains and undistributed profits are included higher Gini coefficients and shares of 1% are calculated, but CASEN does not include the necessary data to perform an imputation.

Finally, Table 7 contains the independent income adjustment factors that result from including different fractions of the distributed income of companies (D.42) in their comparable aggregate in national accounts. All the values found for the independent workers decisively depend on this fraction[37].

7.3 Implications For Inequality

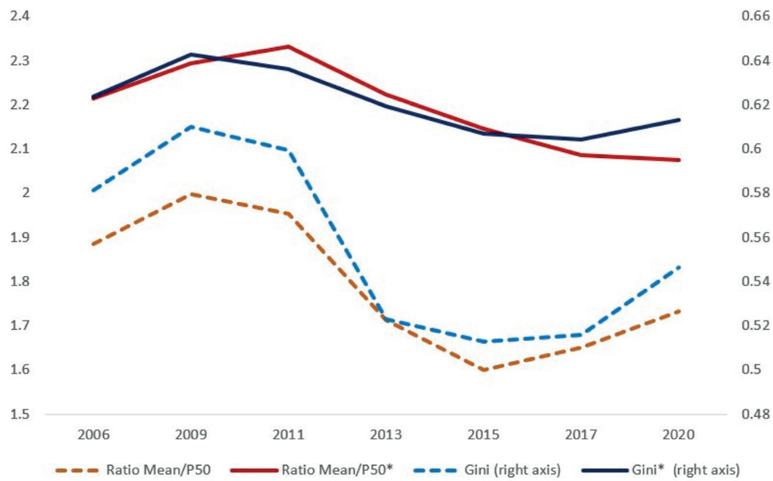
The methodology and results we develop are designed to measure income accurately and reliably especially for the middle of the distribution. However, it is constructive to analyze the impact of all adjustments on the level and trend of inequality indicators. In addition to the standard Gini coefficient, we consider useful giving attention to the mean-median ratio as simpler inequality indicator. Figure 12 illustrates that the reduction of inequality between 2006 and 2017 is robust in our computation with the trend observed in CASEN. However, this reduction is smaller than the original survey. Between 2006 and 2017, the uncorrected mean-median ratio of gross income falls from 1.89 to 1.65 (12%) while the Gini coefficient did so from 0.58 to 0.52 (11%). Considering corrections, the ratio only falls from 2.21 to 2.09 with the Gini coefficient going from 0.62 to 0.6. Looking at 2017, the correction causes a 26% and 17% increase in the ratio and Gini, respectively. The rapid decline in inequality observed after 2011 is substantially offset when including top income and national accounts corrections. In 2020 there was a partial reversal in the decline in inequality as COVID-19 had a heterogeneous impact across households and workers, related to the ability to work from home, age, gender and other characteristics.

---

[37] See **Section 4.2**. for details on the construction of the comparable aggregate of national accounts for independent workers.

FIGURE 12

GINI AND MEAN-MEDIAN RATIO OF GROSS INCOME



Note: Own elaboration based on data from the CASEN survey, the Central Bank of Chile, and the SII.
* Asterisks and solid lines denote series corrected for missing rich and adjusted to match national accounts. Dashed lines correspond to CASEN survey income with simulated taxes and social security.

Besides its simplicity, the ratio of the mean to the median provides a reasonably good indicator of a country's income distribution (Birdsall and Meyer, 2015). Gini has more robust theoretical properties, as, for example, the ratio fails the Pigou-Dalton transfer principle.[38] However, the mean-median ratio is easier to understand and, in usual distributions (such as the lognormal and Weibull) is associated directly to inequality. Another benefit in our context is that it is not sensitive to changes in the share of the top 1% that hold the average, unlike the Gini coefficient. We emphasize that the Gini coefficient is more sensitive to the assumptions made for allocating capital income and to the top of the distribution in general.

---

38 Namely that a transfer of income from a richer to a poorer person, so long as that transfer does not reverse the ranking of the two, will result in greater equity.

8. Concluding Remarks

This is the first study in Chile to compare and combine administrative records, national accounts, and surveys in a consistent theoretical framework to obtain estimates of dependent and independent worker income in the middle of the distribution. By harnessing the advantages and correcting the limitations of these data sources it is possible to generate better estimates than when using each source separately. The methodology serves as a basis to implement distributional national accounts in countries with similar data quality limitations.

We find that the average gap between income in the CASEN survey and national accounts is larger for independent workers than it is for dependent workers, 59% and 40% respectively. The missing rich phenomenon explains 10% of this gap for dependent income and 5% for independent income. Considering this, in the main scenario for 2017, the median gross income was approximately CLP 600,000 for dependent workers (employees) and CLP 440,000 for independent workers (self-employed). This equates to a joint median gross income of CLP 570,000. In each scenario analyzed, between 2006 and 2017, the median real wage grew by between 4% and 5% annually, which is 2% more than both the income of the 90*th* percentile and real GDP per capita. Between 2017 and 2020, the median real wage had a similar growth and the 90*th* percentile grew by nearly 10% but the number of active employees declined by 12%. The trend in our inequality estimations fall between Fairfield and Jorratt (2016) (flat) and official data from the world bank (important fall) with a reduction from 0.62 in 2006 to 0.60 in 2017 instead of 0.58 to 0.52 in the CASEN survey with imputed taxes.

The methodology applied in this study addresses the fact that underreporting of income is particularly prevalent among high income earners, although the practice is present throughout the whole distribution. The tax and social security simulation performed generates revenues similar to the real administrative records. Social security contributions are realistically imputed and scaled up to match administrative records, resulting in greater reliability. To account for the numerous assumptions still required to get precise results, a sensitivity analysis was proposed and implemented to consider deviations from the central estimations.

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. A larger income, whether 40% more or even 20% more, significantly impacts the standard of living for low- and middle-income families. Despite the progress made in this investigation, the data series that have been presented here should be understood as prototypes for which it is possible and desirable to incorporate improvements. The same is also true for the national accounts and other sources of information. Our results depend on some relatively arbitrary parameters; however, there are four dimensions where it is possible to improve the methodology used if more data were available. First, precision for independent (self-employed) workers would be substantially improved if primary source data were available for mixed income, separated from household operating surplus and quasi-corporation income. Second, the imputation used for capital income is unrealistic and is only functional when studying income close to the median. Third, the point at which survey data are mixed with tax data—that is, the merging point of the two data sources—can be determined endogenously by considering the two types of biases present, as discussed in the methodology section. Finally, the simulation of the tax structure can also be adjusted by modeling tax evasion in a more sophisticated way or, ideally, directly imputed by matching individuals from the survey data to the tax data.

The question of whether economic growth experienced by Chile has benefited different socioeconomic groups and to which extent is a controversial issue. This paper sheds light on the matter by arguing that median income is much higher than what CASEN shows, and which is used in popular media. There is evidence also that inequality has declined, but to a lesser extent than in CASEN data. Whether this is enough or not is a value judgment, but the discussion needs to be grounded in sound evidence making clear the data limitations. Beyond these considerations, Chile is still a country with relatively high levels of inequality. Future research should aim at a more detailed treatment of top income earners and impute a distribution to the excluded components of GNI, such as direct taxes and withheld capital income, to improve comparability with other countries. Future stages of this project will aim to incorporate tax microdata and unpublished national accounts.

References

Altimir, O. (1986). Estimaciones de la distribución del ingreso en la Argentina, 1953-1980. *Desarrollo económico*, 521–566.

Altimir, O. (1987). Income distribution statistics in Latin America and their reliability. *Review of income and wealth*, Vol. 33(2), 111–155.

Alvaredo, F., Atkinson, A. B., Blanchet, T., Chancel, L., Bauluz, L., Fisher-Post, M., Flores, I., Garbinti, B., Goupille-Lebret, J., Martínez-Toledano, C., et al. (2020). Distributional National Accounts Guidelines, Methods and Concepts Used in the World Inequality Database.

Atkinson, A. B. (2017). Monitoring global poverty: Report of the commission on global poverty. World Bank, Washington.

Atria, J., Flores, I., Sanhueza, C., Mayer, R., et al. (2018). Top incomes in Chile: a historical perspective of income inequality (1964-2015).

Bandourian, R., McDonald, J., and Turley, R. S. (2002). A comparison of parametric models of income distribution across countries and over time.

Birdsall, N. and Meyer, C. J. (2015). The median is the message: A good enough measure of material wellbeing and shared development progress. *Global Policy*, Vol. 6(4), 343–357.

Blanchet, T., Flores, I., and Morgan, M. (2022a). The weight of the rich: improving surveys using tax data. *The Journal of Economic Inequality*, Vol. 20(1), 119–150.

Blanchet, T., Fournier, J., and Piketty, T. (2022b). Generalized pareto curves: theory and applications. *Review of Income and Wealth*, Vol. 68(1), 263–288.

Bourguignon, F. (2011). Non-anonymous growth incidence curves, income mobility and social welfare dominance. *The Journal of Economic Inequality*, Vol. 9(4), 605–627.

Bravo, D. and Valderrama, J. A. (2011). The impact of income adjustments in the Casen Survey on the measurement of inequality in Chile. *Estudios de Economía*, Vol. 38(1), 43–65.

Campos, J. A., Foster, W., et al. (2013). Medición de la Pobreza. Consecuencias de compatibilizar ingresos de encuestas de hogares con cuentas nacionales. *Estudios Públicos*, 130.

Candia, B. (2018). Taxes, Transfers and Income Distribution in Chile.

Central Bank of Chile. (2017). CUENTAS NACIONALES DE CHILE: Métodos y Fuentes de Información. Technical report.

Chotikapanich, D., Valenzuela, R., and Rao, D. P. (1997). Global and regional inequality in the distribution of income: estimation with limited and incomplete data. *Empirical Economics*, Vol. 22(4), 533–546.

De Rosa, M., Flores, I., and Morgan, M. (2020). Inequality in Latin America revisited: insights from distributional national accounts. Technical Note, (2020/02).

Deaton, A. and Dreze, J. (2002). Poverty and inequality in India: a re-examination. *Economic and political weekly*, 3729–3748.

Del Castillo, M. (2015). La magnitud de la desigualdad en el ingreso y la riqueza en México: una propuesta de cálculo.

Dowrick, S. and Akmal, M. (2005). Contradictory trends in global income inequality: A tale of two biases. *Review of Income and Wealth*, Vol. 51(2), 201–229.

Eurostat (2018). Methodological note Comparison of household income: European Union Statistics on Income and Living Conditions and National Accounts.

Fairfield, T. and Jorratt, M. (2016). Top income shares, business profits, and effective tax rates in contemporary Chile. *Review of Income and Wealth*, Vol. 62, S120–S144.

Ffrench-Davis, R., Vivanco, D., et al. (2016). Depreciación del capital físico, inversión neta y producto interno neto. Santiago, Corporación de Estudios para Latinoamérica (CIEPLAN)/Universidad de Talca, Santiago, noviembre.

Fitoussi, J. P., Durand, M., and Stilitz, J. (2018). Beyond GDP measuring what counts for economic and social performance: measuring what counts for economic and social performance. OECD Publishing.

Guerriero, M. (2019). The labor share of income around the world: Evidence from a panel dataset.

Hurst, E., Li, G., and Pugsley, B. (2014). Are household surveys like tax forms? Evidence from income underreporting of the self-employed. *Review of economics and statistics*, Vol. 96(1), 19–33.

Hussmanns, R. (2004). Measuring the informal economy: From employment in the informal sector to informal employment. Policy Integration Department, Bureau of Statistics, International Labour Office.

INE, I. N. d. E. (2018). Informe de Principales Resultados VIII Encuesta de Presupuestos Familiares (EPF). Technical report.

INE, I. N. d. E. (2019). Encuesta Suplementaria de Ingresos. Technical report.

Lakner, C. and Milanovic, B. (2016). Global income distribution: from the fall of the Berlin Wall to the Great Recession. *The World Bank Economic Review*, Vol. 30(2), 203–232.

Larrañaga, O. and Rodríguez, M. E. (2014). Desigualdad de ingresos y pobreza en Chile, 1990 a 2013 (Documento de Trabajo), Santiago: PNUD.

Loayza, N. V. and Rigolini, J. (2011). Informal employment: safety net or growth engine? *World Development*, Vol. 39(9), 1503–1515.

López, R., Figueroa, E., and Gutiérrez, P. (2013). La "parte del león": Nuevas estimaciones de la participación de los súper ricos en el ingreso de Chile.

López, R. E., Figueroa, E., and Gutiérrez, P. (2016). Fundamental accrued capital gains and the measurement of top incomes: An application to Chile. *The Journal of Economic Inequality*, Vol. 14(4), 379–394.

Lubrano, M. (2017). Lorenz curves, the Gini coefficient and parametric distributions. The Econometrics of Inequality and Poverty.

Lustig, N. et al. (2020). The "missing Rich" in Household Surveys: Causes and Correction Approaches. Technical report, ECINEQ, Society for the Study of Economic Inequality.

Milanovic, B. (2012). Global income inequality by the numbers: in history and now–an overview–. World Bank Policy Research Working Paper, (6259).

Moore, J. C., Stinson, L. L., and Welniak, E. J. (2000). Income measurement error in surveys: A review. *Journal of Official Statistics-Stockholm*, Vol. 16(4), 331–362.

Morgan, M. (2017). Extreme and persistent inequality: new evidence for Brazil combining national accounts, surveys and fiscal data, 2001-2015.

Piketty, T., Yang, L., and Zucman, G. (2019). Capital accumulation, private property, and rising inequality in China, 1978–2015. *American Economic Review*, Vol. 109(7), 2469–96.

Pinkovskiy, M. and Sala-i-Martin, X. (2009). Parametric estimations of the world distribution of income. Technical report, *National Bureau of Economic Research*.

Pinkovskiy, M. and Sala-i-Martin, X. (2016). Lights, camera… income! Illuminating the national accounts-household surveys debate. *The Quarterly Journal of Economics*, Vol. 131(2), 579–631.

Ravallion, M. (2000). Should poverty measures be anchored to the national accounts? *Economic and Political Weekly*, 3245–3252.

Ravallion, M. (2012). Benchmarking global poverty reduction. *World Bank Policy Research Working Paper*, (6205).

Ravallion, M. (2016). Toward better global poverty measures. *The Journal of Economic Inequality*, Vol. 14(2), 227–248.

Ruiz, N. and Woloszko, N. (2016). What do household surveys suggest about the top 1% incomes and inequality in OECD countries?

Szekely, M. and Hilgert, M. (1999). What's Behind the Inequality We Measure? An Investigation Using Latin American Data. An Investigation Using Latin American Data (December 1999). *IDB-OCE Working Paper*, (409).

Szekely, M., Lustig, N., Cumpa, M., and Mejía, J. A. (2000). Do we know how much poverty there is?

UNDP (2017). Desiguales. Orígenes, cambios y desafíos de la brecha social en Chile. UNECE (2011). *Handbook on household income statistics*.

United Nations (2008). System of National Accounts 2008. Technical report.

Zwijnenburg, J., Bournot, S., and Giovannelli, F. (2017). Expert group on disparities in a national accounts framework: Results from the 2015 exercise.

9. Tables

TABLE 1

AVERAGE MONTHLY NET INCOME (CLP) ACCORDING TO CASEN SURVEY (C) AND SII (I), 2017

2017	Average	P50	P90	N	Share 1%
Wages of dep. workers (C)	CLP 589,397	CLP 400,000	CLP 1,101,667	5,793,314	8.9%
Total wages (C)	CLP 564,112	CLP 378,333	CLP 1,027,500	6,011,892	9.0%
Income of ind. workers (C)	CLP 493,262	CLP 285,623	CLP 1,000,000	1,803,263	16.5%
Total independent income (C)	CLP 477,361	CLP 274,923	CLP 1,000,000	2,282,220	16.7%
Net taxable income (C)	CLP 529,451	CLP 335,000	CLP 1,000,500	9,339,690	11.3%
Net taxable income (I)	CLP 564,169	CLP 299.118*	CLP 1,286,763	9,941,065	13.1%

Note: Own elaboration based on data from CASEN survey, and SII for the last row.
*Calculated from a generalized Pareto interpolation with tabulated data from P75. P50 and P90 denote the corresponding percentiles so that P50 equals the median. N corresponds to the number of individuals represented by the CASEN survey using the standard (regional) expansion factor. The difference between the number of individuals in the first column and the second column, or the third and fourth, is due to the fact that some individuals that receive salaries despite being independent in their main occupation, and vice versa.

TABLE 2

NOMINAL MONTHLY INCOME PER CAPITA IN NATIONAL ACCOUNTS AND CASEN SURVEY (CLP)

	2006	**2009**	**2011**	**2013**	**2015**	**2017**
1. Total household income CASEN (CLP)	185,609	226,351	254,549	324,769	371,638	433,577
2. Subtotal CASEN and NA complements*(CLP)	329,935	407,943	479,009	566,054	662,850	743,372
3. GNI (CLP)	381,558	456,637	570,134	641,966	741,037	813,582
1. : 3. (Ratio HS to NA)	0.49	0.50	0.45	0.51	0.50	0.53
(3. - 2.) : 1. (HS to NA gap/total HS income)	0.28	0.22	0.36	0.23	0.21	0.16

Note: Central Bank of Chile and CASEN Surveys. GNI corresponds to the annual Gross National Income divided into twelve from National Accounts.
*The online appendix provides details on the components and magnitudes of this line.

TABLE 3

NATIONAL ACCOUNTS ADJUSTMENT FACTORS, WITH AND WITHOUT BFM CORRECTION

	2006	**2009**	**2011**	**2013**	**2015**	**2017**	**2020**
Wages	1.26	1.43	1.43	1.43	1.44	1.41	1.38
Wages BFM	1.22	1.34	1.38	1.34	1.30	1.29	1.21
Self-employed income	1.46	1.35	1.42	1.85	1.82	1.63	1.37
Self-employed income BFM	1.49	1.33	1.57	1.76	1.64	1.46	1.18

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

TABLE 4

RATIO SIMULATIONS TO ADMINISTRATIVE RECORDS

	2006	**2009**	**2011**	**2013**	**2015**	**2017**	**2020**
Tax revenue IGC BFM	1.18	0.94	0.94	0.84	0.83	0.91	0.91
Social Security Evasion BFM	0.24	0.21	0.35	0.36	0.38	0.35	0.55

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

TABLE 5

LABOR INCOME OF ACTIVE DEPENDENT AND INDEPENDENT WORKERS, 2017

	Mean (CLP)	Median (CLP)	P90 (CLP)
Net Income CASEN	579,357	370,167	1,090,000
Gross Income CASEN	664,149	408,615	1,252,511
Net Income BFM NA	886,833	511,861	1,768,609
Gross Income BFM NA	967,029	568,414	2,004,139
Net Income NA	884,302	544,119	1,633,849
Gross Income NA	966,826	593,956	1,828,474

Note: Own elaboration based on CASEN, SII and Central Bank. BFM stands for the Blanchet et al. (2022a) correction and NA stands for National Accounts corrections. All figures are in 2017 CLP.

TABLE 6

SENSIBILITY ANALYSIS FOR GROSS INCOME, 2017

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

Note: Own elaboration based on data from the CASEN survey, the Central Bank of Chile, and the SII. The parameter p corresponds to the percentile from which the BFM correction is applied, and k to the proportion of capital attributed to the highest income quintile. For a more detailed description see the methodology section.
1: CASEN estimates presented in the first row include taxes and imputed contributions as well as all rows.
2: Last row shows the estimates without correcting for missing rich and adjusting for national accounts. All figures are in CLP.

TABLE 7

INDEPENDENT INCOME ADJUSTMENT FACTORS FOR DIFFERENT FRACTIONS OF D.42

%D.42	2006	2009	2011	2013	2015	2017	2020
0%	1.15	1.00	1.01	1.33	1.30	1.19	1.08
20%	**1.46**	**1.34**	**1.42**	**1.85**	**1.82**	**1.63**	**1.36**
40%	1.77	1.69	1.83	2.38	2.34	2.08	1.65
60%	2.08	2.03	2.24	2.91	2.86	2.52	1.94
80%	2.39	2.38	2.65	3.44	3.38	2.97	2.22
100%	2.70	2.72	3.06	3.96	3.90	3.42	2.51

Note: Own calculations based on the Central Bank and CASEN. D.42 corresponds to the distributed income of corporations (D.421) and quasi-corporations (D.422).

Appendix

10. Parametric Estimation Of Median Income

This section details the theoretical foundations used to elaborate Figure 1 and the results presented in the introduction. We propose a simple exercise is to illustrate the implications of the gap between numbers reported by national accounts and by CASEN. The exercise consists of imputing a parametric distribution to the aggregate income of national accounts, assuming a given level of inequality, and comparing the result with the distribution actually captured in the 2017 CASEN survey.[39] Using a parametric distribution, it is possible to obtain an approximation of median income from labor and capital that is consistent with the mean of national accounts at different levels of inequality. Given an average income, greater inequality (represented by the Gini coefficient) is necessary for a lower median, leaving the type of distribution constant. This exercise is an extremely simplified version of DINA that uses only two variables and a fixed distribution as inputs. This is only an exercise to show inconsistencies because to have a better assessment DINA uses many variables and sources to impute a micro distribution to macroaggregates in a much more sophisticated and precise way.

Figure 1 shows the gross median income for Chile calculated using a parametric distribution, different Gini coefficients and the average national income as reported by national accounts.[40] It also displays the median of the income from labor and capital as reported in the 2017 CASEN survey as comparison (in red). According to national accounts, the average national income, for 8.2 million factor income earners, is close to CLP 1,350,000 per month, and the GNI components not included, indirect taxes and capital depreciation, amount to about CLP 350,000. To combine this parameters, Pinkovskiy and Sala-i Martin (2009) argue that the lognormal and Weibull distributions are the best choice to fit to empirical data. We select the Weibull distribution because it has lower medians for a given mean and the displayed Gini coefficients.[41]

The true or empirical distributions of a country's income have different degrees of fit to known probability distribution functions. For the purposes of this exercise, only two parameter distributions are used, as they are the most commonly used type.

39 A similar endeavor, on a global scale, is undertaken in Pinkovskiy and Sala-i Martin (2009) using average income from national accounts and different parametric distributions.

40 The comparable income of national accounts equals GNI minus capital depreciation and indirect taxes, which is equivalent to the factor compensation of capital and labor, net of depreciation and indirect taxes.

41 Although this relationship is not monotonic and does not hold outside the interval of Gini coefficients we display.

In past research, the Weibull distribution was found to be the best fit for data from OECD countries (Bandourian et al., 2002). In Pinkovskiy and Sala-i Martin (2009), a larger sample of 191 countries is considered and they find that the lognormal is generally more precise. An additional benefit is that there are analytical expressions of these distributions for all statistics of interest, which clarifies the analysis.

Table 8 shows the density, mean, median, and Gini function in terms of the parameters k, α, σ and μ, and of Γ and Φ, which correspond to the cumulative standard normal and gamma functions respectively. These two distributions have the property of being closed for multiplication—that is, $X \sim f(x) \rightarrow cX \sim \mathrm{f}(x)$—which is attractive if you want to make a proportional fit. The Gini coefficient (for any distribution) is also insensitive to scale.

TABLE 8
DENSITY FUNCTION AND STATISTICS FOR WEIBULL AND LOGNORMAL DISTRIBUTIONS

	Weibull	**Lognormal**
Density function	$k\alpha\left(kx\right)^{\alpha-1}\left(e^{-kx}\right)^{\alpha}$	$\left(x\sigma\sqrt{2\pi}\right)^{-1} exp\left(-\frac{\left(ln\left(x\right)-\mu\right)^{2}}{2\sigma^{2}}\right)$
Mean	$\Gamma\left(1+1/\alpha\right)$	$\exp\left(\mu+\frac{\sigma^{2}}{2}\right)$
Median	$\log\left(2\right)^{\{1/\pm\}}$ / k	$\exp\left(\mu\right)$
Gini	$1-2^{\left(-\frac{1}{\alpha}\right)}$	$2\Phi\left(\sigma/\surd 2-1\right)$

Note: Own elaboration based on Lubrano (2017).

Regarding the information in Table 8, if two of these statistics are known, it is possible to clear all the parameters of the distribution and, consequently, obtain the missing statistic. The median calculated in this way complies with being monotonically increasing in the mean and monotonically decreasing in the Gini coefficient, both good properties. It is desirable to perform goodness-of-fit measures to find the most appropriate distribution for the population's income.

TABLE 9

PARAMETRIC MEDIAN FROM MONTHLY GNI PER WORKER 2017

Gini	Median1	Median2
0.45	CLP 893,515	CLP 980,653
0.5	CLP 810,413	CLP 885,336
0.55	CLP 721,858	CLP 779,589
0.6	CLP 629,014	CLP 665,474
0.65	CLP 533,267	CLP 545,873
0.7	CLP 436,281	CLP 424,657

Note: Own calculations based on data from the Central Bank and Ffrench-Davis et al. (2016) using Table 8. Average monthly income corresponds to GNI less indirect taxes and capital depreciation. This amount was divided by 12 to obtain monthly data. Median1 assumes a Weibull distribution, and Median2 assumes a lognormal distribution. All monetary figures are in CLP.

Table 9 shows the median calculated for different Gini coefficient values, leaving the mean fixed. For the average, GNI is used, discounting indirect taxes and 12.3% of GNI for capital depreciation. The value for depreciation is obtained from Table 4 in Ffrench-Davis et al. (2016), averaging the depreciation calculated for the 2011–2015 interval, using 2008 prices. Total income is divided by the number labor and capital income earners. In other words, all the calculations consider that the population consist of all individuals with an autonomous income stricter than zero.[42]

The rationale for discounts made the income is that most micro definitions of income do not include income that is used to finance capital depreciation. For example, according to the SNA 2008, income is defined as the maximum amount that a household or other entity is capable of consuming in goods and services without reducing its stock of assets or increasing its financial and non-financial liabilities. It is reasonable to assume that a very small fraction of the income reported in surveys (including earnings, withdrawals, and other income) is used to later cover capital depreciation.

42 The chosen distributions are defined for positive values, and the probability defined for x = 0 is identical to zero. In order to include values equal to zero, censored distributions must be used, which complicates the analysis unnecessarily.