

# Estimando la Desigualdad Salarial en Chile: Un enfoque con datos administrativos

Febrero, 2022

- ▶ La desigualdad se ha vuelto un tópico de creciente interés en la política pública.
- ▶ Sin embargo, esta se ha calculado tradicionalmente a través de encuestas que implican ciertos problemas al momento de obtener resultados.
  - ▶ Proporción significativa de *income non-response* (Moore *et al.*, 2000).
  - ▶ Subrepresentación de individuos de alto ingreso (Atkinson & Piketty, 2007).
  - ▶ Errores asociados al reporte del ingreso que se correlacionan fuertemente con las variables bajo estudio (Bound *et al.*, 2001).
  - ▶ Baja frecuencia entre cada encuesta.

- ▶ Larrañaga *et al.* (2021) propone el estudio de la desigualdad mediante la corrección del uso de la CASEN a través de la imputación de distintas fuentes.
- ▶ En este espíritu, se propone construir indicadores a partir de la base de datos del Seguro de Cesantía; un insumo actualmente subutilizado.
  - ▶ Esta ofrece información de naturaleza administrativa y frecuencia mensual.

## II. Sobre los datos

- ▶ La base observa al total de afiliados al Seguro y excluye a:
  1. Trabajadores menores de 18 años.
  2. Trabajadores de casa particular.
  3. Pensionados.
  4. Trabajadores independientes o por cuenta propia.
  5. Funcionarios del sector público.
- ▶ La base incluye a todo aquel que en algún momento ha estado afiliado al seguro.

## II. Sobre los datos

- ▶ A junio de 2020 la base tiene las siguientes características:
  1. Mujeres representan el 39,3%.
  2. La edad promedio es de 39 años.
  3. El ingreso imponible medio es de \$834.519.
  4. El ingreso mediano es de \$584.990.
- ▶ La base incluye a todo aquel que en algún momento ha estado afiliado al seguro.

### III. El ejercicio

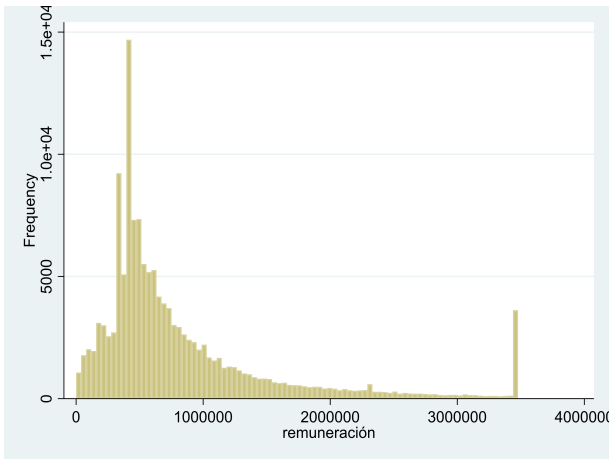
- ▶ Lograr retratar la desigualdad salarial de trabajadores afiliados al Seguro.
- ▶ Una derivada de lo anterior es incluir a quienes hoy no se encuentran afiliados pero en algún momento sí lo estuvieron.
  - ▶ Dado que quienes no cotizan dejan de observarse, esto requiere una rectangularización de la
- ▶ Para los siguientes ejercicios se consideró el periodo entre enero 2017 y junio 2020, último mes de la muestra disponible en la página web de la Superintendencia de Pensiones.
- ▶ Por motivos computacionales, el cálculo se realizó con bloques de seis meses cada uno.
  - ▶ Dado que la rectangularización considera solo aquellos individuos presentes en por lo menos un mes del bloque, este método disminuye el número de no cotizantes.
- ▶ Se trabajó con una muestra del 3% del total. En promedio, esto significa 883.000 observaciones mensuales.

## IV. Desafíos: Tope imponible

- ▶ Salarios que superen el tope imponible de cotización se encuentran censurados en este monto.
- ▶ A junio 2020 el tope fue de \$3.455.140
- ▶ En todo el periodo analizado, los datos censurados representan al 2,63% del total.

## IV. Desafíos: Tope imponible

La existencia del tope imponible significa que, a junio de 2020, la distribución de ingresos era la siguiente:





## IV. Desafíos: Tope imponible

- ▶ No encargarse de este problema llevaría a una subestimación de la desigualdad.
- ▶ Una solución es ajustar una distribución paramétrica a través de un modelo que considere datos censurados al estilo de Jenkins *et al.* (2009).
  - ▶ Es decir, replicar teóricamente la distribución real del ingreso sin tope imponible.
- ▶ Una dificultad es encontrar un método que logre hacer el ajuste considerando datos censurados.

## IV. Desafíos: Tope imponible

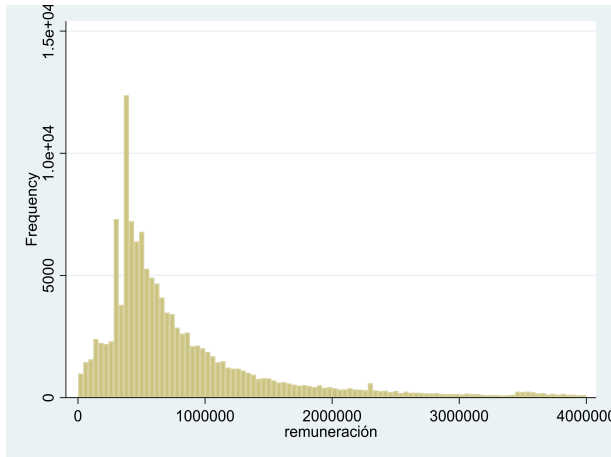
- ▶ Para ello se utilizó el paquete *SurPyval* de Python, ajustando una distribución *Weibull* de la forma:
- ▶  $f(x) = \lambda\alpha(\lambda\alpha)^{\alpha-1}e^{-(\lambda\alpha)^\alpha}$
- ▶ McDonald (1984) y Ripsy *et al.* (2002) comparan la utilidad de distintas distribuciones para replicar distribuciones de ingreso.
- ▶ “*Among the two-parameter models, the Weibull provides the best fit in sixty two percent of the cases. . . .*” (Ripsy *et al.*, 2002).
- ▶ Esto no quita que puedan existir distribuciones más adecuadas (y complejas).

## IV. Desafíos: Tope imponible

- ▶ El método para imputar valores es:
  - ▶ De la distribución, se toma un valor aleatorio.
  - ▶ Si este es menor al valor del tope imponible, se descarta y se toma un nuevo valor.
  - ▶ Si el valor es mayor al tope imponible, se le asigna al ingreso de una observación que tenga tope activo.

## IV. Desafíos: Tope imponible

Aplicado el método, la distribución de ingresos toma la siguiente forma:



## IV. Desafíos: Individuos no cotizantes

- ▶ Solucionado el problema del tope imponible, ya es posible calcular el índice de trabajadores cotizantes.
- ▶ Sin embargo, una derivada interesante es incorporar a aquellos trabajadores que alguna vez se afiliaron al seguro pero que no se encuentran cotizando por estar empleados de manera informal o por cuenta propia.
- ▶ Esto permitiría un cálculo más realista de la desigualdad salarial en Chile.

## IV. Desafíos: Individuos no cotizantes

- ▶ Es necesario buscar un método de imputación de datos: *Heckman*, *hotdeck*, OLS estándar, OLS aleatorios...
- ▶ Sin embargo, el patrón de *missings* no es ignorable, dado que son todos aquellos individuos no cotizantes. Con ello, cualquier método de imputación a partir de la misma base producirá resultados sesgados.
- ▶ Para corregir el sesgo, se necesita imputar a partir de una base que retrate ingresos de individuos no cotizantes: ESI

## IV. Desafíos: Individuos no cotizantes

- ▶ Corregido este problema, es necesario elegir el método de imputación más adecuado para el cálculo de la desigualdad; particularmente uno que no afecte la varianza: *Hotdeck*
- ▶ “*Correction methods such as the hot-deck, the two-step regression or the random OLS model perform better as they introduce relatively more variation in imputed incomes than, for instance, the standard OLS model. . . hot-deck, for instance, produces the most accurate (Gini) estimations in all cases.*” (Ripsy et al., 2002)

#### IV. Desafíos: Individuos no cotizantes

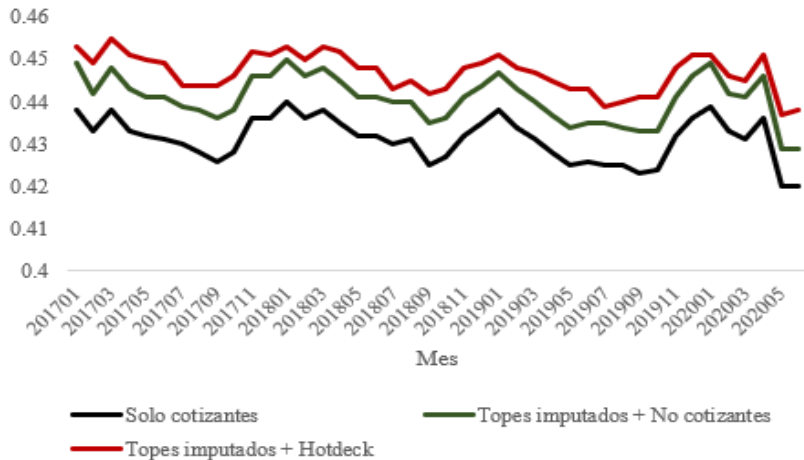
- ▶ El método consiste en imputar valores aleatorios de acuerdo a sub grupos.
  - ▶ Los subgrupos se conforman con base en sexo y grupos etarios.
- ▶ Por ejemplo, si en la base se tiene una mujer no cotizante de entre 18 y 25 años, se le imputará aleatoriamente el ingreso de una mujer -no cotizante- de entre 18 y 25 años a partir de la ESI.
- ▶ Naturalmente, el resultado involucra imputar ingreso \$0 en algunos casos. Dado que el objetivo es el cálculo de desigualdad de la población activa, estos son posteriormente descartados.
- ▶ Entre enero y junio de 2020, los no cotizantes equivalieron al 16% del total. De este 16



- ▶ Se presentan los resultados considerando tres escenarios:
  - ▶ Considerando solo cotizantes, sin imputación de topes.
  - ▶ Considerando solo cotizante, imputando los topes.
  - ▶ Considerando cotizantes con topes imputados y no cotizantes -teóricamente- ocupados.

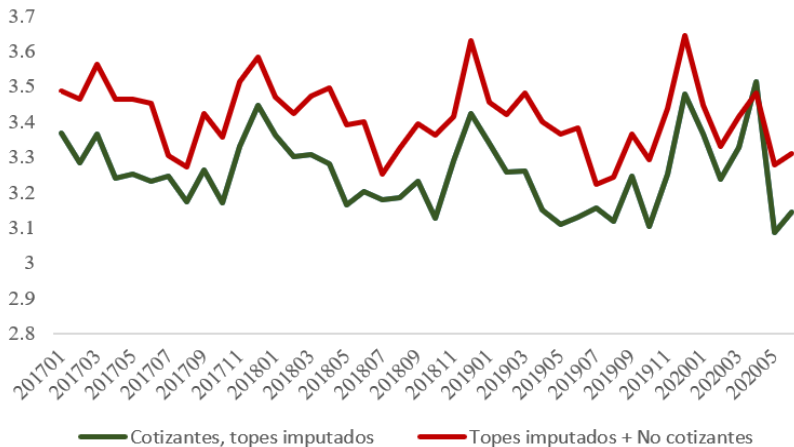
## V. Resultados: Índice Gini

≡ HORIZONTAL



## V. Resultados: Razón interpercentil p80/p20<sup>1</sup>

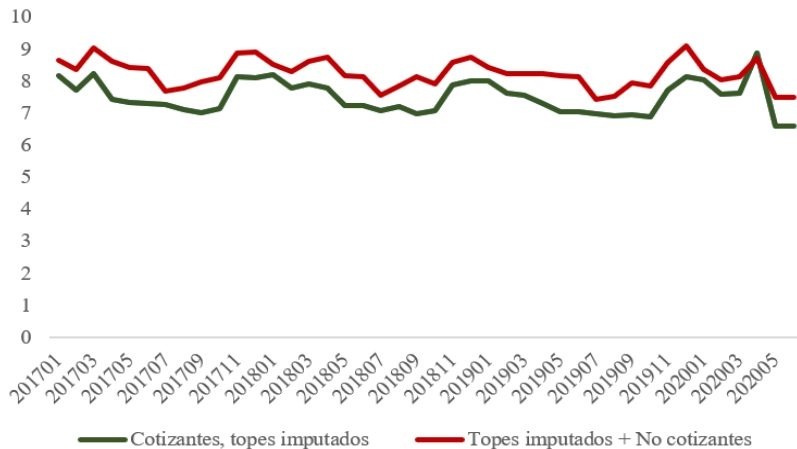
≡ HORIZONTAL



<sup>1</sup>Dado que valores con topes imputados pertenecen en general al 3% más rico, no existe diferencia en el resultando considerando cotizantes con topes con y sin imputar.

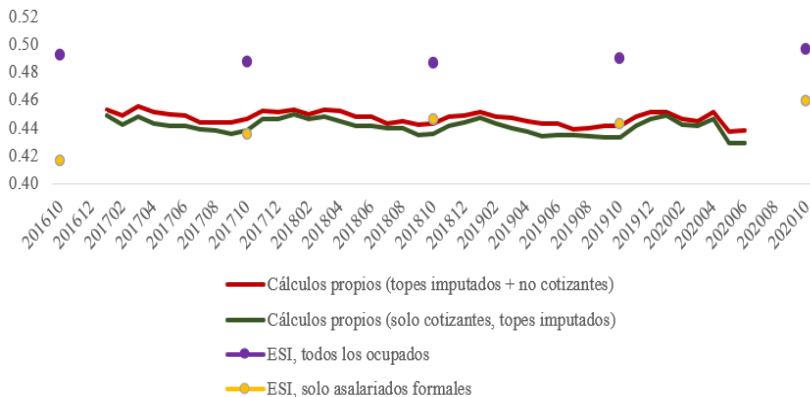
## V. Resultados: Razón interpercentil p90/p10

≡ HORIZONTAL



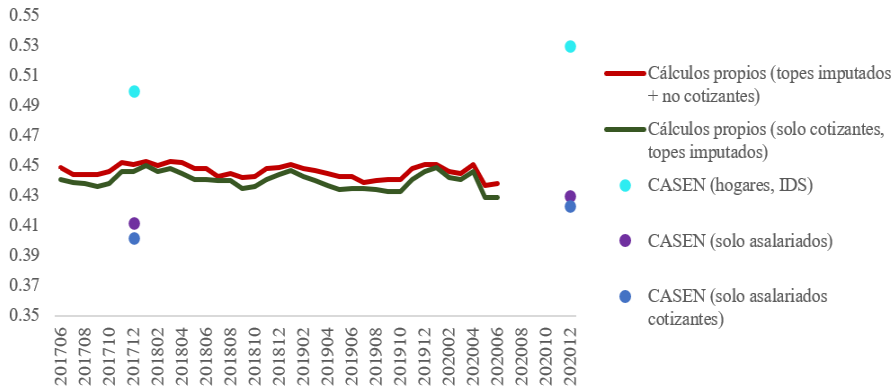
## V. Resultados propios vs. ESI

Gini: Cálculos propios vs. ESI



## V. Resultados propios vs. CASEN

Gini: Cálculos propios vs. CASEN



# VI. Conclusiones

- ▶ El Seguro de Cesantía se valida como una opción para la estimación de la desigualdad.
- ▶ La mayor frecuencia presenta dinámicas no observables de otras fuentes en información.
- ▶ El resultado dependerá de la fuerza de trabajo considerada. Considerando asalariados formales, los cálculos parecen ofrecer un buen ajuste respecto a otras fuentes de información como la ESI.

## VII. Desafíos y mejoras pendientes

- ▶ Especificar de mejor manera la imputación *hotdeck*, ocupando región, comunta, *etc.*
- ▶ Utilizar distintas distribuciones en la imputación de topes, comprobando la sensibilidad de los resultados.
- ▶ Aplicar *imputación múltiple*, repitiendo cada ejercicio de imputación un número de veces y calcular el promedio.
- ▶ Comparar la *variación porcentual* de los resultados con otras fuentes de información
- ▶ Replicar el ejercicio para la totalidad del periodo disponible en la base de datos del Seguro.



- ▶ Atkinson, A. & Piketty, T., 2007. *Top Incomes Over the Twentieth Century: a contrast between continental European and English-speaking countries*. Oxford University Press.
- ▶ Bound, J., Brown, C. & Mathiowetz, N., 2001. *Measurement Error in Survey Data*. Handbook of Econometrics, 5.
- ▶ Jenkins, S., Burkhouse, R., Feng, S. & Larrimore, J., 2009. *Measuring Inequality Using Censored Data: A Multiple Imputation Approach*. Institute for the Study of Labor - Discussion Paper, 4011.
- ▶ Larrañaga, O., Echeopar, B. & Grau, N. 2021. *Una nueva estimación de la desigualdad de ingresos en Chile*. Serie de Documentos de Trabajo. Departamento de Economía, Universidad de Chile.

- ▶ Moore, J., Stinson, L. & Welniak, E., 2000. *Income Measurement Error in Surveys: A Review*. Journal of Official Statistics, 16.
- ▶ McDonald, J., 1984. *Some Generalized Functions of the Size Distribution of Income*. Econometrica, 52.
- ▶ Paraje, G. & Weeks, M., 2010. *Income Nonresponse and Inequality Measurement*. Revista de Análisis Económico, 25 (2).
- ▶ Ripsy, B., McDonalds, J. & Turley, R., 2002. *A Comparison of Parametric Models of Income Distribution Across Countries and Over Time*. Luxembourg Income Study Working Paper, 305.