

Reconciliando fuentes en la práctica:

Problemas, soluciones, y más problemas

Ignacio Flores¹

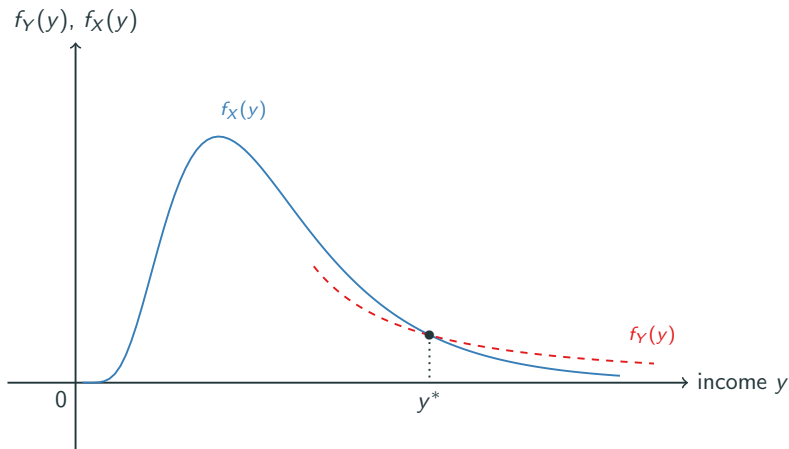
Seminario “Conciliación de fuentes para medir la distribución del ingreso”
CEPAL, 12 Octubre 2023

¹Coordinador regional en World Inequality Lab
PSE - CUNY

- ¿Cómo combinar microdatos?
- ¿Cómo imputar y ajustar flujos totales?
(ingreso, impuestos, transferencias en especie)

Combinar microdatos

Escena típica: Encuestas vs. Tax



Fuente: Blanchet, Flores y Morgan (2022)

Sampling

- Small sample bias (Taleb and Douady, 2015)

Non-Sampling

- Misreporting (Bound and Krueger, 1991; Bollinger, 1998; Angel et al., 2017; Paulus, 2015)
- Heterogeneous response rates across the income distribution (Korinek, Mistiaen and Ravallion, 2006; Johansson and Klevmarken, 2007; Bollinger et al., 2015; Chenevert et al., 2016)

Non-Response

- $f_Z(y) = f_Y(y)\theta(y) \Rightarrow \theta(y) = f_Z(y)/f_Y(y)$

Misreporting

- $f_Z(y) = f_Y(y)(1 - p(y)) + f_M(y)\bar{p}$

Non-Response & Misreporting

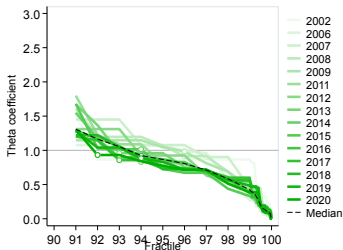
- $f_Z(y)/f_Y(y) = \theta(y)(1 - p(y)) + f_M(y)/f_Y(y)\bar{p}$

Conclusiones

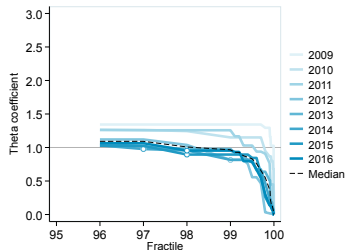
- Si ambos existen, se confunden *ex-post*.
- Misreporting sólo se puede resolver con *matching* individual

La forma del sesgo

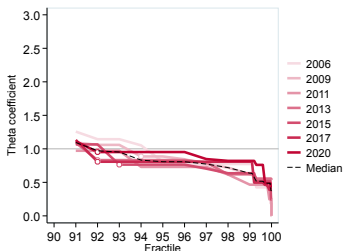
Brasil



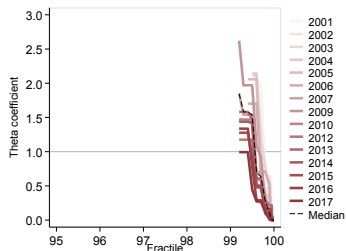
Uruguay



Chile



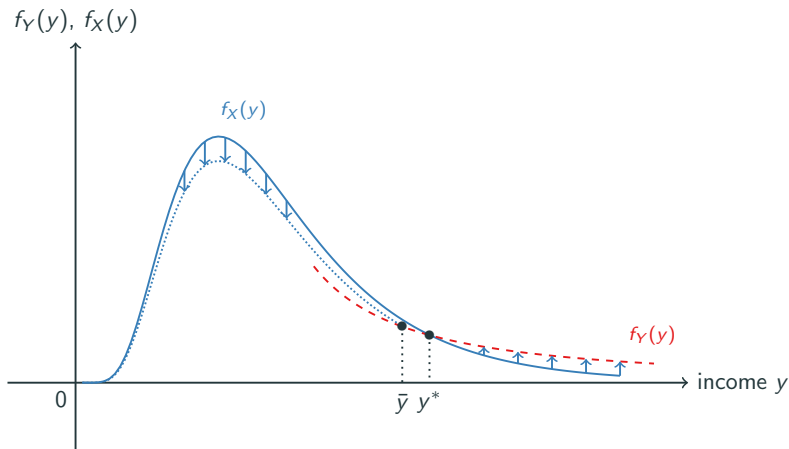
El Salvador



Dos hipótesis fundamentales:

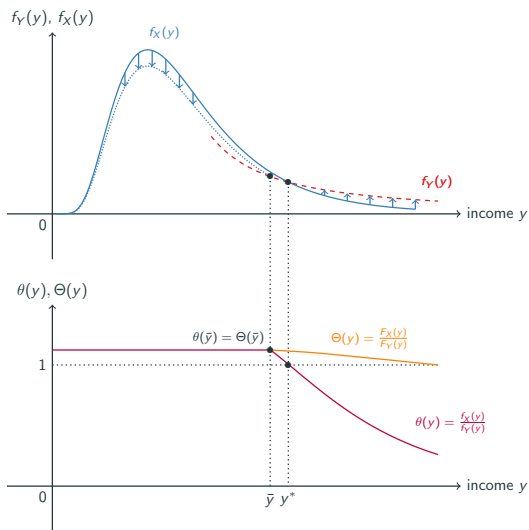
1. Declaraciones fiscales son un *lower bound* para las frecuencias en parte alta de la distribución
2. No cuestionamos las respuestas aportadas por individuos encuestados

Corrección BFM: Intuición



Fuente: Blanchet, Flores y Morgan (2022)

Corrección BFM: Intuición



Fuente: Blanchet, Flores y Morgan (2022)

- Metodología tradicional (e.g. Deville and Särndal, 1992)
- Variable en encuesta x_1, \dots, x_n , con *design weights* d_1, \dots, d_n
- Encontrar nuevos weights w_1, \dots, w_n que entreguen los totales observados en otra fuente:

$$\sum_{k=i}^n w_i x_x = T$$

- Condición: minimizar distancia entre weights originales y corregidos:

$$\min_{w_1, \dots, w_n} \sum_{i=1}^n \frac{(w_i - d_i)^2}{d_i}$$

⇒ Solución se interpreta cómo un modelo de *no-respuesta*.

- **Reweighting**

⇒ Korinek, Mistiaen, and Ravallion (2006); Hlasny and Verme (2017; 2018); Alvaredo (2011, for Argentina)...

- **Replacing**

⇒ Burkhauser et al. (2016); Piketty, Yang, and Zucman (2017); Chancel and Piketty (2017); Czajka (2017); DWP (2015); Alvaredo (2011, for U.S.); Burkhauser et al. (2018); Jenkins (2017)...

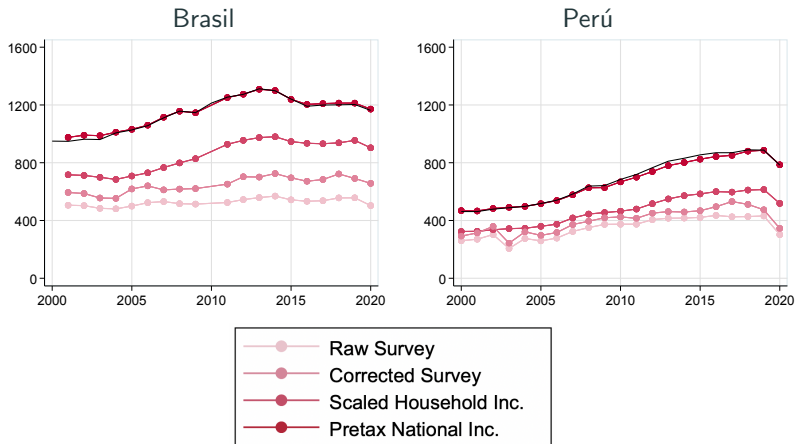
- **Hybrids**

⇒ Bourguignon (2018); Medeiros et al. (2018); *nuestro método*

Nuestro método híbrido usa datos fiscales, encuentra un 'merging point' no-arbitrario y preserva la estructura de micro-datos, incluyendo totales de población (entre otros)

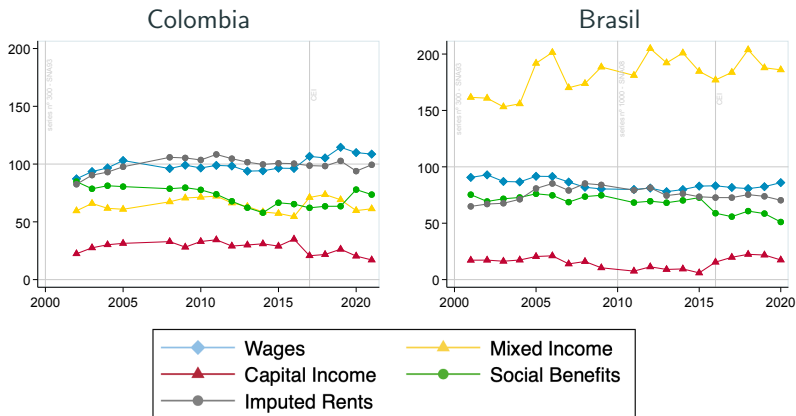
Escalado de flujos

Al escalar, agregamos la mayor cantidad de ingresos



Ingreso promedio en Euros PPP 2020.
Fuente: De Rosa, Flores y Morgan (2022)

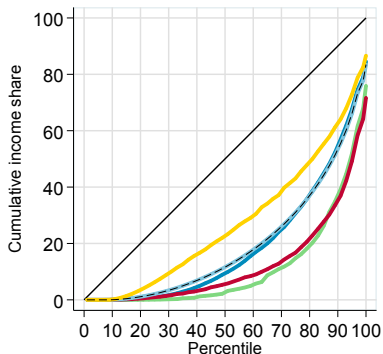
Comparación de totales por tipo de ingreso



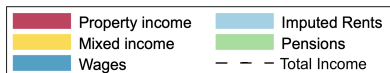
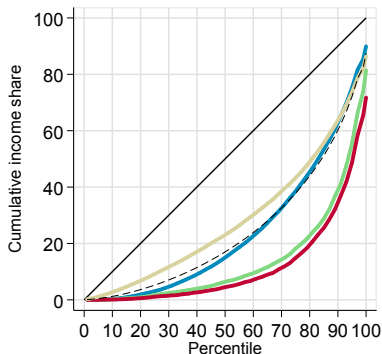
Fuente: Alvaredo et al. (2022)

El nivel de agregación es crucial

Colombia 2020



Ecuador 2020



Fuente: De Rosa, Flores y Morgan (2022)

Coincidencia conceptual de ingresos

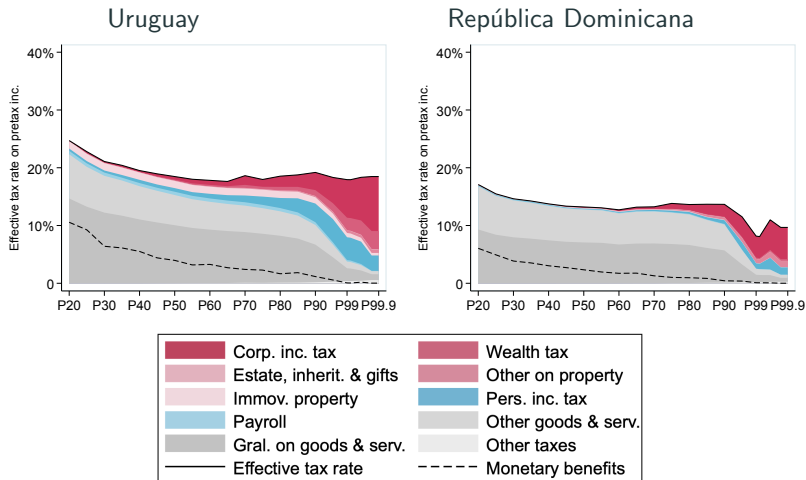
Income in -2* Survey [1]	National Accounts (SNA08) [2] = [3] + [4]	Matching definitions in SNA and Survey [3]	Non-matching definitions in SNA and Survey [4]
Dependent work	Compensation of employees (D1)	Wages, salaries (D11)	Social security contributions (D61)
Imputed rent	Operating surplus (B2)	Rent of owner occupiers	Rental income from dwellings
Capital income	Property income (D4)	Interests received (D41r), Dividends (D42)	Interests paid (D41u) Rent of natural resources (D45) Investment income of insurance policy holders (D441) Investment income of pension funds (D442) Investment income of investment funds (D443)
Independent work	Mixed income (B3)	Self-employed income	Rent of non-dwelling buildings
Benefits	Social transfers (D62)	Pension benefits, Other cash benefits	Sick-leave Unemployment insurance

(*) Sick-leave is part of social insurance benefits in SNA, while it is part of Salaries in surveys.

(**) SNA does not deduct costs when deriving property income.

Fuente: Alvaredo et al. (2022)

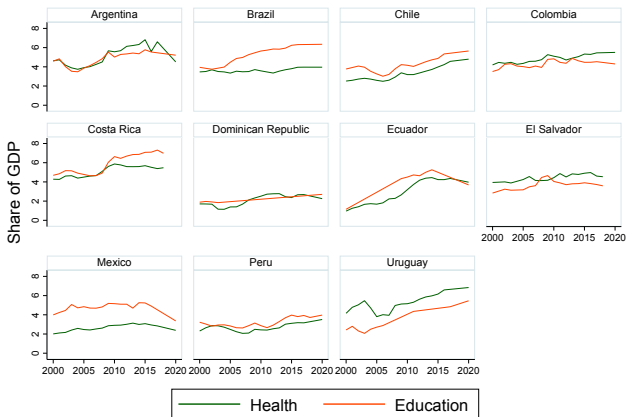
Distribuir la carga fiscal



Fuente: De Rosa, Flores y Morgan (2022)

- Aquí también importa el nivel de agregación

Distribuir transferencias en especie



Graphs by country

Fuente: World Bank

- Uso de perfiles CEQ
- Corrección por consumo en mercados informales?

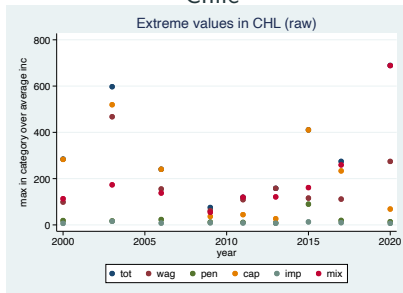
- Serie de decisiones "de expertos" es necesaria
- Las suposiciones / límites datos tienen impacto mayor
- Es necesario abrir el debate y experimentar
- Acceso a más y mejores datos genera oportunidades
- Queda pendiente la distribución de stocks

Gracias!

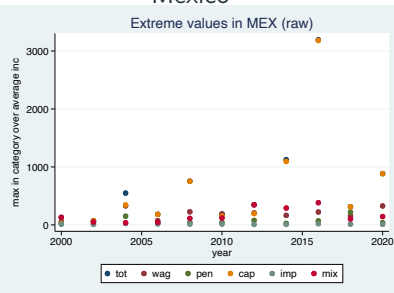
Appendix

Bonus: Extrema sensibilidad a valores extremos

Chile



Mexico



- Censuramos los valores > 1000 veces el promedio de ingreso total

- **The weight of the rich: Improving surveys using tax data**
Thomas Blanchet, Ignacio Flores and Marc Morgan in the Journal of Economic Inequality, 1-32 (2022)
- **More Unequal or Not as Rich? Revisiting the Latin American exception** Mauricio De Rosa, Ignacio Flores and Marc Morgan in Stone Center Working Paper Series (2022)
- **The Inequality (or the Growth) we Measure: Data Gaps and the Distribution of Incomes** Facundo Alvaredo, Mauricio De Rosa, Ignacio Flores and Marc Morgan, in Stone Center Working Paper Series (2022)