Personal income distribution in a peripheral economy: Peru, 2004-2019

Brown bag Seminar - Economics Department, NSSR

César J. Castillo-García

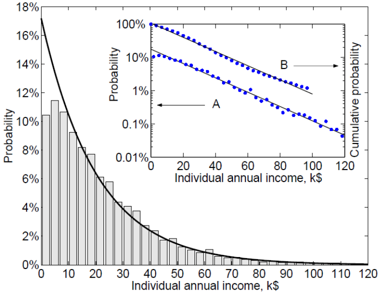
Department of economics, New School for Social Research

November, 2021

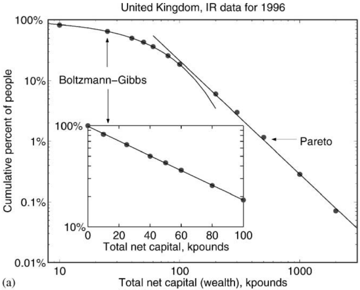
Contents

- Statistical mechanics: income distribution, Lorenz curve and parametrization
- Methodology: maximum entropy model applied to gross household income data
- The Peruvian case 2004-2019: stylized facts
- Analysis of the ENAHO (Encuesta Nacional de Hogares de Perú) and results
- Further steps
- Conclusions

Statistical mechanics: income distribution, Lorenz curve and parametrization



(Dragulescu and Yakovenko, 2001a)



(Dragulescu and Yakovenko, 2001b)

Statistical mechanics: income distribution, Lorenz curve and parametrization

Exponential probability density function for household income (x), μ is the average income

$$f^*(x) = \frac{1}{\mu} e^{\frac{-x}{\mu}}; x \geq 0$$

Lorenz curve (Dragulescu and Yakovenko, 2001b)

$$m = n + (1-n)\log(1-n)$$

$$m = \frac{\int_0^a xf(x)dx}{\int_0^\infty xf(x)dx}$$

$$n = \int_0^a f(x)dx$$

Statistical mechanics: income distribution, Lorenz curve and parametrization

Pareto law (Dragulescu and Yakovenko, 2003; Silva and Yakovenko, 2005; Banerjee and Yakovenko, 2010)

$$m = (1-q)(n + (1-n)\log(1-n)) + q\Theta(n-1)$$

$$q = \frac{\langle x \rangle - \mu_{exp}}{\langle x \rangle} \text{ ; } \textit{income share behaving as a Pareto law}$$

$$\Theta(n-1) = \textit{step function, a jump in the Lorenz curve}$$

$$f(x) \sim \frac{1}{x^{\alpha+1}} \text{ ;} \textit{Pareto probability density function}$$

$$Pr(X > x) = 1 - Pr(X \leq x) \sim \frac{1}{x^{\alpha}} \text{ ;} \textit{Pareto cumulative density function}$$

Methodology: maximum entropy model for gross household income

- Shanon entropy (Shanon, 1948)

$$H[(f_1, f_2, f_3, ..., f_n)] = -\sum f_i log[f_i]; \; where \; 0 \log[0] \equiv 0$$

- Entropy is a concave function of the frequencies, and shows a maximum in $log[n], when \; f_1 =:::= f_n = \frac{1}{n}$
- Entropy is a function of the uncertainty included in an information set.
- When the probabilities of a subset are higher (lower), the entropy diminishes (increases).
- When the probabilities are almost similar, (as in a uniform probability density function), there is maximum entropy (uncertainty).

Methodology: maximum entropy model for gross household income

- In 1950, E. T. Jaynes used a similar principle to derive the probability distribution of the Boltzmann laws according to Gibb's interpretation.
- The maximum entropy model in the info-metrics approach:

$$\underset{\{f_i \geq 0\ \forall i \in \mathbb{N};\ x \in X\}}{\text{Max}} \quad H[(f_1, f_2, f_3, ..., f_n)] = -\sum_{x_i \in X} f(x) \log f(x)$$

subject to;

$$\sum_{k=1}^{K} f_k g_m(x_k) = y_m;\ \ m = 1, 2, 3, ..., M$$

Methodology: maximum entropy model for gross household income

Maximum entropy program for both the exponential distribution and the Pareto law (Foley, 2020; Golan, 2018; Kaniadakis, 2009; Silva, 2005)

$$\underset{\{f(x)\geq 0; g(x)\geq 0; x\in[0,\infty[\}}{\text{Max}} \quad H(x) = -\int_0^b f(x)\log f(x)dx - \int_b^\infty g(x)\log g(x)dx$$

subject to;

$$\int_0^b f(x)dx = p$$

$$\int_b^\infty g(x)dx = q \;;\; p+q=1$$

$$\int_0^b xf(x)dx = \mu_{exp}$$

$$\int_b^\infty log(x)g(x)dx = \mu_{par}$$

Methodology: maximum entropy model for gross household income

Maximum entropy program optimization

$$\mathcal{L}(f,g,\lambda,\delta) = -\int_0^b f(x)\log f(x)dx - \int_b^\infty g(x)\log g(x)dx +$$
$$(\lambda_0 - 1)\left(p - \int_0^b f(x)dx\right) + (\delta_0 - 1)\left(q - \int_b^\infty g(x)dx\right) +$$
$$\lambda_1\left(\mu_{exp} - \int_0^b xf(x)dx\right) + \delta_1\left(\mu_{par} - \int_b^\infty log(x)g(x)dx\right)$$

Methodology: maximum entropy model for gross household income

Maximum entropy program optimization

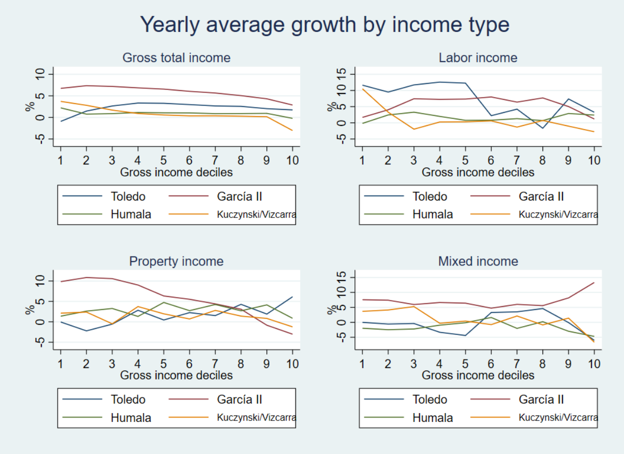
$$\frac{\partial \mathcal{L}(f,g,\lambda,\delta)}{\partial f(x)} = 0$$
$$\frac{\partial \mathcal{L}(f,g,\lambda,\delta)}{\partial g(x)} = 0$$
$$\frac{\partial \mathcal{L}(f,g,\lambda,\delta)}{\partial \lambda_i} = 0 \ ; \ \mathrm{i} = 0, 1$$
$$\frac{\partial \mathcal{L}(f,g,\lambda,\delta)}{\partial \delta_i} = 0 \ ; \ \mathrm{i} = 0, 1$$

$$f^*(x) = \frac{1}{\mu} e^{\frac{-x}{\mu}} ; 0 \leq x \leq b$$
$$g^*(x) \sim \frac{1}{x^{\alpha+1}} ; \ x > b$$

Methodology: maximum entropy model for gross household income

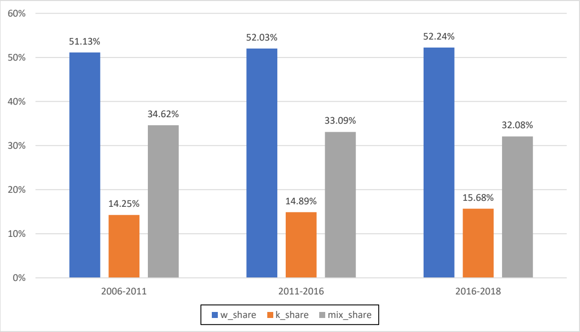
- Measured concept: total gross income (gross household income)
- Assumption 1: equal income distribution in within a household
- Assumption 2: do not correct by the problem of unreported income
- Assumption 3: Household ranked according to gross income, yet it is possible to rank by either wages or capital income
- State space: gross income expressed in terms of Peru's national currency (soles, S/.)

Results are different according to the type of income analyzed for across the households by deciles



Source: Instituto Nacional de Estadística e Informática (INEI) (2004-2019)

Household income source, average 2004-2019

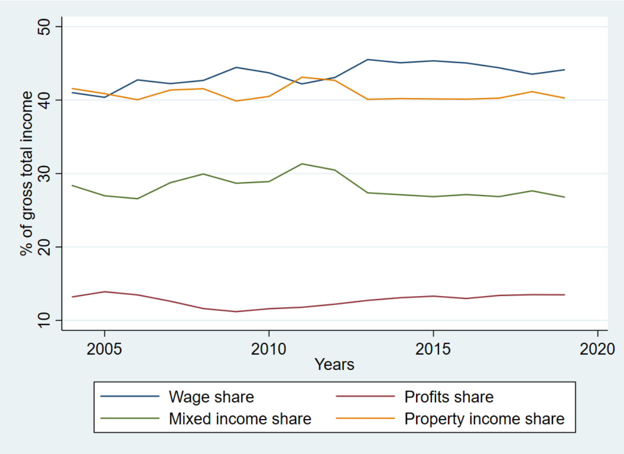


Source: Instituto Nacional de Estadística e Informática (INEI) (2004-2019)

The Peruvian case 2004-2019: stylized facts

Functional income distribution of Peruvian households

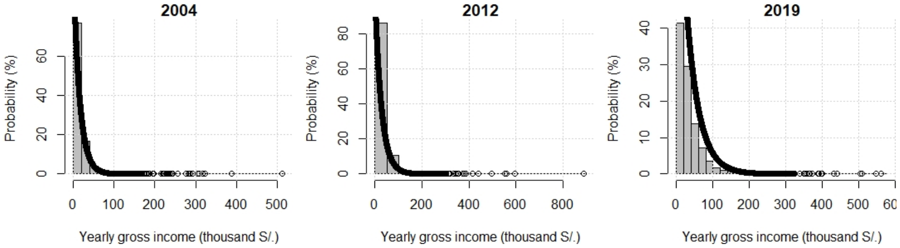
Household income source, average 2004-2019



Source: Instituto Nacional de Estadística e Informática (INEI) (2004-2019)

Analysis of the ENAHO (Encuesta Nacional de Hogares de Perú) and results

Probability density function for gross income in Peru



Source: Instituto Nacional de Estadística e Informática (INEI) (2004-2019)

Analysis of the ENAHO (Encuesta Nacional de Hogares de Perú) and results

Plotting the exponential fitting curve

Exponential density function

$$f(x) = \frac{1}{\mu} e^{\frac{-x}{\mu}}; x \geq 0$$

Exponential cumulative density function

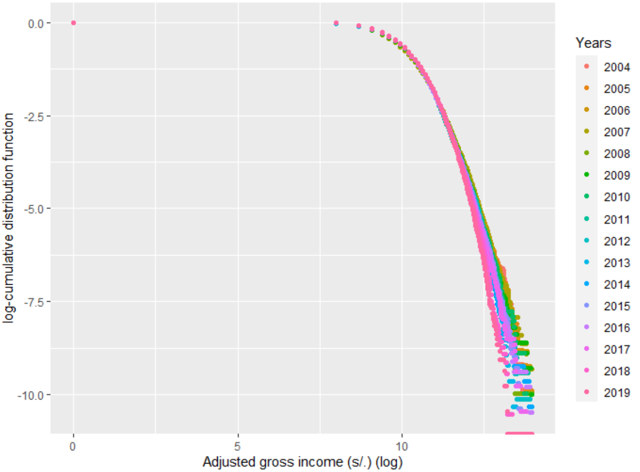
$$\mathrm{F(x)}=\mathrm{Pr(X}\leq x) = 1 - e^{\frac{-x}{\mu}}; x \geq 0$$

Fitting curve graph

$$gr(F(x_t)) := \{(\log(\frac{x_t \mu_{2019}}{\mu_t}), \log(1 - F(\frac{x_t \mu_{2019}}{\mu_t})) \in \mathbb{R}^+ \times \mathbb{R}^- : x_t \in \mathbb{R}^+ \cup \{0\}\}$$

Analysis of the ENAHO (Encuesta Nacional de Hogares de Perú) and results

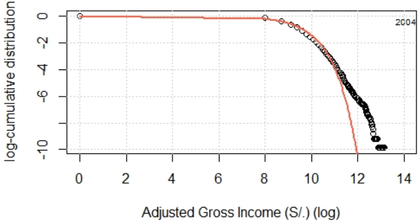
Normalized exponential fitting curves



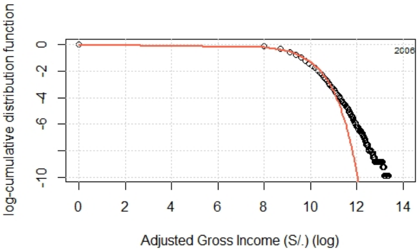
Source: Instituto Nacional de Estadística e Informática (INEI) (2004-2019)

Analysis of the ENAHO (Encuesta Nacional de Hogares de Perú) and results

Exponential distribution fitting (1)



(a) Peru,2004

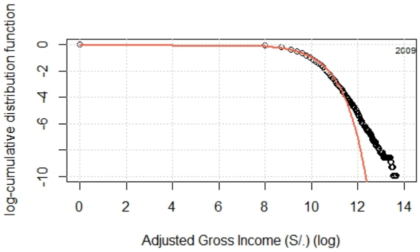


(b) Peru,2006

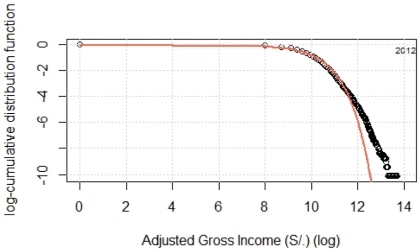
Figure: Cumulative distribution function for the gross income (log-log scale)

Analysis of the ENAHO (Encuesta Nacional de Hogares de Perú) and results

Exponential distribution fitting (2)



(a) Peru,2009

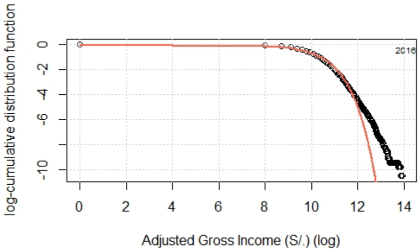


(b) Peru,2012

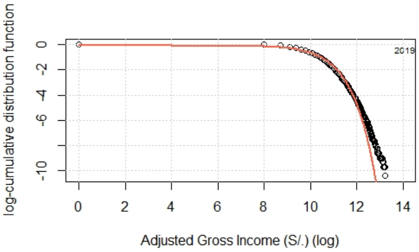
Figure: Cumulative distribution function for the gross income (log-log scale)

Analysis of the ENAHO (Encuesta Nacional de Hogares de Perú) and results

Exponential distribution fitting (3)



(a) Peru,2016



(b) Peru,2019

Figure: Cumulative distribution function for the gross income (log-log scale)

Analysis of the ENAHO (Encuesta Nacional de Hogares de Perú) and results

Corrigendum:exponential distribution and Pareto law

Pareto distribution

$$g(x) = \frac{\alpha b^{\alpha}}{x^{\alpha+1}} \sim \frac{1}{x^{\alpha+1}}; \int_b^{\infty} g(x)dx = q = 0.03 \ (\textit{Shaikh and Jacobo},\ 2019)$$

Pareto cumulative distribution function

$$\mathrm{G(x)}{=}\mathrm{Pr(X}{\leq}\ x) = 1 - Pr(X > x) = 1 - (\tfrac{b}{x})^{\alpha} \sim \tfrac{1}{x^{\alpha}}; \quad \alpha = \tfrac{\log q}{\log q - \log p}$$

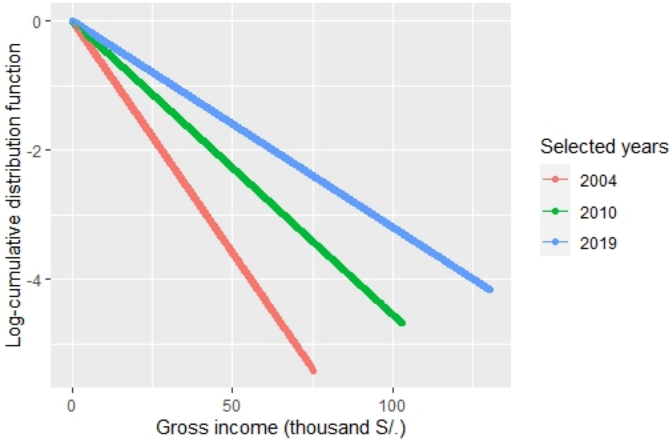
Fitting curve graphs: exponential and Pareto distributions

$$gr(F(x_t)) := \{(\tfrac{x_t}{\mu_t}, \log(1 - F(\tfrac{x_t}{\mu_t}))) \in \mathbb{R}^+ \times \mathbb{R}^- : 0 \leq x_t \leq b\}$$

$$gr(G(x_t)) := \{(log(\tfrac{x_t}{\mu_t}), \log(G(\tfrac{x_t}{\mu_t}))) \in \mathbb{R}^+ \times \mathbb{R}^- : x_t > b\}$$

Analysis of the ENAHO (Encuesta Nacional de Hogares de Perú)

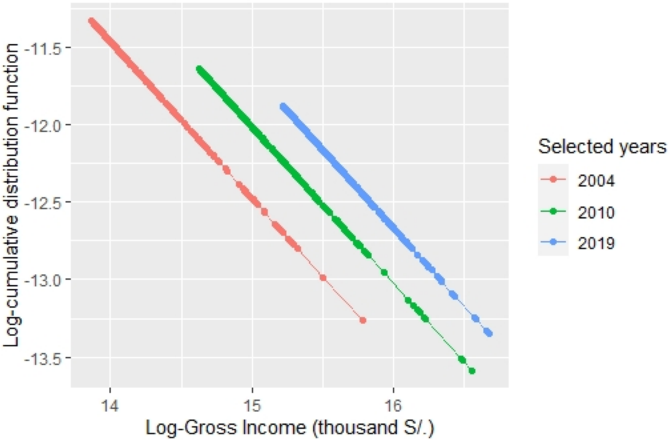
Cumulative exponential distribution (bottom 97%) (log-linear scale)



Fuente: Instituto Nacional de Estadística e Informática (INEI) (2004-2019)

Analysis of the ENAHO (Encuesta Nacional de Hogares de Perú) and results

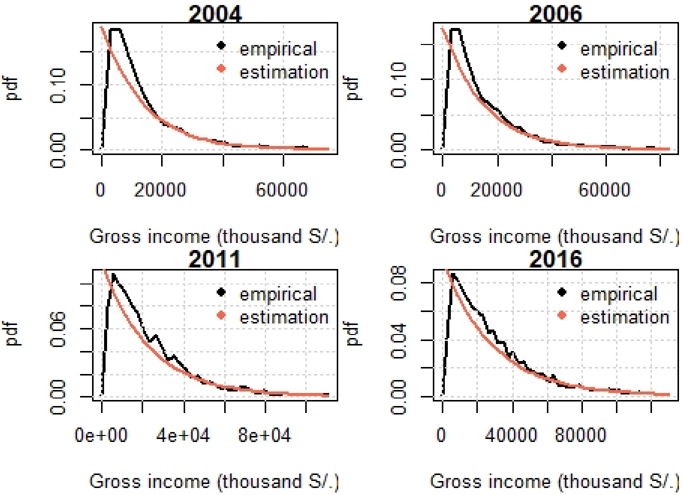
Cumulative Pareto distribution (top 3%) (log-log scale)



Source: Instituto Nacional de Estadística e Informática (INEI) (2004-2019)

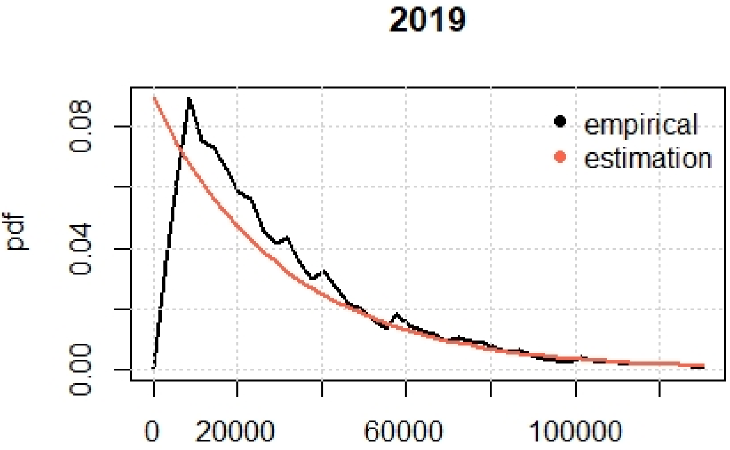
Analysis of the ENAHO (Encuesta Nacional de Hogares de Perú) and results

Computations-exponential distribution share



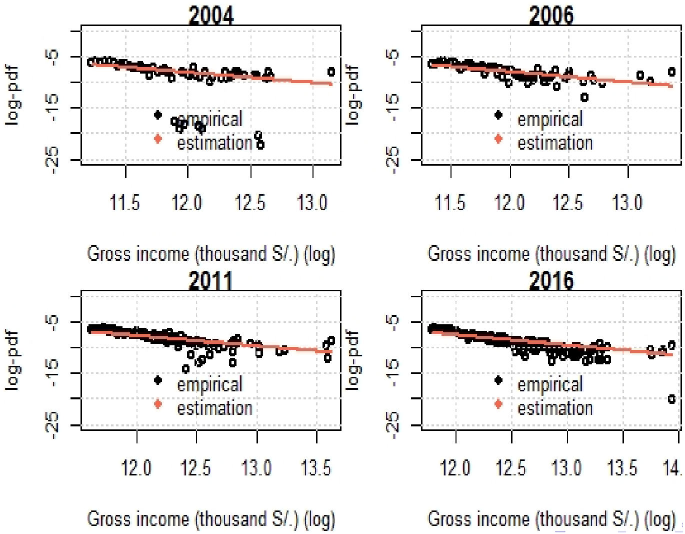
Analysis of the ENAHO (Encuesta Nacional de Hogares de Perú) and results

Computations-exponential distribution share



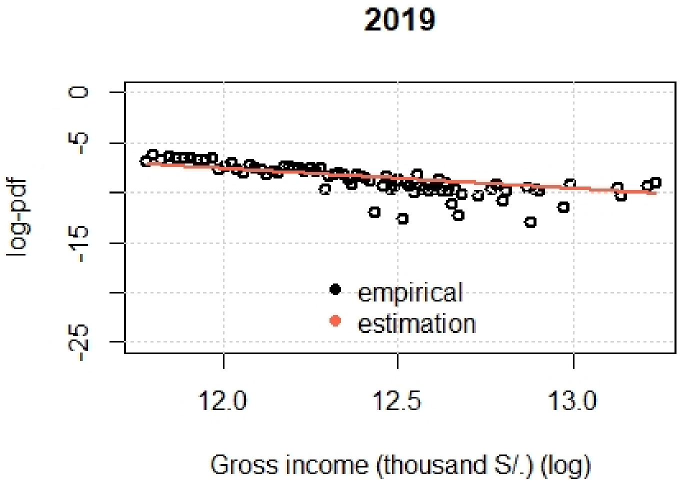
Analysis of the ENAHO (Encuesta Nacional de Hogares de Perú) and results

Computations-Pareto distribution share



Analysis of the ENAHO (Encuesta Nacional de Hogares de Perú) and results

Computations-Pareto distribution share



Information loss metric

$$D_{KL}(f||h) = \sum_{i=1}^{N} f(x_i) * \log \frac{f(x_i)}{h(x_i)} = \sum_{i=1}^{N} f(x_i) * (\log f(x_i) - \log h(x_i))$$

Kullback-Leibler divergence for the analyzed case

$$D_{KL}(observed||max - ent) = \int_0^b f_{obs}(x) * (\log f_{obs}(x) - \log f_{exp}(x))dx \\ + \int_b^{\infty} f_{obs}(x) * (\log f_{obs}(x) - \log g_{par}(x))dx$$

Analysis of the ENAHO (Encuesta Nacional de Hogares de Perú) and results

Kullback-Leibler divergence

Años	Exp	Exp ($\leq 97\%$)	Pareto ($\geq 97\%$)	Exp&Pareto
2004	1.809	0.211	0.015	0.178
2005	2.417	0.204	0.009	0.138
2006	1.908	0.188	0.014	0.148
2007	2.321	0.161	0.011	0.106
2008	1.819	0.143	0.013	0.118
2009	1.872	0.146	0.013	0.110
2010	1.648	0.142	0.013	0.121
2011	1.723	0.133	0.012	0.109
2012	1.736	0.114	0.013	0.1
2013	1.360	0.120	0.013	0.117
2014	1.868	0.120	0.012	0.112
2015	1.508	0.125	0.012	0.121
2016	1.821	0.114	0.011	0.112
2017	1.649	0.110	0.012	0.109
2018	1.269	0.115	0.015	0.118
2019	1.150	0.122	0.015	0.124

Further steps

- Modify the maximum entropy model to add deciles and other quantiles as constraints
- Analyze the implications of assessing the functional income distribution in personal income distribution measured by the maximum entropy methods
- Revisiting the personal income distribution and the Gini coefficient for the Peruvian case: it is necessary to merge different data sources
- Expand the income data set to include other country cases with similar characteristics as Peru (Latin America and middle-income economies)

Conclusions

- From a macroscopic perspective, the gross household income distribution fits an exponential probability distribution. This is an "obvious" statistical result because income distributions are positive skewed.
- Yet, higher percentiles do not completely fit the exponential distribution shape. Hence, it makes sense to state that the upper-tail section of the income distribution behaves as a Pareto probability function.
- The analysis of ENAHO suggests that the concentration of income in the higher percentiles is not a necessarily consequence of the accumulation of property income.

Conclusions

- The maximum entropy estimation shows that the total gross income for Peru (including mostly labor and mixed income) fits an exponential shape at the bottom part of distribution and a Pareto distribution in the upper-tail. Such a result replicates previous findings of Draguslecu and Yakovenko (2003), Silva and Yakovenko (2005), Banerjee and Yakovenko (2010).
- Following the methodology of Shaikh and Jacobo (2019), we find that the bottom 97% of the overall distribution of the Peruvian gross household income is well approximated by an exponential distribution, while top 3% is well approximated by a power law. The evaluation of Kullback-Leibler divergences for the state space confirms such a result.
- Although Peruvian labor income for the households at the lower percentiles have improved in the last 5 years, the decrease in other income sources and the relative stabilization of the percentiles in the middle part of the distribution have generated a concentration of gross income in the upper-tail.


THAT'S ALL FOLKS