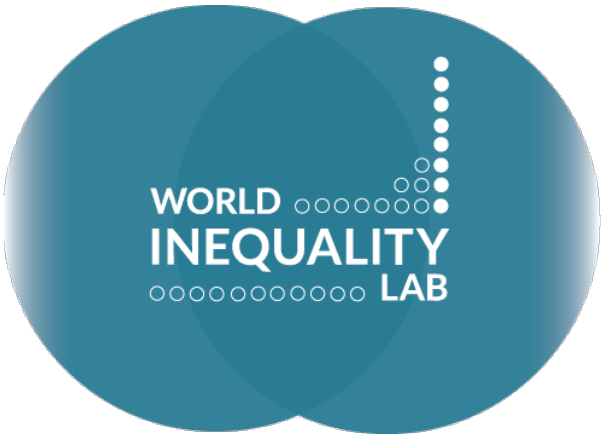
World Inequality Lab – Working Paper N° 2021/18

Income and Wealth Inequality in Hong Kong, 1981-2020: The Rise of Pluto-Communism?

Appendix

Thomas Piketty
Li Yang

June 2021



WID.WORLD
THE SOURCE FOR
GLOBAL INEQUALITY DATA

Appendix A-E

Appendix A: Data Description

Our study relies on below five main data sources.

1. HK Salaries Tax Assessment (1973 – 2018) and HK Corporate Profit Tax (1989-2018)
2. HK Census Micro data for 1981, 1986, 1991, 1996, 2001, 2006, 2011, 2016.
3. HK Rich List (1996 – 2020)
4. HK National Account (1980-2020)
5. HK pre-nomination Survey for 2016 Legislative Council Election

Below we will explain each data source in details.

1. HK Salaries Tax Assessment (1973-2017) and HK Corporate Profit Tax (1989-2018)

Both data sets are provided by the HK Inland Revenue Department Annual Report. HK Salaries Tax Assessment covers only the salaries income (including employer's contribution for retirement schemes). Both total income and chargeable income are reported. Hong Kong corporate profits tax covers profits from both domestic and foreign corporations as well as domestic and foreign unincorporated businesses in Hong Kong. Unincorporated business does not possess a separate legal identity from its owner(s). The owner(s) bear full liability for any action or inaction of the business: they may sue and be sued for business activity or inactivity. Unincorporated enterprises include sole proprietor-ships, partnerships and family trusts. For details, please see https://www.ird.gov.hk/eng/paf/pam.htm#stpa.

2. HK Census

Micro sample data set is purchased from Census and Statistic Department. For wave 1981 and 1986, the sample data set covers 1% of the Census. For the rest of the waves, the sample data set covers 5% of the Census.

a. Sampling approach

The sampling approach has been changed periodically. Caution is required when comparing statistic results across different waves.

1981, 1986, 1991: de facto enumeration approach.
1996: de jure enumeration approach.
2001, 2006, 2011, 2016: Resident population approach.

b. Income Variables

The census income variables have two constraints:

First, both household income and individual income are top-coded. Taking individual income from employment as an example, monthly income above 99,998 HKD is top-coded in wave 1981 to 1991; monthly income above

150,000 is top-coded in wave 1996-2016. The share of population with top-coded income varies from 1% to 2.5%. Nationally, on option to correct such census data is to apply the top correction using salary tax data.

Second, there are only two income sources for individual income:

Income from employment: for employers or self-employed persons, this is the amount earned excluding expenses incurred in running their main business; for employees, this is the total amount earned from their main employment including salary or wage, bonus, commission, overtime, housing allowance, tips and other cash allowances.

Income from other sources: rent income, interest, dividend, education grants (excluding loan), regular/monthly pensions, social security payment, old age allowance, disability allowance, comprehensive social security assistance, scholarships, regular, contribution from persons outside the household, contribution from charities.

Potentially, one could use the information contained in Employment Status to split income from employment into wage income and business income.

3. HK Rich List (1988 – 2019)

Wealth data in HK is very limited. Our rich list is retrieved from Forbes magazine. Before 2007 HK data is from the world billionaires list. From 2008 to 2012, the data is from the Hong Kong's 40 Richest. Since 2013 data is from the Hong Kong's 50 Richest.

4. HK National Account (1980-2016)

Our data is from UN and Census and Statistic Department. The National Accounts of Hong Kong are highly aggregated. Only decomposition of primary income account and external account for total economy are available. For details, please Appendix D: HKG_NA.xlsx

5. HK Survey for 2016 Legislative Council Election

HK election survey data is from the Public Opinion Program (POP) of The University of Hong Kong (HKU).

The Pre-nomination Survey for 2016 Legislative Council Election was conducted between June 20 and July 8, 2016 over 5,084 adult registered voters in Hong Kong who speak Cantonese. The pre-nomination survey contains information of interviewee of gender, age, education level, occupation, income level, and residential district, and political inclination.

The 2016 Election Rolling Survey is a daily rolling survey, conducted from July 30 to September 2 2016. For each day it interviews daily interviewing at least 200 adult registered voters in Hong Kong who speak Cantonese. it provides less detailed information on education levels and occupation than the pre-nomination survey.

For more details of both surveys, pleases see https://develo.pori.hk

Appendix B: Codebook of harmonized dataset of Hong Kong Census

1. year: census year
2. DIST1: residential area, 1981, 1986, 1996, 2001, 2006, 2011, 2016
   Hong Kong Island
   Kowlong
   New Territories
3. DIST2: residential area, sub-district level, 1981, 1986, 1996, 2001, 2006, 2011, 2016

2	Central and Western	中西区
3	Wan Chai	湾仔区
4	Eastern	东区
5	Southern	南区
6	Yau Tsim Mong	油尖旺区
7	Sham Shui Po	深水涉区
8	Kowloon City	九龙城区
9	Wong Tai Sin	黄大仙区
10	Kwun Tong	观塘区
11	Sai Kung (Tseung Kwan O)	西贡区 1
12	Sai Kung (Other Areas)	西贡区 2
13	Tsuen Wan (New Town)	荃湾区 1
14	Tsuen Wan (Other Areas)	荃湾区 2
15	Tuen Mun (New Town)	屯门区 1
16	Tuen Mun (Other Areas)	屯门区 2
17	Yuen Long (New Town)	元朗区 1
18	Yuen Long (Tin Shui Wai)	元朗区 2
19	Yuen Long (Other Areas)	元朗区 3
20	North (New Town)	北区 1
21	North (Other Areas)	北区 2
22	Tai Po (New Town)	大埔区 1
23	Tai Po (Other Areas)	大埔区 2
24	Sha Tin (New Town)	沙田区 1
25	Sha Tin (Ma On Shan)	沙田区 2
26	Sha Tin (Other Areas)	沙田区 3
27	Kwai Tsing (Kwai Chung)	葵青区 1
28	Kwai Tsing (Tsing Yi)	葵青区 2
29	Kwai Tsing (Other Areas)	葵青区 3
30	Islands	离岛区
91	Marine	

4. lq: living quarter ID
5. hh: household ID
6. ind: individual ID. In 1986, 3345 missing value.
7. wt_lq: living quarter weight. It is equal to 1 for all years.
8. wt_hh: household weight. It is equal to 1 for year 1981, 1986
9. wt_ind: individual weight. It is equal to 1 for year 1981, 1986
10. tnr: tenure type (household level)
    1. Own the premises
    2. Sole tenant
    3. Co-tenant
    4. Main-tenant
    5. Sub-tenant
    6. Rent free
    7. Provided or subsidized by employer
    8. N.A.
11. rent: monthly Household Rent paid. It includes rates for that month but excludes payments for water, electricity and management fee.
12. mrtg: monthly mortgage paid, 2001, 2006, 2011, 2016.
13. mrtgyr: mortgage year left, 2001, 2006, 2011, 2016.
14. hhinc: monthly household income. There are missing values
15. hhsize: household size
16. hhadt: number of adult in household
17. age: age, when age=75, it includes 75+.
18. pob: place of birth. In 1986, only including 3 categories HK, China (main), Other. In 1981 China (main) including China mainland and Macro; in the rest of the years China (main) including China mainland and Macro and Taiwan.
    1. HK
    2. China(main)
    3. Japan
    4. South Asia
    5. UK
    6. North America
    7. Australia
    8. Others
19. nat: nationality, 1991, 1996, 2001, 2006, 2011, 2016.
    a. British
    b. HK Chinese
    c. Other Chinese
    d. Filipino
    e. N. American
    f. S. Asian

g. Thai
h. Japanese
i. Austria
j. Other

20. lan: language spoken at home, 1991, 1996, 2001, 2006, 2011, 2016.
    a. Cantonese
    b. English
    c. Putonghua
    d. ChiuChau
    e. Hakka
    f. Fukien
    g. SzeYap
    h. Shanghainess
    i. OtherChineseDialects
    j. Filipino
    k. Japanese
    l. Others
21. mar: marital status
    1. Single
    2. Married
    3. Widowed
    4. Divorced/Separated
22. edu: educational attainment (highest level attended)
    1. No Schooling
    2. Primary
    3. Secondary
    4. Craft
    5. Colleges and Polytechnic
    6. University
    7. Post-graduate
23. emp: employment status. In 1991, selfemployed (Hawking) and selfemployed (others) are combined in one category, selfemployed.
    1.1 Employees
    1.2 Employer
    1.3 Selfemployed (Hawking)
    1.4 Selfemployed (Others)
    1.5 Unpaid worker
    2. Unemployed
    3.1 Home-maker
    3.2 Student
    3.3 Of Independent means

3.4 Retired

3.5 Others

24. occ: occupation. For all years, “Worker” includes sales worker, services worker, skilled worker and elementary worker

    1. Professions

    2. Adm & Managerial

    3. Clericals

    4. Worker

        4.1 Sales Worker

        4.2 Services Worker

        4.3 Skilled Worker

        4.4 Elementary Worker

25. inds: Industry.

    a. Agriculture[1]

    b. Minging

    c. Manufacturing

    d. Electricity, Gas and Water

    e. Construction

    f. Wholesale and Retail

    g. Restaurant and Hotel

    h. TSC

    i. Financial

    j Servince

26. inc_wage: monthly income of employees

27. inc_busi: monthly income of employer and self-employed person

28. inc_oth[2]: monthly income from other sources. See definition for various years below

    1981: rent, dividend, interest, pension, remittance, social welfare payment, ect.

    1986: secondary employment earnings, interest, rent, social welfare payment, old age allowance ect.

    1991-1996: interest, dividend, rental, welfare payment, old allowane etc.

    2001-2016: rent income, interest, dividend, education grants (excluding loan), regular/monthly pensions, social security payment, old age allowance, disablity allowance, comprehensive social security assistance, scholarships, regular, contribution from persons outside the household, contribution from charities.

[1] Industrial code, 2016:HSIC v2.0 (adjusted using HISC v1.1); 2011: HSIC v1.1; 1991, 1996: SIC1991

[2] Note that inc_wage and inc_oth includes income for retired person.

Appendix C: Methodology for constructing wage and total income series

In this section, we describe step-by-step how we construct the inequality series of wage and total income in Hong Kong.

Wage Distribution

Step 1. We start by estimating the wage distributional series (**svy_wage**[3]) using Hong Kong income survey data (sample datasets from Hong Kong population census). However, there are two concerns:

i) the wage income (as well as other income variables) in the survey are top-coded and the share of top-coded observations of the sample increases along the time, i.e. in 1991 0.07% of the sample are top-coded with wage income, in 2016 the share increased 5 times to 0.35%. (For more details on top-coding, please see Appendix D, Sheet AX5.) Naturally top-coded sample will generate a downwards bias to the top of the distribution.

ii) Surveys often fails at capturing the top income because individuals from high-income strata typically have very low response rates in surveys.

Step 2. We first correct the top coding issue. We assume that the top of the wage distribution of Hong Kong follows a Pareto distribution. This allows us to estimate the Pareto coefficient α using the data (shortly) before the cut off C. Then using estimated $\hat{\alpha}$ and C, we could estimate the wage mean of the top-coded observations, which is subsequently assigned to all observations in the top-coded group. We then estimate the wage distributional series based on the uncoded survey (uncoded_svy_wage). See example below.

As an example, the graph below shows the how we estimate the Pareto coefficient α for wage distribution in 2016.

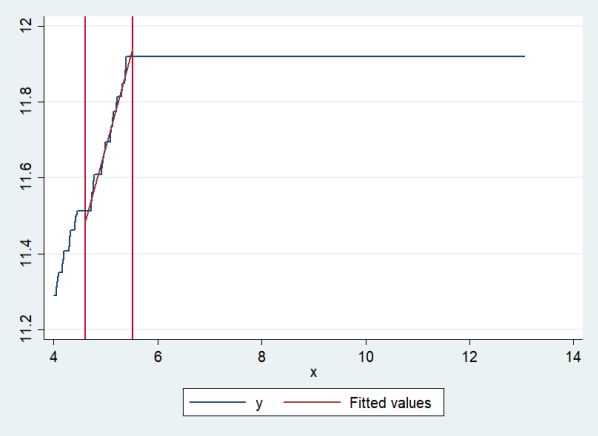
i) Using the observations from p99-p99.6, we estimate the Pareto coefficient α = 0.498.

ii) We calculate b coefficient using α, b = 1/(1-α) = 1.99.

*iii) We estimate the mean for the top-coded group with b and the threshold of cutoff (1.8 million HKD), Mean_top=1.8*b = 3.587 million HKD*

iv) Replace wage to 3.587 million HKD if they were equal to 1.8 million HKD.

[3] Including wage and pension income



Step 3. To overcome the problem regarding missing survey data for the top incomes, we correct the uncoded survey using fiscal data. To do so, we adopt the novel reweighting method proposed in Blanchet, Flores, Morgan (2018). The basic idea is to adjust the weight of the survey[4] using fiscal data in a nonparametric manner[5] under the assumption that tax data sets a credible lower bound on the number of people with given levels of income. Rather than directly making assumptions regarding the behavior of complex statistics such as quantiles or bracket averages (i.e., Piketty, Yang, and Zucman, 2019; Chancel and Piketty, 2017; Czajka, 2017; Morgan, 2017), their method makes easily interpretable assumptions at the observation level. Most importantly, this method allows us to correct the income distribution without losing the richness of the information in surveys. Therefore, in the adjusted survey data, the representativeness in terms of age, gender, or ethnicity, in the case of this research, will be maintained. We then estimate the wage distributional series based on the adjusted survey (**adjusted_svy_wage**).

Step 4. Until now we use individual adult as the unit of observation. In this step, we convert the individual adult wage into the wage equal-split among the adults in the household, and estimate the wage distributional series for survey (**svy_wage_eq**), uncoded survey (**uncoded_svy_wage_eq**), and adjusted survey (**adjusted_svy_wage_eq**) , which we will use as our bench mark series.

We provide the full results of the correction and the corresponding computer code used for applying the method in DataAppenidx.

[4] In addition to adjusting the weight of the survey, the reweighting method also adjust the very top incomes by adding observations.

[5] in contrast to previous reweight methods used in the literature, Blanchet, Flores, Morgan (2018) use a data-driven method, instead of relying on ad hoc assumptions and parameters, to determine where the bias starts in the survey data and beyond which point incomes are merged from tax data into the survey data.

Total Income Distribution

In our sample we define total income equal to sum of wage income, business income, other income. Other income includes transfer income, rent and capital income (excl. pension).

Step 1. We start with raw survey data (**svy_inc**). Following the similar approach above we correct the top-coded observations on wage, business and other income (**uncoded_svy_inc**).

Step 2. Using uncoded survey, we then convert the individual adult total income into the total income equal-split among the adults in the household (**uncoded_svy_inc_eq**).

Step 3. We generate upgrade factor for each g-percentile using equal split wage distribution from uncoded survey (**uncoded_svy_wage_eq**) and adjusted survey (**adjusted_svy_wage_eq**).

Upgrade factor (by g-percentile) = (mean_wage in adjusted survey)/ (mean_wage in uncoded survey)

Step 4. Then we apply the upgrade factor to each g-percentile of total income in uncoded survey and estimate the corrected total income distribution series (**adjusted_svy_inc_eq**).

Appendix D: Income and Wealth Data Set

See file HKG_NA.xlsx

Appendix E: Main Figures and Tables

See file MainFiguresTables.xlsx