China Economic Review 77 (2023) 101912

Contents lists available at ScienceDirect

China Economic Review

journal homepage: www.elsevier.com/locate/chieco

Top incomes and income polarisation in China

Chen Zhang [a,1], Yangcheng Yu [b,1,*], Qinghai Li [c,*]

[a] Institute of Population and Labour Economics, Chinese Academy of Social Sciences, Beijing, China
[b] School of Economics, Shanghai University of Finance and Economics, Shanghai, China
[c] School of Economics, Nanjing University of Finance and Economics, Nanjing, China

ARTICLE INFO

JEL:
C81 D31 D63

Keywords:
Top-income data
Household survey data
Income polarisation
DER index
Relative distribution

ABSTRACT

Income polarisation is normally measured using household survey data, but such data often provide insufficient coverage of top incomes. This paper combines data from the Chinese Household Income Project and Top Incomes in China databases for 2002, 2013, and 2018 to study the trends in, reasons for, and heterogeneity of income polarisation in China between 2002 and 2018. Our main findings are as follows. (1) Introducing external top-income data increases the estimated level of polarisation in each of the three surveyed years, but polarisation trends are not affected. (2) Polarisation increases significantly between 2002 and 2013, due to a rise in polarisation among poor residents. Polarisation remains stable between 2013 and 2018, due to the expansion of the middle-income group; the convergence to the middle of the distribution is attributable mainly to the poor, rather than the rich. (3) In 2018, levels of polarisation are higher among male, well educated, and urban residents than among female, less educated, and rural residents, respectively.

1. Introduction

Narrowing China's income gap has long drawn attention from policymakers and researchers. At the 18th National Congress of the Chinese Communist Party (CCP) in 2012, the Chinese government made achieving "common prosperity" a priority. Today, having largely won the fight against extreme poverty and promoted development in all areas of society, China has achieved favourable conditions for promoting common prosperity. However, many scholars have pointed out that China's wide income gap poses a serious obstacle to achieving common prosperity. Early research found that China's Gini coefficient increased from 0.31 in the early 1980s to 0.45 in the 2000s (Ravallion & Chen, 2007), but this increasing trend has been disrupted in recent years. The National Bureau of Statistics of China reported that the Gini coefficient reached a peak of 0.491 in 2008 and subsequently showed a downwards trend, falling to 0.468 in 2020 (NBS, 2021). Combining survey data with high-income data from tax records, Piketty, Yang, and Zucman (2019) reported that inequality in China, as measured by top 10% income shares, remained stable between 2006 and 2015.

Although income inequality has received a great deal of attention in the literature, income polarisation is less well explored, especially in China. Tan, Zeng, and Zhu (2018) analysed statistical data from the 2011 Chinese Household Finance Survey and found that different indicators of income inequality produced largely consistent results. However, as Luo (2018) noted, the Gini coefficient measures the dispersion of an income distribution, whereas income polarisation reflects clustering in the distribution. Hence, many

* Corresponding authors.
E-mail addresses: yyc961220@163.sufe.edu.cn (Y. Yu), 9120131013@nufe.edu.cn (Q. Li).
[1] Chen Zhang and Yangcheng Yu are joint first authors.

https://doi.org/10.1016/j.chieco.2022.101912
Received 18 October 2021; Received in revised form 1 November 2022; Accepted 20 December 2022
Available online 22 December 2022
1043-951X/

scholars have emphasised that further research on income polarisation should be conducted (Schettino, Gabriele, & Khan, 2021; Wang & Wan, 2015). A series of studies have shown that income polarisation hinders social and economic development. It reduces income mobility and hinders the formation of an "olive-shaped" social structure (Foster & Wolfson, 2010). In addition, it is not conducive to poverty reduction or economic growth (Clementi, Vasco, & Francesco, 2018; Ezcurra, 2009), and it may threaten social stability and even induce conflict (Duclos, Esteban, & Ray, 2004; Indra, Hartono, & Sumarto, 2019). Further research on income polarisation in China will be of great theoretical and practical value in helping the government to evaluate China's income gap and further reform its income distribution system.

Income polarisation is normally measured using household survey data. However, such data often provide insufficient coverage of top incomes, which may result in underestimation of the income gap (Luo, 2019). Conventional methods of collecting top-income data include backward extrapolation from inheritance tax data, inference from governments' administrative data, and prediction from individual behaviour (Hurst, Li, & Pugsley, 2014; Kopczuk, 2015). However, these methods are not feasible in China, because the country has no inheritance tax system, and its platforms for cross-departmental administrative information linkage are not mature (Li, Li, & Wan, 2020). For a more realistic and objective assessment of income polarisation in China, extra data on top-income earners are needed to compensate for the insufficient coverage of top incomes in household survey data. This approach is challenging but feasible. In 2013, the China Institute for Income Distribution (CIID) of Beijing Normal University established the Top Incomes in China (TIC) database, which has since been greatly expanded and refined. The central purpose of the TIC database is to collect as much data as possible on top incomes in various industries, thereby providing a rough overall picture of top incomes in China.

Using TIC data, studies have corrected for the Gini coefficient and top income shares (Li et al., 2020; Li, Yu, & Li, 2021), but the literature has rarely focused on the effects of top-income data on measurements of income polarisation. After the inclusion of top-income data, how do estimations of trends in income polarisation change? What are the reasons for these changes? Are there any differences in polarisation between sub-groups of residents? This paper aims to answer these questions. In order to obtain representative income data for Chinese residents, we combine data from the Chinese Household Income Project (CHIP) with TIC data for 2002, 2013 and 2018. The trends in, reasons for, and heterogeneity of income polarisation are explored using "identification and alienation" framework and the relative distribution method. Compared with previous studies, the contributions of this paper are as follows.

First, studies of China's top-income groups have primarily relied on data from the Hurun China Rich List and the Forbes China Rich List. However, these data sources focus on the super-rich, excluding other top income earners (Luo, 2019). This paper supplements prior studies by including data on rich but not super-rich people, extending previous measures of income polarisation. Second, there has been insufficient research on income polarisation in China since China reformed its income distribution system in 2013.[2] Some studies since 2013 have used recent household survey data to measure polarisation, but they have not incoporated top incomes (e.g. Luo, 2018; Wan & Clementi, 2021).[3] By combining the CHIP data with the TIC data, we are able to identify trends in the income polarisation of Chinese residents, especially after 2013, greatly assisting our efforts to determine the current income gap of Chinese residents. Third, most relevant research has presented only statistics on income polarisation, whereas we use the relative distribution method to explore changes in different parts of the income distribution over time. Finally, while most analyses of the heterogeneity of residents' income polarisation have focused on regional differences, we consider gender and education as well as region.

The remainder of the paper is organised as follows. In Section 2, we summarise the literature on income polarisation. Section 3 first briefly introduces the CHIP database and the TIC database, then discusses the methodology used to combine the two sources of data, and finally introduces our measures of income polarisation. Section 4 first explores the trends in, and reasons for polarisation among Chinese residents based on the DER index and the relative distribution method. Next, it reports our robustness tests. Finally, it discusses the heterogeneity of polarisation in terms of gender, education, and region. We conclude the paper and discuss its implications in Section 5.

2. Literature review

2.1. Concept and measurement of income polarisation

The concept of polarisation was first developed by sociologists and politicians in the 19th century. It was not noticed by economists until Levy (1987) found that no existing income inequality index could properly account for the disappearing middle class in the United States. Income polarisation was not formally analysed by economists until the 1990s (Wolfson, 1994; Esteban & Ray, 1994). Since then, the concepts of income polarisation and income inequality have often been used interchangeably to describe and analyse income distributions and income gaps. However, there are significant differences between them. The Gini coefficient—the most commonly used indicator of income inequality—measures the degree of income dispersion among residents. Income polarisation is

[2] In "Opinions on Deepening Reform of the Income Distribution System," issued by the State Council of the PRC on February 3, 2013, supporting low-income groups, expanding middle-income groups, and restricting top-income groups were established as the focus of China's income distribution system reform for the foreseeable future. The release of this important policy document, which had been in preparation for many years, marked a new round of reform of China's distribution system.

[3] The importance of adding top incomes to household survey data has been verified by many studies. For example, Luo (2019) found that China's Gini coefficient for 2013 increased by around 6% after including top-income data. Li et al. (2020) found that after incorporating top incomes, China's 2016 Gini coefficient increased from 0.46 to 0.64. However, few studies have discussed whether and how the inclusion of top incomes affects measurements of income polarisation.

also based on the income distribution, but it additionally accounts for aggregation at certain local points in the distribution. "Income bi-polarisation" describes a situation in which people diverge to the upper and lower tails of the income distribution and the middle-income class disappears. More generally, polarisation can also be multidimensional ("multi-polar polarisation"), such as when people converge to more than two groups.

Several indicators have been used to measure income polarisation. The first indicator is the FW index proposed by Foster and Wolfson (1992) to study bi-polarisation. The phenomenon of bi-polarisation is of great importance because the middle class plays a key role in economic growth, social development, and political stability (Easterly, 2001). Another indicator, which deals with multi-polar polarisation, measures the degree to which income distributions converge into minority groups. This measure does not limit polarisation to two groups (Esteban & Ray, 1994). Duclos et al. (2004) proposed a continuous polarisation index expressed by the density function of the income distribution, referred to as the DER index, which is the most commonly used measure of polarisation. In addition, Esteban and Ray (1994) proposed the "identification and alienation" framework for analysing income polarisation, which has since been further developed and enriched.

Recent studies have used the non-parametric relative distribution method proposed by Handcock and Morris (1998, 1999) to compare an income distribution at two points in time. By comparing the income distribution between a comparison population and a reference population, changes in the income distribution can be distilled into the location effect and the shape effect. "The location effect" refers to changes in the average or median income, whereas "the shape effect" refers to changes in variation, skewness, higher moments, and other characteristics of the distribution. Other relative distribution indices have been developed, including the median relative polarisation index (MRP). This index is symmetric and invariant of monotonic transformation in the distribution and can be further decomposed into the upper relative polarisation index (URP) and the lower relative polarisation index (LRP), which reflect the segments of the distribution above and below the median.

2.2. The level of income polarisation

Scholars have studied income polarisation in different countries and regions. The first wave of these studies focused on North America and Europe. Esteban, Gradín, and Ray (2007) analysed the income polarisation of five OECD countries using the extended DER index and found that the degree of income polarisation in the United States increased continuously between 1974 and 1997. During the same period, however, the degree of income polarisation in Canada remained stable. Foster and Wolfson (2010) also found that income polarisation showed a slight upwards trend in the United States in the 1980s, although they detected a slight decrease in Canada in the same period. Using the relative distribution method, Schettino and Khan (2020) analysed ASEC-CPS data from the United States from 1998 to 2018. They found a quasi-monotonic increase in income polarisation, because many middle-income Americans became low-income earners during this period. Europe is another hotspot for income polarisation research. Using data on 28 European Union countries and three non-European Union countries from 2004 to 2012, Wang, Caminada, Goudswaard, and Wang (2017) found that income polarisation, as measured by the DER index, remained fairly stable in Europe during this period, despite the Great Recession. In a case study of Russia, Nissanov and Pittau (2016) used the relative distribution method to analyse income polarisation from 1992 to 2008. They found an expansion of the middle class in the first six years, before incomes began to polarise in 1998, with polarisation reaching its highest level in 2008.

The second wave of studies focused on income polarisation in Asia, Africa, Latin America, and the Caribbean. Using micro-level data, Gochoco-Bautista, Bautista, Maligalig, and Sotocinal (2013) investigated changes in the degree of income polarisation in some Asian countries. They found that countries with higher GDP growth rates and higher GDP per capita tended to have a lower degree of income polarisation. In India, the degree of income polarisation decreased between 1983 and 1993 and increased between 1994 and 2005 (Motiram & Sarma, 2014). In a case study of Africa, Clementi, Dabalen, Molini, and Schettino (2017) used the relative distribution method to show that the degree of income polarisation of Nigerian residents increased significantly from 2003 to 2013. Latin America and the Caribbean are also important regions for studying income polarisation. Using a large set of household survey data, Gasparini, Horenstein, Molina, and Olivieri (2008) found high levels of polarisation in Latin America in the 1990s, reporting that 10 out of 16 countries experienced an increase in income polarisation during this decade. In a case study of Brazil, Clementi and Schettino (2013) used the non-parametric method to investigate residents' income distribution in 2001–2011. They showed that although the Gini coefficient decreased significantly during this period, income polarisation actually increased. Their conclusion was confirmed by a follow-up study (Clementi & Schettino, 2015), which found a clear rise in income polarisation among Brazilian residents after 2005, becoming even more pronounced after 2007.

2.3. Income polarisation in China

Research on income polarisation in China began later than research on this topic in other regions of the world, but it has gradually increased in recent years. In the following, we review and summarise the relevant research. Zhang and Kanbur (2001) was the first to use quantitative methods to study the income polarisation of Chinese residents. Their paper clarified the difference and relationship between income polarisation index and Gini coefficient. They found that the degree of income polarisation of Chinese residents increased from 1983 to 1995. Bennefond and Clément (2012) analysed CHNS 1989–2006 data and found clear evidence of income polarisation in China. The level of polarisation did not change substantially from 1989 to 1997, but it increased significantly from 1997 to 2006 due to an increase in middle-income and high-income earners. Based on data from CHIP 2002 and CHIP 2007, Wang and Wan (2015) found that the degree of income polarisation in China rose between 2002 and 2007. This increase was significant in eastern and central China but insignificant in western China. After distinguishing income by source, Wang and Wan (2015) found that the rise in polarisation was primarily driven by investment income, while business income helped to reduce the degree of polarisation. Schettino et al. (2021) used the CHNS 1989–2011 datasets to explore the trends in polarisation in China over 20 years. They found that income polarisation climbed in the early 2000s, peaked in the mid-2000s, and then started to decrease.

Due to limited access to recent household survey data, most research has only studied levels of and trends in income polarisation in China before 2013. Some studies conducted since 2013 have determined that income polarisation has tended to ease or decline in recent years. For example, Luo (2018) analysed CFPS 2012 and CFPS 2016 data and found that both the Wolfson index and the DER index declined between 2012 and 2016. Wan and Clementi (2021) analysed CHIP 1995–2018 data and found that the income polarisation of Chinese residents generally increased from 1995 to 2013, but a "historical reversal" in income polarisation occurred in 2013 and polarisation subsequently decreased until 2018. However, no consensus has been reached on differences and trends in income polarisation over time between sub-groups of residents, especially urban and rural residents. Many studies found that the degree of income polarisation was greater in rural areas than in urban areas (Bennefond & Clément, 2012; Wan & Clementi, 2021; Wang & Wan, 2015; Zhang & Kanbur, 2001). Contrary to these studies, Wang and Li (2013) found that the degree of bi-polarisation in rural areas was higher than that in urban areas between 1995 and 2005, while smaller than that in urban areas between 2008 and 2010. For the polarisation trends of rural residents, Zhang, Peng, and Kong (2019) analysed panel data from the National Fixed-site Rural Survey (NFRS) and found that income polarisation gradually increased in 2009–2015, but this polarisation was regionally heterogeneous. In contrast, Wan and Clementi (2021) found that the FW index decreased but the DER index remained unchanged from 2013 to 2018 among rural residents. This indicated that the level of bi-polarisation decreased while the level of multi-polarisation remained stable.

To summarise, scholars have conducted a series of fruitful studies on the levels of and trends in income polarisation around the world. Studies have addressed income polarisation in China, but few have focused on the period after 2013, when China reformed its income distribution system. In particular, researchers have not reached a consensus on differences in polarisation between rural and urban residents. Their inconsistent findings may be related to the insufficient coverage of top incomes in household survey data, because most top income earners live in urban areas. This paper explores the trends in, reasons for, and heterogeneity of income polarisation by combining household survey data and external top-income data, enhancing understanding of income polarisation in China and providing guidance for the further reform of its income distribution system.

3. Data and methodology

3.1. Data description

Economists have primarily used household survey data to measure polarisation. However, given the scant coverage by these data of the upper tail of the distribution, it is important to incorporate external top incomes into traditional survey data. In this study, survey data are obtained from the CHIP database, while top-income data are obtained from the TIC database. Both databases are constructed by the CIID, which guarantees the comparability of the data. The two databases are briefly described below.

To track the dynamics of income distribution in China, the CHIP research team conducted six waves of surveys. They collected income and expenditure information in 1988, 1995, 2002, 2007, 2013 and 2018, as well as other household and individual information. The datasets are commonly referred to as CHIP 1988, CHIP 1995, CHIP 2002, CHIP 2007, CHIP 2013 and CHIP 2018, and have provided information for many studies and public policy analyses.

Given the scarce coverage of top incomes in the CHIP data, the CIID set up a research team in 2013 and spent seven years constructing a special database (TIC) of external data on top incomes in China. The TIC database covers several types of top incomes from 31 provinces, autonomous regions, and municipalities across China (excluding Hong Kong, Macao, and Taiwan) from 2002 to 2019. The purpose of establishing this database is to collect information on top income earners from different industries and estimate their annual disposable income based on other information. The samples are weighted according to their types to provide a rough picture of China's top incomes (China Institute for Income Distribution, 2020).

The current paper explores the trends in, reasons for, and heterogeneity of income polarisation in China between 2002 and 2018 by combining the CHIP data with the TIC data for 2002, 2013, and 2018.[4] "Income" here refers to disposable income, which consists of wage income, business income, property income, and transfer income, as well as real subsidies and welfare. Imputed income from the services of certain assets, such as owner-occupied housing, is excluded. Our measure of income is household income per capita. The values are adjusted to 2018 constant prices by the Consumer Price Index and measured in units of 10,000 yuan.

Table 1 shows descriptive statistics for the CHIP data, the TIC data, and the magnate data. For the CHIP data, the sample sizes are 63,157, 62,101, and 69,411, respectively, in 2002, 2013, and 2018. The average annual income rose from 6944 yuan in 2002 to 20,575 yuan in 2013 and 27,822 yuan in 2018. The minimum incomes are below 0 in each of the three years because of negative business income. The Gini coefficient also shows a gradually increasing trend, rising from 0.4279 in 2002 to 0.4357 in 2013 and 0.4458 in 2018.

The TIC data cover four types of top income earners, namely magnates, the CEOs of listed companies, famous actors, and digital economy workers (including online writers and online celebrities), in 2013 and 2018, but the TIC data for 2002 only cover magnates. We also present summary statistics for the magnate data because we combine the CHIP data with the magnate data in the robustness checks. We show in Section 4.2 that the benefits of including more top-income samples and types (using CHIP+TIC) exceed the benefits of keeping top-income type consistent (using CHIP+Magnate).

[4] We exclude CHIP 2007 data from this analysis because the 2007 survey units, sampling schemes and survey ideas were different from those in other years.

Table 1
Descriptive statistics.

Data source	CHIP			TIC			Magnate[a]		
Year	2002	2013	2018	2002	2013	2018	2002	2013	2018
Obs	63,157	62,101	69,411	3447	17,534	28,190	3447	3631	3583
Mean	0.69	2.06	2.78	1078.83	145.48	188.92	1078.83	5471.11	9578.56
SD	0.60	1.79	2.75	1094.12	1239.03	2346.25	1094.12	9134.82	21,627.93
Max.	10.44	32.15	105.36	18,549.80	246,229.61	450,000.00	18,549.80	246,229.61	450,000.00
p90	1.47	4.34	5.70	2003.38	137.65	146.18	2003.38	9849.18	18,333.33
p50	0.49	1.54	2.06	763.63	45.00	34.75	763.63	3283.06	3750.00
p10	0.18	0.44	0.60	369.18	17.51	16.05	369.18	2012.01	3000.00
Min.	−0.06	−27.91	−21.60	56.92	13.13	12.00	56.92	21.18	36.45
Gini	0.4279	0.4357	0.4458	0.3946	0.7409	0.8287	0.3946	0.4595	0.5800

Units are measured in 10,000 yuan except for years, sample sizes and Gini coefficients.
[a] "Magnates" refers to all those named on the Hurun China Rich List or the Forbes China Rich List between 2002 and 2018. Some magnates may have relatively low incomes following crises such as corporate failure during the period. As we later explain, these relatively low-income samples are not included in the combined dataset.

3.2. *Combination of datasets*

Let us start with intuition before discussing the techniques in detail. Although household survey data are often criticised for their insufficient coverage of the upper tail of the income distribution, most studies have argued that such data adequately represent low incomes, middle incomes, and even high incomes (although not the highest incomes). In contrast, TIC data offer good coverage of top incomes, but their representation of relatively low incomes is questionable (Li et al., 2020). Hence, we form a nationally representative income distribution by combining low, middle and high incomes from the CHIP data with top incomes from the TIC data.

The primary problem when connecting the two datasets is choosing a splicing point. We follow Han and Cheng (2019) in treating the CHIP data as truncated income data, which means that the survey data are perfectly representative up to the survey maximum and fail to include any respondents beyond this point (Blanchet, Fournier, & Piketty, 2022a).[5] We then select samples with incomes higher than the survey maximum from the TIC data and consolidate the resulting top income data with theentire CHIP data. In other words, the sample splicing point is set at the maximum of the CHIP data. In the combined dataset, the samples with incomes below the survey maximum are from the CHIP dataset, and the samples with incomes above the survey maximum are from the TIC dataset. Therefore, the combination of datasets does not suffer from income overlap between sources, because the relatively low-income samples in the TIC data are dropped before integrating the TIC data with the CHIP data. A similar combination of survey data and high-income data was used in Alvaredo's (2011) study in Argentina.

Another issue regarding the combination of datasets is weighting adjustment, because the two datasets were collected according to different sampling methods. Our weighting calibration follows the following three steps. First, the samples in the CHIP data are weighted according to the population shares of urban and rural residents within each province in each year. Second, the sample weights in the TIC data are determined according to their types. For example, the weight for magnates is set at 1, because they are likely to represent the richest people in China. Other top incomes in the TIC dataset are weighted by type to represent the total population in each industry. The weighted TIC dataset roughly represents top incomes in China, especially in the upper tail (Li et al., 2020).[6] Third, when the two datasets are combined, we keep the weights in the TIC data unchanged and adjust the weights for the CHIP data to make the total weights for the combined dataset sum to the national population in the corresponding year.

Our examination of the trends in, reasons for, and heterogeneity of income polarisation in China is based on a combination of the CHIP data and the TIC data (CHIP+TIC).[7] In Section 4.2, we report two robustness checks using a combination of the CHIP data and the magnate data (CHIP+Magnate). This combination ensures the consistency of top-income types across time, but a key disadvantage is that there are fewer data and types of top incomes than in the combined dataset composed of the CHIP data and the TIC data.

3.3. *Measures of polarisation*

The Duclos–Esteban–Ray (DER) index, proposed by Duclos et al. (2004), is commonly used to measure the multipolar polarisation of income. The primary advantages of the DER index are that income is defined as continuous and the estimate does not require an

[5] We assume that the CHIP data are truncated data, but this assumption is still highly stylised. It is also possible that some high-income samples in the CHIP data suffer from income under-reporting and thus are not representative. We follow Piketty et al. (2019) in upgrading incomes above the 90th percentile in the survey data by a factor of 1.5 and combine the adjusted survey data with the TIC data in the same manner. We find that the trends in polarisation are still consistent with those observed by combining the raw CHIP data and TIC data. Therefore, we argue that under-reporting in the upper tail of the survey data does not affect our estimates of polarisation trends.

[6] Appendix Table 1A shows descriptive statistics for different types of top incomes in 2002, 2013 and 2018.

[7] Appendix Table 2A compares the top income shares calculated using CHIP+TIC with those calculated using the generalised Pareto interpolation method. The top income shares are very similar, which implies that the income distribution of our CHIP+TIC dataset is almost the same as the income distribution recovered by the generalised Pareto interpolation method.

arbitrary definition of the number of income groups. Consider two individuals with incomes x and y respectively. The interaction between the two components characterises the antagonism from x to y, which can be represented by the following function:

$$T(I, A) \tag{1}$$

where $I = f(x)$, $A = |y - x|$ and T is a continuous and non-decreasing function. A general polarisation index can thus be defined as the sum of all antagonisms characterising the income distribution:

$$P_\alpha(f) = \iint T(I, A) f(x) f(y) dx dy \tag{2}$$

Considering several axioms to determine the functional form of T, Duclos et al. (2004) proposed the following polarisation index:

$$DER = \iint f(x)^{1+\alpha} f(y) |y - x| dx dy \tag{3}$$

The parameter a reflects the degree of aversion to polarisation, which belongs within the interval [0.25, 1]. In particular, when $\alpha = 0$, the DER index ignores population clusters and is equal to the Gini coefficient. Higher values of α emphasise the growing importance of the formation of income groups in society.

According to Esteban and Ray (1994), the concept of polarisation has three important features: a small number of groups, a degree of homogeneity within each group (so-called identification) and a degree of heterogeneity between groups (so-called alienation). The DER index introduced by Esteban and Ray (1994) can be used to analyse polarisation via the "identification and alienation" framework. The degree of identification is expressed as $\tau(y) = f(y)^\alpha$. The degree of alienation is expressed as $I(y) = \int f(y)|x - y|dy$. The gap between income x and income y is represented by $|x - y|$.

The DER index can be further expressed by formula (4):

$$DER = \overline{\tau(y)} \times \overline{I(y)} \times (1 + \rho) \tag{4}$$

According to formula (4), the DER index is divided into three parts: identification, alienation, and correlation. $\overline{\tau(y)} = \int f(y)^{1+\alpha} dy$ indicates the average degree of social identification after aggregation, $\overline{I(y)} = \iint |x - y| f(x) f(y) dx dy$ indicates the average degree of social alienation after aggregation, and $\rho = cov(\tau(y), I(y)) / (\overline{\tau(y)} \times \overline{I(y)})$ indicates the degree of correlation between identification and aggregation.[8]

Apart from the DER index, distributional analysis of differences over time is often performed using the relative distribution method (Handcock & Morris, 1998, 1999). The non-parametric method tests a hypothesis regarding distributional differences by computing the fractions of the population in the comparison year that fall in each quantile of the population in the reference year. Let Y_0 be a random variable for the reference population and Y be a random variable for the comparison population; the median relative polarisation index (MRP) of Y with respect to Y_0 is defined as:

$$\text{MRP} = \frac{4}{n}\left(\sum_{i=1}^{n}\left|r_i - \frac{1}{2}\right|\right) - 1 \tag{5}$$

where r_i is the proportion of adjusted reference incomes that are below the i^{th} income in the comparison population. A positive MRP means that people are moving from the middle to one or both tails of the distribution, whereas a negative MRP implies that people are converging to the centre of the distribution.

The MRP index can be separated into a lower relative polarisation (LRP) index and an upper relative polarisation (URP) index to investigate the changes in overall polarisation due to incomes in the upper and lower tails of the distribution, respectively. These indices can be calculated by

$$\text{LRP} = \frac{8}{n}\left(\sum_{i=1}^{\frac{n}{2}}\left|\frac{1}{2} - r_i\right|\right) - 1 \tag{6}$$

$$\text{URP} = \frac{8}{n}\left(\sum_{i=1}^{n/2}\left|r_i - \frac{1}{2}\right|\right) - 1 \tag{7}$$

All of the three estimates range from -1 to 1, and it can be shown that MRP $= \frac{1}{2}(LRP + URP)$.

[8] For instance, consider two income sequences (1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10,000) and (2000, 2000, 2000, 2000, 2000, 8000, 8000, 8000, 8000, 8000). The levels of identification, alienation, and correlation are 0.7827, 0.3000 and − 0.0074, respectively, for sequence one. The levels of identification, alienation and correlation are 0.7562, 0.3000 and − 0.0000, respectively, for sequence two. The level of alienation is the same for the two sequences, but the DER index for sequence one (0.2331) is higher than that for sequence two (0.2268).

4. Empirical evidence using income data from China, 2002–2018

4.1. Changes in income polarisation after including top incomes

Table 2 presents summary statistics for the CHIP data and the CHIP+TIC data. After the inclusion of top incomes, the average incomes increase slightly in each of the three years because the top incomes account for a relatively small proportion of the whole population. The standard deviations and maximums increase sharply, while the minimums remain at −0.6, −279.1, and − 216.0 thousand yuan in 2002, 2013, and 2018, respectively.

Table 3 presents the DER indices as well as the identification and alienation indices for each of the three years. Using the CHIP data, the DER index increases significantly from 0.3177 in 2002 to 0.3215 in 2013, before declining slightly to 0.3209 in 2018. Using the combined dataset, the DER index shows a similar upwards trend between 2002 and 2013, but it continues to rise slightly, increasing from 0.3439 in 2013 to 0.3442 in 2018. According to the estimates from the two datasets, we can confirm that polarisation increases between 2002 and 2013; however, the change in polarisation between 2013 and 2018 remains unclear. Further statistical tests show that neither the downwards trend derived from the CHIP data nor the upwards trend derived from the combined data is statistically significant, even at the 10% significance level. Hence, we conclude that income polarisation remains stable between 2013 and 2018 even when top incomes are included.

Using the "identification and alienation" framework, we decompose the DER index into alienation, identification, and correlation. Based on the combined dataset, we find a persistent growth in alienation. The levels of alienation are 0.4485, 0.4994, and 0.5123 for 2002, 2013 and 2018, respectively, with increases by 4.81%, 14.62% and 14.92% compared with the CHIP data. The identification index calculated with the combined dataset shows a trend opposite to that for the CHIP data; it experiences a sharp decline from 2002 to 2013 but rises to 0.7442 in 2018. This implies that the degree of association within the identified groups narrows from 2002 to 2013 and then increases between 2013 and 2018.

Table 4 presents relative distribution indices to track the changes in the income distributions.[9] For the CHIP data, the MRP index in 2013 is significantly positive, which implies a dispersion from the middle of the income distribution towards either or both tails between 2002 and 2013. Meanwhile, the LRP index and URP index indicate that the lower tail of the distribution is positively polarised whereas the upper tail is negatively polarised. In 2018, all the indices are significantly negative when 2013 is used as the reference year. This indicates that people converge towards the centre of the distribution from both tails between 2013 and 2018.

When top incomes are included, the relative distribution indices retain the same signs. The MRP in 2013 remains positive but is larger than that derived from the CHIP data, which indicates a faster divergence towards either or both tails of the distribution between 2002 and 2013. The poor still become more polarised and the rich still become less polarised, but the pace is slower for both, as reflected by the lower LRP index and URP index in absolute values. Compared with the survey-based estimates, the corrected MRP index in 2018 shows a more obvious convergence to the middle of the income distribution. In addition, the convergence to the middle from the poor becomes slower while the convergence from the rich becomes faster after introducing the top incomes.

According to the DER indices and the relative distribution indices, we provide some explanations for the changes in income polarisation between 2002 and 2018. Using both the survey data and the combined dataset, we find that polarisation rises significantly between 2002 and 2013. The relative distribution method shows that the rich become less polarised during this period; thus, the increase polarisation is caused by the rise in polarisation among the poor. From 2013 to 2018, the DER indices calculated by the CHIP data and the combined data present opposite trends. However, neither change is statistically significant. All three of the relative distribution indices support the convergence to the centre of the distribution. Hence, the stabilised polarisation between 2013 and 2018 can be explained by the expansion of the middle-income group. In particular, the LRP index is lower than the URP index, which shows that the poor become less polarised more quickly than the rich did in this period.

A possible reason for the stabilisation of polarisation between 2013 and 2018 is the new round of income distribution system reforms implemented by the CCP in 2013. The Decision of the Central Committee of the CCP on Several Major Issues Concerning the Comprehensive Deepening of Reform, issued in 2013, pointed out the need to standardise the order of income distribution, improve the institutional mechanism and policy system for income distribution regulation and control, and gradually form an olive-shaped distribution pattern across society by increasing the income of low-income earners and expanding the proportion of middle-income earners. The continuous improvement of the institutional policy system and the steady advancement of income redistribution measures may have mitigated the rise of polarisation.

4.2. Robustness check

Our results in Section 4.1 are based on the CHIP+TIC dataset. In this section, we conduct two robustness checks to test whether our conclusions above are reliable. We also give some reasons for using the CHIP+TIC results (rather than using the CHIP+Magnate results) as our benchmark.

In robustness check I, we combine the CHIP data with only one type of top incomes in the TIC database—magnate. As mentioned in Section 3.1, four types of top incomes are included in the TIC 2013 and the TIC 2018, but only magnates are included in the TIC 2002. Hence, the combination of CHIP+Magnate ensures the consistency of the top-income data in our analysis.

[9] The income distributions are displayed in Appendix Fig. 1A.

Table 2
Summary statistics for the CHIP data and the CHIP+TIC data.

Data source	CHIP			CHIP+TIC		
Year	2002	2013	2018	2002	2013	2018
Obs	63,157	62,101	69,411	66,604	75,540	79,657
Mean	0.69	2.06	2.78	0.72	2.32	3.16
SD	0.60	1.79	2.75	7.57	55.09	115.91
Max.	10.44	32.15	105.36	18,549.80	246,229.61	450,000.00
Min.	−0.06	−27.91	−21.60	−0.06	−27.91	−21.60

Units are measured in 10,000 yuan except for years and sample sizes.

Table 3
DER index and its components.

Data source	CHIP			CHIP+TIC		
Year	2002	2013	2018	2002	2013	2018
Obs	63,157	62,101	69,411	66,604	75,540	79,657
DER	0.3177	0.3215***	0.3209	0.3252	0.3439***	0.3442
Alienation	0.4279	0.4357	0.4458	0.4485	0.4994	0.5123
Identification	0.8245	0.8309	0.8181	0.7924	0.7176	0.7442
Correlation	−0.0997	−0.1118	−0.1202	−0.0851	−0.0406	−0.0972

For the DER indices, *** indicates that changes in the indices are statistically significant at the 1% level compared with the indices in the reference year. We follow Wan and Clementi (2021) to use log income data with an additive median shift.

Table 4
Relative distribution indices.

Data source	CHIP			CHIP+TIC		
Year	2002	2013	2018	2002	2013	2018
Obs	63,157	62,101	69,411	66,604	75,540	79,657
MRP	–	0.0516***	−0.0340***	–	0.0528***	−0.0349***
LRP	–	0.1312***	−0.0400***	–	0.1308***	−0.0397***
URP	–	−0.0280***	−0.0281***	–	−0.0252***	−0.0301***

For the MRP, LRP, and URP indices, *** indicates that changes in the indices are statistically significant at the 1% level compared with the indices in the reference year.

In robustness check II, we retain the CHIP+TIC combination but recalculate the incomes of magnates before combining them with the survey data. In the TIC database, the annual incomes of magnates are estimated from their wealth, with an annual rate of return of 5% (Li et al., 2020). As this rate of return may be regarded as too low, we reset the annual return rate to 10% to test whether this affects our results.

Figure 1 presents the estimates based on the four datasets. In the first panel, we find that the DER index from the CHIP data shows an increasing trend between 2002 and 2013 but decreases slightly from 2013 to 2018 (although this is not significant). Analyses of the other three datasets reveal that the DER index keeps rising between 2002 and 2018. Statistical tests confirm that the changes in the DER index between 2002 and 2013 are significant, but the changes between 2013 and 2018 are non-significant (see Appendix Table 3A). The relative distribution indices from the two robustness checks have exactly the same signs as those in our benchmarking results. Hence, the two robustness checks support our conclusion regarding the trends in and reasons for income polarisation between 2002 and 2018.

We wish to emphasise three advantages of using CHIP+TIC rather than CHIP+Magnate. First, the inclines in the DER index are sharper using CHIP+TIC, but polarisation still remains stable between 2013 and 2018. As we focus on polarisation trends rather than levels, we feel that the benefits of including more top-income samples and types exceed the benefits of keeping top-income type consistent. Second, the MRP indices and LRP indices show a "square" shape, but the URP indices from the four datasets are not all the same. The CHIP data and robustness check I give similar results, but are different from those for CHIP+TIC. This indicates the importance of using CHIP+TIC, because using CHIP+Magnate cannot reflect changes in the upper tail of the distribution. Third, robustness check II gives the same URP indices as CHIP+TIC. This implies that the potential income underestimation for magnates does not affect the reasons for polarisation between 2002 and 2018 when we use CHIP+TIC.

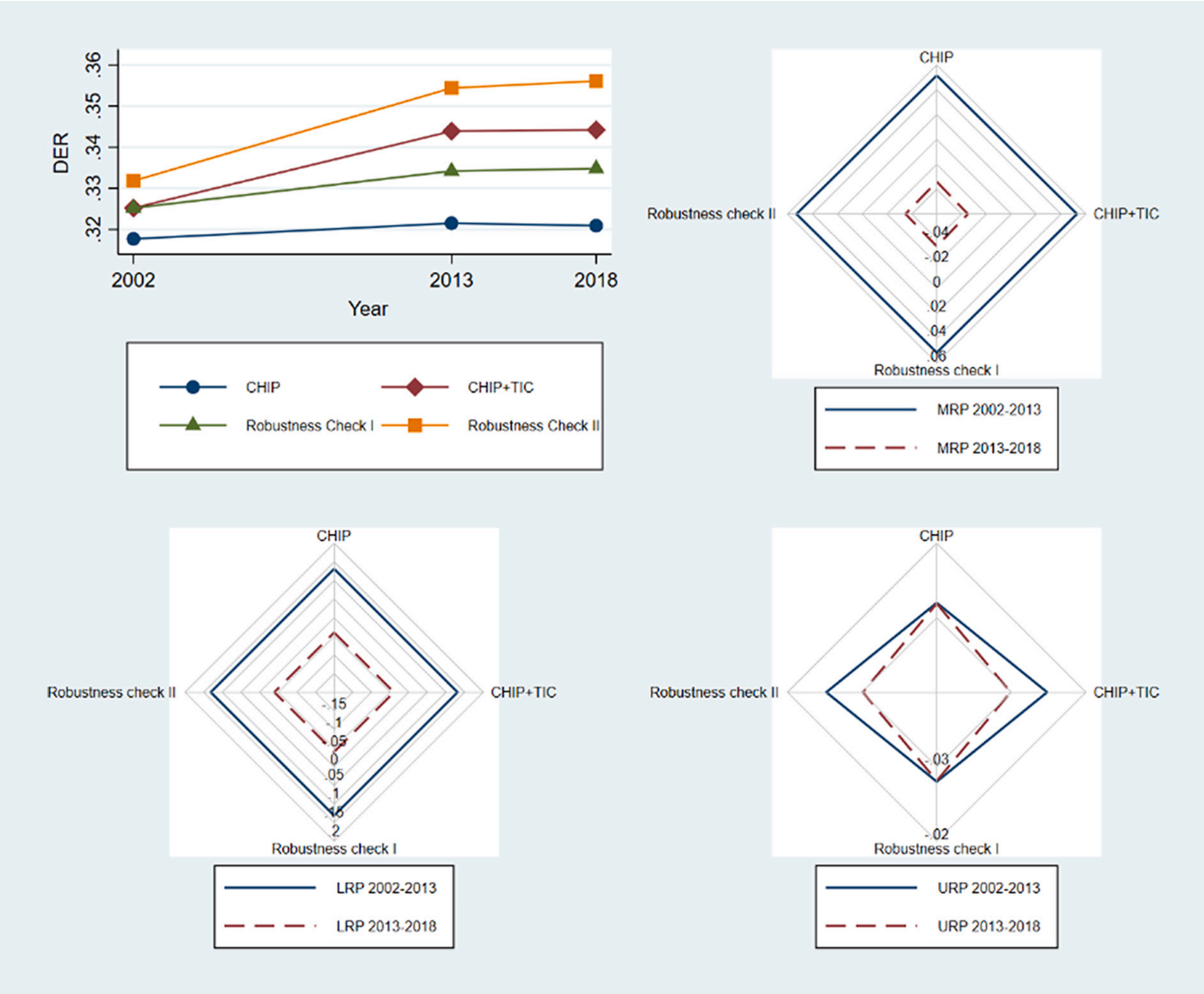

Fig. 1. Robustness check.

4.3. *Heterogeneity analyses*

In this section, we examine polarisation heterogeneity in terms of gender, education, and region using the CHIP+TIC dataset.[10] In Fig. 2, the DER indices for males are always greater than those for females between 2002 and 2018, which indicates higher levels of income polarisation among males. The polarisation of the male group increases between 2002 and 2013 but decreases in 2018. Although the DER index for females shows a consistently increasing trend, statistical tests show that the level of polarisation increases significantly from 2002 to 2013 but remains stable between 2013 and 2018. The MRP indices are positive in 2013 and negative in 2018 for both males and females. This indicates that people diverge from the middle of the distribution between 2002 and 2013 but tend to converge to the centre between 2013 and 2018. In 2012, the 18th CCP National Congress stressed the importance of narrowing the income gap. With the development and reform of economic and legal systems, males and females began to move from the tails of the income distribution to the centre, contributing to the expansion of the middle-income group.

In Fig. 3, the polarisation of people with college degrees (well educated) increases between 2002 and 2018, as reflected by the rising DER index. The positive MRP indices show that people are moving towards the tails of the income distribution at a low speed during this period. The level of polarisation among those who have not attended college (less educated) shows an upwards trend between 2002 and 2013 but remains stable between 2013 and 2018. Compared with that for well educated people, the DER index for less educated people is higher in 2002 but lower in 2013 and 2018. In addition, the negative MRP for less educated people in 2018 indicates an expansion of the middle-income group between 2013 and 2018. With the rise of the platform economy and the increasingly active urban informal job market, people without college degrees have received more income generation opportunities (Choudary, 2018). Some of them have even obtained significantly higher incomes and are gradually becoming part of the middle-

[10] Please refer to Appendix Table 4A for details.

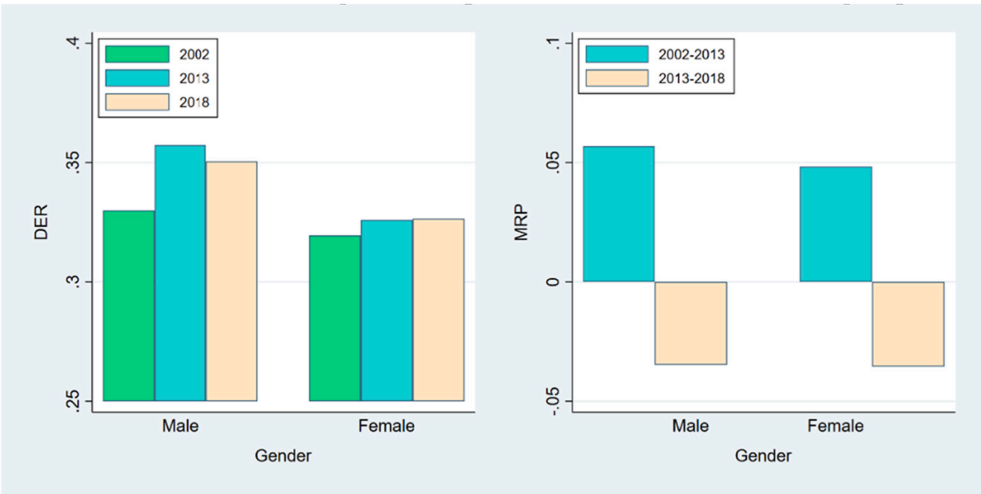

Fig. 2. Gender heterogeneity.
The income distributions are displayed in Appendix Fig. 2A.

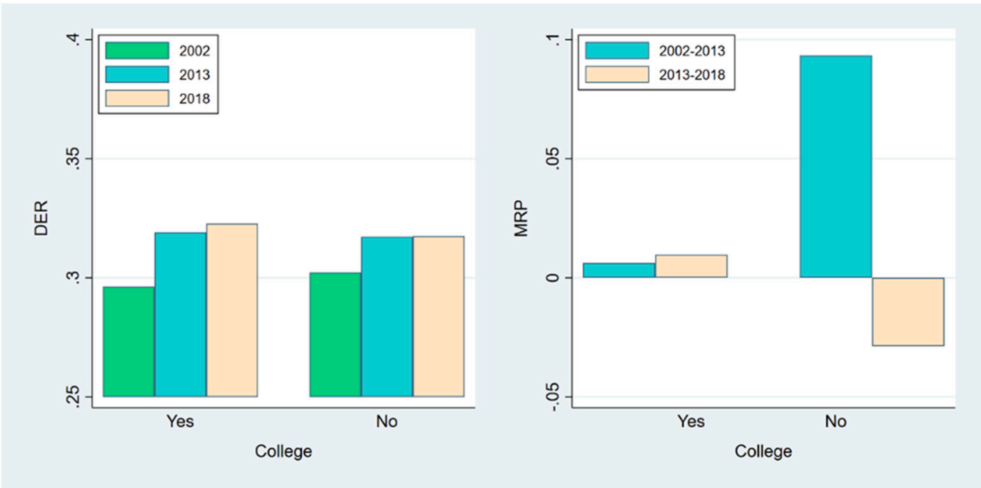

Fig. 3. Educational heterogeneity.
The income distributions are displayed in Appendix Fig. 3A.

income group.

In Fig. 4, we show that the DER index for rural residents is higher than that for urban residents in 2002. However, in 2013 and 2018, the DER indices for urban residents exceed those for rural residents. Our findings for regional heterogeneity are not consistent with those of some studies based solely on household survey data (e.g., Wan & Clementi, 2021). One noticeable difference is that our analysis uses a combination of CHIP data and TIC data. With the rapid development of the economy, a large number of top income earners has emerged in urban areas, but few of them live in rural areas. This explains the more serious income polarisation among urban residents. The MRP indices for urban and rural residents between 2002 and 2013 and 2013 to 2018 are consistently positive. A movement from the centre of the distribution to both tails is evident between 2002 and 2013; however, the diverging trend considerably slows between 2013 and 2018.

5. Conclusion and implications

China has made great advancements in economic growth since its reform and opening up, but these achievements have been accompanied by the risk of a widening income gap. Many studies have found stable Gini coefficients and top income shares in China in the past decade, but few have focused on income polarisation, particularly trends in polarisation after the reform of the income distribution system in 2013. This paper makes the first attempt to explore income polarisation among Chinese residents by combining

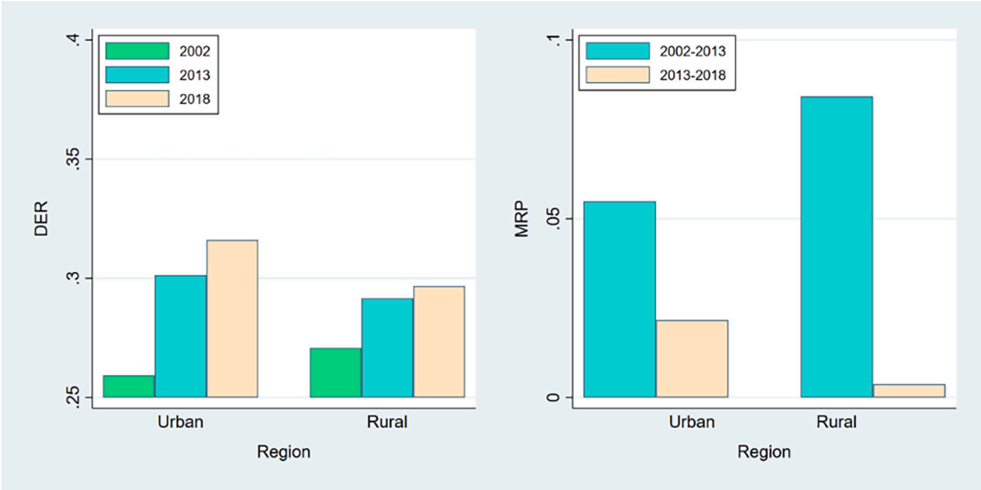

Fig. 4. Regional heterogeneity.
The income distributions are displayed in Appendix Fig. 4A.

household survey data with external top-income data. We incorporate the CHIP data with the TIC data for 2002, 2013, and 2018 to explore the trends in, reasons for, and heterogeneity of polarisation between 2002 and 2018. The main conclusions are as follows.

First, according to the CHIP data, the DER index increases significantly from 0.3177 in 2002 to 0.3215 in 2013 but declines slightly to 0.3209 in 2018. After including data on top incomes, the revised DER index unsurprisingly rises in all three years. However, the decreasing DER index between 2013 and 2018 derived from the CHIP data is reversed. Statistical tests confirm a significant rise in the DER index between 2002 and 2013, but neither the decrease nor the increase between 2013 and 2018 is statistically significant, even at the 10% level. Although household survey data are often regarded as unable to properly measure polarisation due to their scarce coverage of top incomes, we draw the same conclusion from the CHIP data and the combined dataset: income polarisation rises significantly between 2002 and 2013 but remains stable between 2013 and 2018.

Second, the positive MRP index supports the increase in polarisation between 2002 and 2013. The LRP index and the URP index show that the poor are positively polarised whereas the rich are negatively polarised. This indicates that the increase in polarisation between 2002 and 2013 is caused by a rise in polarisation among the poor. From 2013 to 2018, the change in the DER index is non-significant. The relative distribution indices imply an expansion of the middle-income group, with people converging to the middle of the income distribution from both tails. The LRP index and URP index are negative and the LRP index is higher than the URP index in absolute values. Compared to the rich, the polarisation of the poor declines more rapidly, which contributes to the stable polarisation during the period.

Third, heterogeneity analyses are conducted according to gender, education, and region. In terms of gender, men show higher levels of polarisation than women during the whole period. The size of the middle-income group increases between 2013 and 2018 for both males and females. In terms of educational difference, the income polarisation of people with college degrees is lower than that of people without college degrees in 2002, but the opposite is the case in 2013 and 2018. The change in income distribution for highly educated people is minor across the whole period. However, less educated people diverge greatly from the middle to the tails between 2002 and 2013, and converge to the centre of the distribution between 2013 and 2018. In terms of regional differences, the income polarisation of urban residents is lower than that of rural residents in 2002 but higher in 2013 and 2018. The polarisation of both groups climbs significantly between 2002 and 2013 and continues to rise slowly between 2013 and 2018.

Finally, we would like to state some comments. Combining the CHIP data and the TIC data, we find that income polarisation increases significantly between 2002 and 2013 but remains almost stable between 2013 and 2018. Our hypothesis is that the 2013 reform of China's income distribution system may have mitigated the rise of polarisation. However, the DER index shows a consistently increasing, although non-significant, trend between 2013 and 2018. We encourage more efforts to support the low-income group, expand the middle-income group, and restrict the top-income group to form an olive-shaped income distribution. Lastly, our combination of CHIP data and TIC data is still imperfect because the top incomes collected are from a limited range of industries. Apart from household disposable income per capita, the TIC database lacks other information on top incomes, limiting our ability to decompose their incomes by source or detect the drivers of income polarisation. It would be worthwhile to collect more data on top incomes in different industries to get a better overall picture of the top-income group in China.

Data availability

The data that has been used is confidential.

Acknowledgements

The authors are grateful for the data collected by the members of the CHIP team and the TIC team. Any remaining errors are solely ours. The views expressed in this paper are those of the authors and do not necessarily reflect the views of any related institutions. This research is supported by the Major Project of National Social Science Fund of China (# 19ZDA116), the National Natural Science Foundation of China (# 72203232), the National Natural Science Foundation of China (# 72141310), the National Natural Science Foundation of China (# 71773074), and the National Natural Science Foundation of China (# 72273081).

Appendix A. Appendix

Table 1A. Descriptive statistics for the sub-groups of the TIC data, 2002–2018.

Year	2002	2013				2018			
Type	Magnates	Magnates	CEO of Listed Companies	Famous actors	Digital economy workers	Magnates	CEO of Listed Companies	Famous actors	Digital economy workers
Obs	3447	3631	10,620	3136	147	3583	11,649	3727	9231
Mean	1078.83	5471.11	66.84	219.75	24.66	9578.56	85.35	984.70	43.36
SD	1094.12	9134.82	223.14	353.01	16.00	21,627.93	718.51	1531.11	52.25
Max.	18,549.80	246,229.61	20,350.88	3561.46	304.24	450,000.00	75,303.41	18,176.80	5343.89
p90	2003.38	9849.18	115.07	543.67	40.91	18,333.33	146.03	2428.80	74.34
p50	763.63	3283.06	44.65	96.96	20.31	3750.00	51.50	468.00	29.74
p10	369.18	2012.01	18.28	18.39	13.94	3000.00	18.49	64.00	15.31
Min.	56.92	21.18	13.13	13.13	13.17	36.45	12.00	14.40	12.00
Gini	0.3946	0.4595	0.4599	0.6245	0.2786	0.5800	0.5241	0.6196	0.4139

Units are measured in 10,000 yuan except for sample sizes and Gini coefficients.

Table 2A. Comparison with the generalised Pareto interpolation method.

	CHIP+TIC				Generalised Pareto Interpolation			
Year	Top 1%	Top 5%	Top 10%	Top 50%	Top 1%	Top 5%	Top 10%	Top 50%
2002	8.41	21.20	32.70	80.11	8.41	21.20	32.69	80.11
2013	15.53	26.96	37.56	82.57	15.57	27.00	37.60	82.62
2018	17.61	29.35	39.60	82.64	17.61	29.36	39.59	82.64

The estimates on the left are calculated using the combined dataset in this paper. The estimates on the right are computed using the generalised Pareto interpolation method (Blanchet, Fournier, & Piketty, 2022b).

Table 3A. Robustness check.

Data source	Indicator	2002	2013	2018
CHIP	Obs	63,157	62,101	69,411
	DER	0.3177	0.3215***	0.3209
	MRP	–	0.0516***	−0.0340***
	LRP	–	0.1312***	−0.0400***
	URP	–	−0.0280***	−0.0281***
CHIP+TIC	Obs	66,604	75,540	79,657
	DER	0.3252	0.3439***	0.3442
	MRP	–	0.0528***	−0.0349***
	LRP	–	0.1308***	−0.0397***
	URP	–	−0.0252***	−0.0301***
Robustness check I	Obs	66,604	65,731	72,993
	DER	0.3252	0.3342***	0.3348
	MRP	–	0.0516***	−0.0340***
	LRP	–	0.1312***	−0.0400***
	URP	–	−0.0280***	−0.0281***
Robustness check II	Obs	66,604	75,540	79,657
	DER	0.3318	0.3544***	0.3561
	MRP	–	0.0528***	−0.0349***
	LRP	–	0.1308***	−0.0397***
	URP	–	−0.0252***	−0.0301***

For the DER, MRP, LRP, and URP indices, *** indicates that the changes in the indices are statistically significant at the 1% level compared with the indices in the reference year.

Table 4A Heterogeneity analyses.

Heterogeneity	Type	Indicator	2002	2013	2018
Gender	Male	DER	0.3299	0.3573***	0.3505***
		MRP	–	0.0569***	−0.0347***
	Female	DER	0.3195	0.3259***	0.3265
		MRP	–	0.0483***	−0.0355***
College	Yes	DER	0.2963	0.3191***	0.3227***
		MRP	–	0.0062***	0.0097***
	No	DER	0.3022	0.3173***	0.3175
		MRP	–	0.0933***	−0.0287***
Region	Urban	DER	0.2593	0.3013***	0.3161***
		MRP	–	0.0549***	0.0216***
	Rural	DER	0.2708	0.2916***	0.2968***
		MRP	–	0.0842***	0.0037***

For the DER, MRP, LRP, and URP indices, *** indicates that the changes in the indices are statistically significant at the 1% level compared with the indices in the reference year.

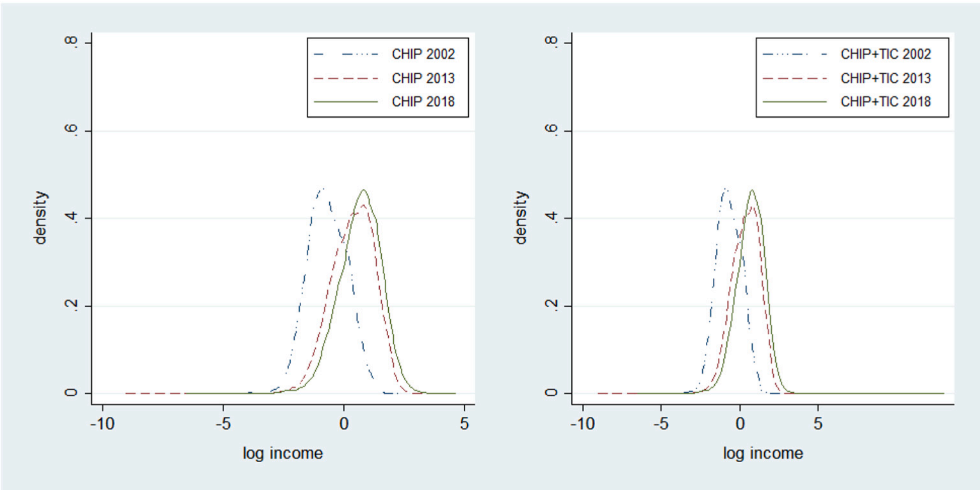

Fig. 1A Income distributions for the CHIP data and the CHIP + TIC data.

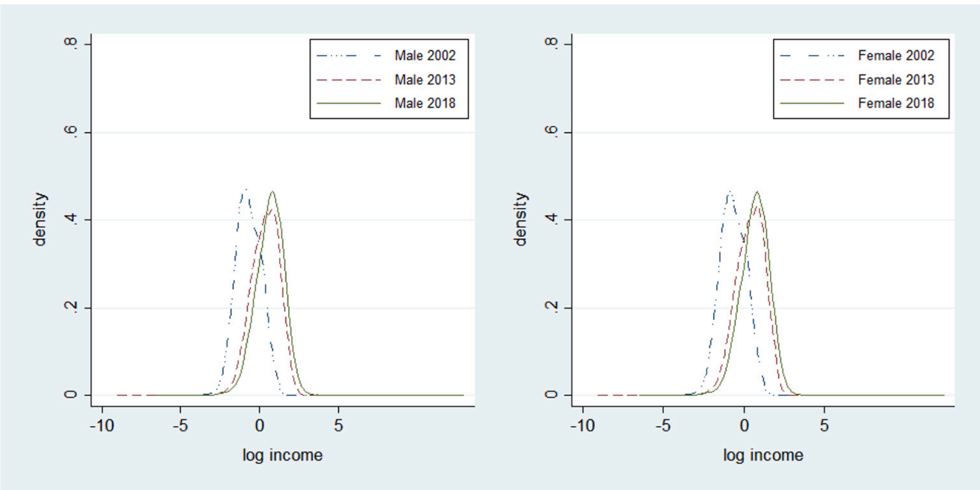

Fig. 2A Income distributions for males and females.

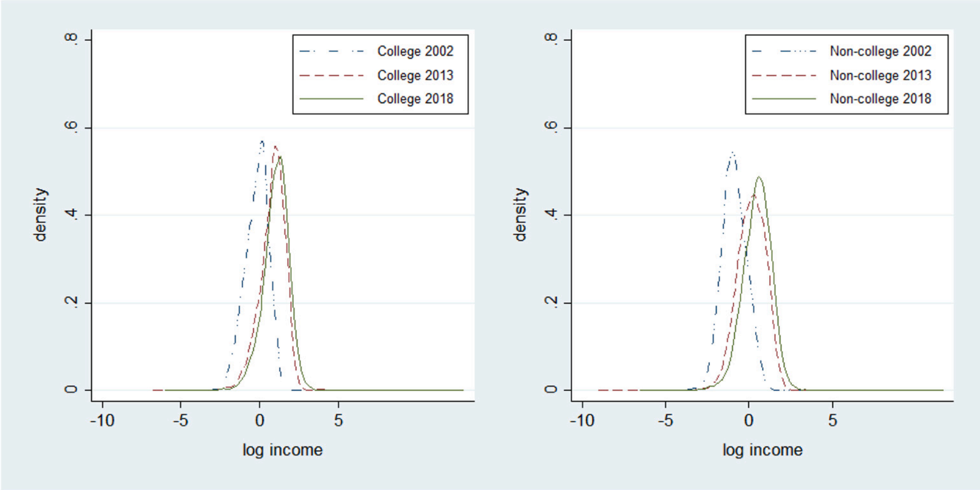

Fig. 3A Income distributions for residents with and without college degrees.

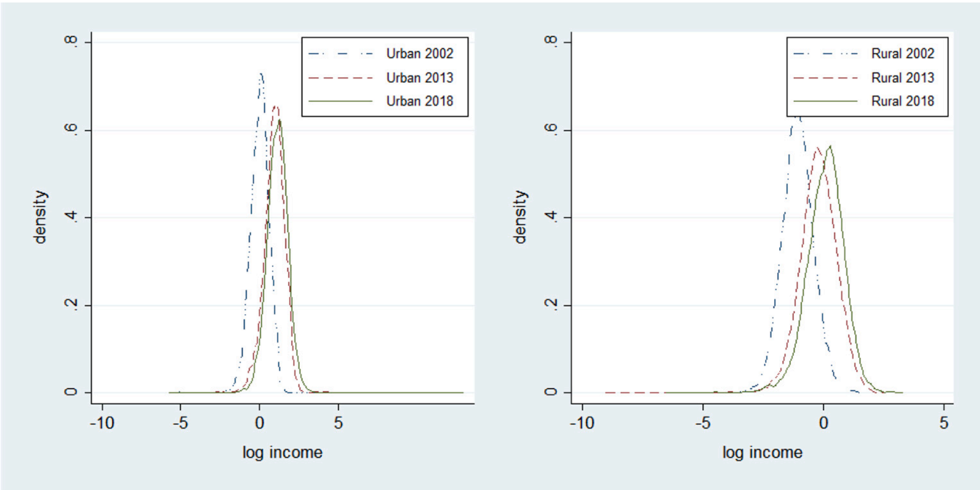

Fig. 4A Income distributions for urban and rural residents.

References

Alvaredo, F. (2011). A note on the relationship between top income shares and the Gini coefficient. *Economics Letters, 110*(3), 274–277.

Bennefond, C., & Clément, M. (2012). An analysis of income polarization in rural and urban China. *Post-Communist Economies, 24*(1), 15–37.

Blanchet, T., Fournier, J., & Piketty, T. (2022a). The weight of the rich improving surveys using tax data. *The Journal of Economic Inequality, 20*(1), 119–150.

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

China Institute for Income Distribution. (2020). *Annual report on income distribution of Chinese residents 2019* (pp. 267–280). Beijing: Social Sciences Academic Press.

Choudary, S. P. (2018). *The architecture of digital labour platforms: Policy recommendations on platform design for worker well-being* (p. 3). Geneva: ILO Future of Work Research Paper, No.

Clementi, F., & Schettino, F. (2013). Income polarization in Brazil, 2001-2011: A distributional analysis using PNAD data. *Economics Bulletin, 33*(3), 1796–1815.

Clementi, F., & Schettino, F. (2015). Declining inequality in Brazil in the 2000s: What is hidden behind? *Journal of International Development, 27*(7), 929–952.

Clementi, F., Dabalen, A. L., Molini, V., & Schettino, F. (2017). When the center cannot hold: Patterns of polarization in Nigeria. *Review of Income and Wealth, 63*(4), 608–632.

Clementi, F., Vasco, M., & Francesco, S. (2018). All that glitters is not gold: Polarization amid poverty reduction in Ghana. *World Development, 102*, 275–291.

Duclos, J. Y., Esteban, J., & Ray, D. (2004). Polarization: Concepts, measurement, estimation. *Econometrica, 72*(6), 1737–1772.

Easterly, W. (2001). The middle class consensus and economic development. *Journal of Economic Growth, 6*(4), 317–335.

Esteban, J., Gradín, C., & Ray, D. (2007). An extension of a measure of polarization, with an application to the income distribution of five OECD countries. *Journal of Economic Inequality, 5*(1), 1–19.

Esteban, J. M., & Ray, D. (1994). On the measurement of polarization. *Econometrica, 62*, 819–851.

Ezcurra, R. (2009). Does income polarization affect economic growth? The case of the European regions. *Regional Studies, 43*(2), 267–285.

Foster, J. E., & Wolfson, M. C. (1992). *Polarization and the decline of the middle class: Canada and the US*. Mimeo: Vanderbilt University.

Foster, J. E., & Wolfson, M. C. (2010). Polarization and the decline of the middle class: Canada and the US. *The Journal of Economic Inequality, 8*(2), 247–273.

Gasparini, L., Horenstein, M., Molina, E., & Olivieri, S. (2008). Income polarization in Latin America: Patterns and links with institutions and conflict. *Oxford Development Studies, 36*(4), 461–484.
Gochoco-Bautista, M. S., Bautista, C. C., Maligalig, D. S., & Sotocinal, N. R. (2013). Income polarization in Asia. *Asian Economic Papers, 12*(2), 101–136.
Han, X., & Cheng, Y. (2019). Does the "missing" high-income matter? -income distribution and inequality revisited with truncated distribution. *China Economic Review, 57*, Article 101337.
Handcock, M. S., & Morris, M. (1998). Relative distribution methods. *Sociological Methodology, 28*(1), 53–97.
Handcock, M. S., & Morris, M. (1999). *Relative distribution methods in the social sciences*. New York, NY: Springer-Verlag Inc.
Hurst, E., Li, G., & Pugsley, B. (2014). Are household surveys like tax forms? Evidence from income underreporting of the self-employed. *Review of Economics and Statistics, 96*(1), 19–33.
Indra, N. S., Hartono, D., & Sumarto, S. (2019). Roles of income polarization, income inequality and ethnic fractionalization in social conflicts: An empirical study of Indonesian provinces, 2002–2012. *Asian Economic Journal, 33*(2), 165–190.
Kopczuk, W. (2015). What do we know about evolution of top wealth shares in the United States? *Journal of Economic Perspectives, 29*(1), 47–66.
Levy, F. (1987). The middle class: Is it really vanishing? *The Brookings Review, 5*(3), 17–21.
Li, C., Yu, Y., & Li, Q. (2021). Top-income data and income inequality correction in China. *Economic Modelling, 97*, 210–219.
Li, Q., Li, S., & Wan, H. (2020). Top incomes in China: Data collection and the impact on income inequality. *China Economic Review, 62*, Article 101495.
Luo, C. (2018). The polarization of household income distribution and property distribution in China. *Statistical Research, 35*(11), 82–92 (In Chinese).
Luo, C. (2019). The missing top income group in household surveys and the underestimation of income inequality. *Economic Perspectives, 1*, 15–27 (In Chinese).
Motiram, S., & Sarma, N. (2014). Polarization, inequality, and growth: The Indian experience. *Oxford Development Studies, 42*(3), 297–318.
National Bureau of Statistics. (2021). *China household survey yearbook 2021* (p. 371). Beijing: China Statistics Press.
Nissanov, Z., & Pittau, M. G. (2016). Measuring changes in the Russian middle class between 1992 and 2008: A nonparametric distributional analysis. *Empirical Economics, 50*(2), 503–530.
Piketty, T., Yang, L., & Zucman, G. (2019). Capital accumulation, private property and rising inequality in China, 1978-2015. *The American Economic Review, 109*(7), 2469–2496.
Ravallion, M., & Chen, S. (2007). China's (uneven) progress against poverty. *Journal of Development Economics, 82*(1), 1–42.
Schettino, F., & Khan, H. A. (2020). Income polarization in the USA: What happened to the middle class in the last few decades? *Structural Change and Economic Dynamics, 53*, 149–161.
Schettino, F., Gabriele, A., & Khan, H. A. (2021). Polarization and the middle class in China: A non-parametric evaluation using CHNS and CHIP data. *Structural Change and Economic Dynamics, 57*, 251–264.
Tan, J., Zeng, T., & Zhu, S. (2018). *Earnings, income, and wealth distributions in China: Facts from the 2011 China household finance survey* (Working paper).
Wan, H., & Clementi, F. (2021). The long-term evolution of income polarisation in China, 1995-2018. *Journal of Development Studies, 57*(11), 1945–1972.
Wang, C., & Li, M. (2013). Development of China's middle-income group under the effect of polarization. *The Journal of Quantitative & Technical Economics, 30*(6), 51–64 (In Chinese).
Wang, C., & Wan, G. (2015). Income polarization in China: Trends and changes. *China Economic Review, 36*, 58–72.
Wang, J., Caminada, K., Goudswaard, K., & Wang, C. (2017). Income polarization in European countries and Europe wide, 2004–2012. *Cambridge Journal of Economics, 42*(3), 797–816.
Wolfson, M. C. (1994). When Inequalities Diverge. *American Economic Review, 84*, 353–358.
Zhang, C., Peng, C., & Kong, X. (2019). The trend and decomposition of household income polarization in China: Evidence from fixed observation points in rural China. *Studies in Labor Economics, 7*(2), 21–41 (In Chinese).
Zhang, X., & Kanbur, R. (2001). What difference do polarization measures make? An application to China. *Journal of Development Studies, 37*, 85–98.