Journal of Asian Economics 81 (2022) 101498

Contents lists available at ScienceDirect

Journal of Asian Economics

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

Inequality of opportunity in earnings in rural China

Xinjie Shi [a,b,c]

[a] China Academy for Rural Development, School of Public Affairs, Zhejiang University, Hangzhou, China
[b] Australian Centre on China in the World & Crawford School of Public Policy, The Australian National University, Canberra, Australia
[c] Institute for Common Prosperity and Development, Zhejiang University, Hangzhou, China

ARTICLE INFO

Keywords:
Inequality of opportunity
Earnings
Rural China

ABSTRACT

This paper explores the role of 'inequality of opportunity' in individual earnings in rural China, which is attributed to 'circumstantial' factors over which individuals have no control, including family background, gender, ethnic minority status and region of birth. These circumstances are contrasted with 'efforts' or choices that individuals make, which also impact on their individual earnings. Utilising the China Labour-force Dynamics Survey (CLDS) 2014, I measure the share of inequality of opportunity in the overall inequality of individual annual earnings for the entire sample and for each of four ten-year birth cohorts in rural China. The empirical results revealed that the share of inequality of opportunity in individual earnings for the full sample is 20.4% of the GE(0) coefficient. The adoption of machine learning methods provides a wide range of estimates between 16.4% (regression tress) and 25.4% (forests). Across all birth cohorts, gender is consistently the largest single contributor to inequality of opportunity, while family background is relatively more important for younger cohorts. A closer investigation indicates that those who find themselves in the worst circumstances are likely to exert lower level of effort, not because they don't want to try harder, but because their circumstances prevent them from doing so.

1. Introduction

Since 1978, China's economic reforms have led to rapid economic growth, which has long been hailed as one of the biggest contributors to the decrease in the global rate of extreme poverty. But there has also been an accompanying downside: a widening income gap. Although the economy has slowed in recent years, income inequality remains high in China. According to Xie and Zhou (2014), China's Gini coefficient nearly doubled from around 0.30 in 1980 to 0.55 in 2012. There may be no other case where a nation's income distribution has deteriorated so much and so dramatically (Naughton, 2007). A rapidly growing literature has examined the determinants of China's high income inequality, of which a substantial part is due to regional disparities (Fan et al., 2009; Gustafsson & Li, 2002; Kanbur & Zhang, 2005), and the rural-urban gap, which is linked to the *hukou* system (Li & Gibson, 2013; Sicular et al., 2007; Yang, 1999). Other factors include gender (Chen et al., 2013; Golley et al., 2019; Matthews & Nee, 2000), political status (Morduch & Sicular, 2000; Walder, 2002), and differences in human capital (Fleisher, Li, & Zhao, 2010).

Inequality in China is far more complex than just the simple rural-urban dichotomy. Even within rural areas, inequality between villages and within villages has increased greatly (Zhou et al., 2008). There is an extensive literature focusing on rural China, examining two main issues: (i) estimations of the level of and changes in inequality, and (ii) the key factors contributing to these trends. On the first issue, despite the diverse data used in the literature, there is broad consensus that rural inequality in China has increased

E-mail address: shixinjie1990@126.com.

https://doi.org/10.1016/j.asieco.2022.101498
Received 29 June 2021; Received in revised form 6 February 2022; Accepted 26 May 2022
Available online 6 June 2022

during the reform period (Benjamin et al., 2005; Gao, Liu et al., 2019; Ha et al., 2016; Liu & Sicular, 2009; McKinley, 2016; Wan & Zhou, 2005). On the second issue, many empirical studies have highlighted the crucial contribution of off-farm income to the increase in rural inequality over time (Benjamin et al., 2005; Bonnefond & Clément, 2012; Howell, 2017; Kung & Lee, 2001; Scharf & Rahut, 2014; Shi et al., 2010). Clément (2016) further points out that while the development of the off-farm sector in rural China was a pivotal factor in explaining the increase in rural inequality until the mid-2000 s, the large-scale diffusion of such off-farm employment partly accounts for the decline in rural inequality since then.

There has been no research to my knowledge, however, on whether the income distribution in rural China is fair or not – that is, whether it constitutes equal opportunity, or 'inequality of opportunity'.[1] A rapidly growing literature addresses this issue from the basic premise that an individual's achievements should depend only upon his or her efforts (and choices), and not on predetermined circumstances over which he or she has no control.

This idea became prominent in the late 1960s and early 1970s when researchers identified different returns arising from identical level of efforts by individuals with different backgrounds (Bowles, 1972; Hanoch, 1967; Weiss, 1970). This literature initiated a new agenda where researchers focused on how family background mattered for an individual's overall economic achievements (e.g. earnings). Analyses of this kind of inequality due to circumstances – called inequality of opportunity—have flourished since the pioneering work of Roemer (1993, 1998), who formalized the concept of unequal opportunities and separated the determinants of an individual's 'advantage' (i.e. desirable outcomes, such as incomes or educational attainments) into 'circumstances' that are exogenous to the individual—for instance, their gender, place of birth and the socioeconomic status of their parents— and 'efforts' that are under the control of the individual—for example, their own educational achievement and their choosing to work in the off-farm sector in the context of rural areas.

The recent popular saying *pingdie* in Chinese –'daddy is the key'—indicates that family background, rather than fair competition among individuals (that is, access to equal opportunities), has become a significant part of individual success in China – and there is ample evidence to support this point (Qin et al., 2016; Song et al., 2016; Zhang & Xie, 2016). This idea clearly extends beyond China, as one's father's education and occupation are two of the key circumstances used in the international empirical literature on inequality of opportunity (Ferreira & Gignoux, 2011; Peragine & Ferreira, 2015; Singh, 2012). The reason is that well-educated and wealthy parents are more likely to invest in educating their children and make use of their social networks to benefit their children as they seek to enter the labour market.

But there is more to inequality of opportunity than just family background. Regional disparities have been identified as a major contributor to income inequality due to a widening divide between industrially developing areas, mostly near the coast or large cities, and areas mainly relying on agriculture, mostly inland and far from major industrial activity (Fleisher et al., 2010; Li & Gibson, 2013; Sicular et al., 2007; Zhou et al., 2008). It is inevitable that a country as large as China will have large spatial differences in socio-economic development. Income gaps between ethnic groups in rural China also exist, with Han Chinese maintaining their traditional lead, as shown in Gustafsson and Shi (2003) and between genders with men more likely to attain higher income than women, as shown in Matthews and Nee (2000) and Hannum (2005). One's age – or birth cohort – is another example of a circumstance over which individuals have no control, and which clearly affects one's earning capacity. This is particularly important in the context of rural China as people born in different periods have faced very different opportunities, for example in terms of access to education and employment opportunities.

Beyond these circumstances, income inequality in rural China, as elsewhere, is also associated with 'efforts' that individuals may exert to influence their own earnings (although as it turns out below, a substantial part of that effort, is in fact influenced by their circumstances). Off-farm employment (and/or migration) is an important mechanism through which one can attain higher income compared to working on the farm (Shi, 2022). One's education—which is well known as one of the most crucial human capital investments—is also likely to be an important choice for individuals to raise income. Another 'effort' variable that seems to matter in the context of rural China is marital status. While the majority of the current literature on inequality of opportunity has not included efforts in the framework, this paper attempts to look closely at the (highly complex) relationship between circumstances and efforts and to examine whether individuals from less fortunate circumstances can exert more effort to offset their circumstances, or instead, whether efforts and circumstances are complementary – and therefore combining to make inequality higher, not lower.

This study uses the China Labor-force Dynamics Survey (CLDS) produced by Sun Yat-sen University in Guangdong, China to investigate inequality of opportunity in earnings in rural China, identifying a key set of circumstances that determine individual annual earnings – and hence a (lower-bound) share of inequality of opportunity in overall earnings inequality. It then identifies a (sub)set of 'effort' variables, which are shown to be at least partially determined by circumstances, but also direct contributors to earnings inequality in their own right. This allows a partial (albeit incomplete) response to the important question of whether and how the least fortunate individuals can offset their circumstances. There is no straightforward answer to this question, but this paper makes a start.

The rest of the paper proceeds as follows. Section 2 reviews the burgeoning literature on inequality of opportunity in a number of nations, including China. Section 3 introduces the conceptual framework and the methodology for the estimation of inequality of opportunity, and for the incorporation of efforts into the analysis. Section 4 describes the data, and presents the empirical results on inequality of opportunity in earnings in China. Section 5 looks more closely at the relationship between efforts and circumstances and the question of whether individuals can overcome their poor circumstances by exerting more effort. Section 6 concludes.

[1] Knight (2017) points out the importance of extending inequality research and research instruments towards an economics of fairness and unfairness in China.

2. Background

In contrast to the broad inequality literature, empirical studies on 'inequality of opportunity' are still relatively rare although growing rapidly. Building on the ideas of a number of philosophers (Arneson, 1989; Cohen, 1989; Dworkin, 1981) who argued that not all differences in particular outcomes or 'advantage' (e.g. incomes, or educational attainments) are unacceptable, an economics literature has evolved that begins with the partitioning of the observed inequality in a particular economic outcome into two components. The first component (acceptable inequalities) stems from a set of factors for which individuals can be held responsible. The second component (unacceptable inequalities) – defined as 'inequality of opportunity' – is attributable to factors over which individuals have no control.

The pioneering work of Roemer (1998) defines these two types of factors as 'efforts' (for example, how long or hard one studies or whether one chooses to find off-farm employment) and 'circumstances' (such as gender, ethnic minority status, or family background) respectively. The main idea of the ex-ante approach[2] to measuring inequality of opportunity is that there is 'equality of opportunity' if a particular outcome (say, earnings) is distributed independently of circumstances, in other words, individuals with identical 'efforts' would achieve the same outcome, regardless of their circumstances. Specifically, Roemer partitions the population into 'types' formed by individuals endowed with the same set of circumstances. For example, two circumstances–gender (male and female) and fathers' education (high level and low level) – will yield four types.[3] The overall inequality in income can be decomposed into a within-type component – which can be explained by variation within each type, attributed to 'effort'– and a between-type component – which is calculated using the mean levels of the outcome for each type as an absolute measure of inequality of opportunity. In the literature, 'relative inequality of opportunity' (referred to as IOR below)—the share of absolute inequality of opportunity to overall inequality—is more widely used for comparison across different datasets and countries.

Building on Roemer's ideas, many papers have estimated inequality of opportunity in different country settings, using a range of parametric and non-parametric techniques. Checchi and Peragine (2010), for instance, use a non-parametric approach to measure opportunity inequality in Italy. They show that the inequality of opportunities generated by family background – proxied by parents' education –accounts for 14.78% of overall inequality, with the South region being characterized by a higher degree of opportunity inequality than the North region. Lefranc, Pistolesi, and Trannoy (2008) use stochastic dominance rankings of distributions conditional on types—which are defined by parental education and occupation—as a measure of inequality of opportunity in nine developed countries during the 1990 s. They find that the U.S. and Italy are the most unequal countries in terms of both outcomes and opportunities (compared with other countries including Belgium, France, West Germany, Great Britain, the Netherlands, Norway, and Sweden.). However, non-parametric approaches like these only work when there are relatively few types, and relatively large datasets. As the number of types increases, the frequency of sample observations per type tends to diminish quite rapidly. Such a restriction demands a large sample for each type, otherwise, it is likely to lead to an underestimate of inequality of opportunity.

Largely due to data limitations, the majority of studies use parametric methods instead. Ferreira and Gignoux (2011), for example, estimate a linear model of earnings as a function of circumstances to construct a simple scalar measure of inequality of opportunity for six Latin American countries. They find that the share of inequality of opportunity (IOR) ranges from one quarter to one half of overall inequality. Singh (2012) adopts the same approach to estimate the inequality of opportunity in India that is associated with several circumstances, including parental education, parental occupation, caste, religion, and place of birth. The relative shares of inequality of opportunity across four cohorts (21–30, 31–40, 41–50, 51–65 years) vary from 18% to 26% of overall earnings inequality for urban India, and from 16% to 21% for rural India.

The summary of the literature above presents *overall* measures of inequality of opportunity based on a set number of circumstances. This literature recognises that more circumstances would increase that overall share (i.e. the lower bound result, discussed further below). There has also been much interest in the partial contributions of each circumstance. Bourguignon et al., (2007) made the first attempt to estimate the contribution of each of the circumstance variables to earnings inequality using a parametric approach. They measure the overall inequality of opportunity in Brazil attributable to five observed circumstances, which they further decompose into partial contributions from each of those five circumstances. However, a corrigendum to this paper (Bourguignon et al., 2013) conceded that their approach was statistically flawed. Another option for estimating partial inequality of opportunity is a Shapley-value decomposition (Shorrocks, 1982, 2012), as recently adopted in Björklund et al., (2011), in their analysis on the contributions of various circumstances on long-run income in Sweden.[4] While acknowledging that there are still technical problems associated with this approach, they find that several circumstances, including parental income and own IQ, are important for long-run income inequality, but that variations in individual effort account for the most part of that inequality. In most of these papers, the role of 'effort' is ignored—or subsumed— in the framework of estimating inequality of opportunity. Bourguignon et al. (2007) attempted to

[2] In contrast, an 'ex-post' measure of inequality of opportunity corresponds to inequality within 'tranches', which is defined in terms of one's relative position in the effort distributions across types (Checchi & Peragine, 2010)

[3] i.e, male individuals whose father has high-level education, male individuals whose father has low-level education, female individuals whose father has high-level education, female individuals whose father has low-level education.

[4] They go further on the contribution of 'effort'.

decompose the effect of circumstances into a direct effect on earnings and an indirect component, which works through the 'effort' variables – however they later conceded that this approach was also flawed. As a result, 'efforts' have tended to be overlooked in the literature – which is unfortunate, as there is much to learn from bringing them back in (while recognising the technical problems of doing so).

To the best of my knowledge, there are only several papers in this literature that focus specifically on China (Shi, 2019; Song & Zhou, 2019; Dai & Li, 2021; Yang et al., 2021; Liu et al., 2022). For instance, Zhang and Eriksson (2010), using data from the China Health and Nutrition Survey collected from nine provinces during the period 1989–2006, find that China has a rising degree of inequality of opportunity in incomes that is largely mirrored by the increase in income inequality. Parental income and parents' type of employer explain about two-thirds of the total inequality of opportunity, whereas parental education plays a minor role. Golley and Kong (2018) measure the inequality of opportunity in individual educational outcomes in aggregate and for each of ten birth cohorts, utilizing the China Family Panel Studies survey for 2010. Their results are based on a parametric approach, and show that the relative index of inequality of opportunity ranges from 25.3% to 42% across the different cohorts. They further show that the *Hukou* system is the dominant circumstance variable in determining the educational outcomes, with father's education, birth cohort, province, parents' Communist Party membership, gender, family size and ethnicity also playing important roles. Using the Survey of Women's Social Status in China (2010), Golley et al. (2019) reveal a considerable relative share of inequality of opportunity in nationwide individual labour earnings of 25%, with gender being the number one contributor among the set of circumstances. Their further investigation into the roles of circumstances and efforts also reveals notable gender differences – circumstances matter more for women, not only directly but also indirectly through their effect. However, neither of these papers focus on *rural* China.

3. Analytical framework: circumstances, effort and inequality of opportunity

Consider a finite population of individuals indexed by $i\epsilon\{1,\ldots,N\}$, each of whom has attained a set of attributes $\{y_i, C_i, E_i\}$, where y denotes a level of income, C denotes a vector of circumstances, which lie beyond the control of the individual, and E denotes a vector of effort, which can be affected by individual choice. C_i consists of J elements corresponding to each circumstance j, and each element C_i takes on a finite number of values, x_j, $\forall i$. This helps partition the population into K types, given by $\Pi \in \{T_1, T_2, \ldots, T_k\}$. T denotes the specific type with distributions $\{y_i^k\}$ where individuals have identical circumstances: $C_t = C_j, \forall t, j | t \in T_k, j \in T_k, \forall k$[5] This results in the maximum possible number of types that is given by $\overline{K} = \prod_{j=1}^{J} x_j$.

Following Roemer (1998), Bourguignon et al. (2007), and Ferreira and Gignoux (2011), the parametric approach starts with the following function.

$$y = f(C, E, u) \tag{1}$$

Where y denotes individual earnings and is a function of circumstance variables C, effort variables E and other unobserved determinants u.

Note that circumstance variables are economically exogenous, but that effort itself can depend on circumstances, as well as other random factors. In this case, (1) can be rewritten as:

$$y = f(C, E(C, e), u) \tag{2}$$

Given that the distribution of earnings is independent of circumstances $F(Y|C) = F(Y)$, Roemer's concept of equality of opportunity would attain when the following two conditions hold (Bourguignon et al., 2007):

(i) $\frac{\partial f(C,E,u)}{\partial C} = 0, \forall C$, circumstances have no direct effect on advantages conditional on efforts.
(ii) $G(E|C) = G(E), \forall E, \forall C$, efforts should be distributed independently from circumstances. In other words, circumstances have no causal effect on efforts.

To measure inequality of opportunity is therefore to measure the extent to which $F(Y|C) \neq F(Y)$. As stated in Section 2, the recent literature mainly comprises two different approaches—parametric and non-parametric—to the measurement of inequality of opportunity. This paper adopts the more common parametric approach. An empirically suitable approximation to Eq. (2) can be obtained by log-linearization:

$$ln(y) = \alpha + \beta C + \gamma E + u \tag{3}$$

$$E = \lambda C + e \tag{4}$$

[5] For instance, in the case of only two circumstances including gender (male and female) and *hukou* status at birth (rural and urban), we obtain four types including rural male (T_1), rural female (T_2), urban male (T_3) and urban female (T_4). Individuals t and j in a specific type (say T_1) have identical circumstance $C_t = C_j$.

where β and γ are two vectors of coefficients, and λ is a matrix of coefficients indicating the mechanism through circumstances affect efforts. Substituting (4) into (3) generates the reduced form of the structural model:

$$ln(y) = \alpha + C(\beta + \gamma\lambda) + \gamma e + u \tag{5}$$

which can be simply estimated by OLS as follows:

$$ln(y) = \alpha + \rho C + \varepsilon \tag{6}$$

where $\rho = \beta + \gamma\lambda$, encompasses both the direct effect of circumstances on income (y), and the indirect effect of circumstances through efforts, and $\varepsilon = \gamma e + u$,denotes the random error term. Using the estimated coefficients ρ and the actual values of circumstances, one can construct a parametric estimate of the smoothed distribution as follows[6]:

$$\{\widehat{y}_i\}, \text{ where } \widehat{y}_i = exp[C\widehat{\gamma}] \tag{7}$$

The parametrically smoothed (direct) estimates for inequality of opportunity indices are then defined as follows[7]:

$$IOA_D = I(\{\widehat{y}_i\}), \quad IOR_D = I(\{\widehat{y}_i\})/I(\{y_i\}) \tag{8}$$

I employ mean log deviation, GE(0), which is decomposable and commonly used for estimating inequality of opportunity (Ferreira & Gignoux, 2011). It is worth noting that the vector of observed circumstances is only a subset of circumstances that influence individual earnings. The estimated $\widehat{\rho}$ in Eq. (6) will be biased if some omitted circumstances that influence income may also influence the level of effort or if some unobserved circumstances are correlated with the observed circumstances. Furthermore, an instrumental variable strategy is not applicable in this framework since 'it is difficult to conceive of correlates of the circumstance variables that would not themselves have any direct influence on earnings' (Bourguignon et al., 2007). In addition, to calculate the index of inequality of opportunity, we focus on more than one circumstance variable, suggesting that an instrumental variable strategy is unlikely to succeed.

The current literature responds to this issue in two ways. First, most studies argue that this is not important since the main focus is on the overall measure of inequality of opportunity rather than the casual relationship between a given circumstance and the outcome. As lower-bound estimates of the 'true inequality of opportunity,' IOR and IOA would increase if more circumstance variables were added to Eq. (6), and the 'true' inequality of opportunity would be measured if all circumstances could be observed (Ferreira & Gignoux, 2011). Golley et al. (2019) further state that 'this lower-bound result holds regardless of the (potentially complex) relationship between circumstances and effort' by giving a relevant example: compared to rural Chinese women, their counterparts (rural Chinese men) are more likely to migrate to cities and earn higher incomes; this tendency may be, in part, because rural parents invest more in their sons, leading to an upward bias on the ρ coefficient for gender via this indirect impact of effort (unobserved in the error term). They argue that 'while this leads us to stress again that the estimates cannot, therefore, be treated as causal, we maintain that a coefficient biased in this way would be entirely appropriate for a genuine understanding of equality of opportunity – because choices made by one's parents are beyond one's own control.'.

Second, another branch of literature has attempted to investigate the likely magnitude of the potential bias in the absence of an instrumental variable strategy (Bourguignon et al., 2007). While the method described above has become dominant in the literature, one disadvantage is that 'effort' variables are omitted – thereby putting aside the important questions about what individuals can actually do to offset – or complement – their circumstances. Furthermore, if the residual in Eq. (6) is regarded entirely as 'effort'—as has been done in many other studies—the proportion of inequality attributed to 'effort' will be larger than the 'true' value. The reason is obvious: the residual is not in fact 'true effort', but instead includes omitted circumstances, as well as random errors. To address this problem, Björklund et al. (2011) suggest measuring effort by a series of behaviours and then ascribing the remaining residual to 'luck'. Following them, and also the original approach by Bourguignon et al. (2007) (leaving aside their flawed decompositions), I attempt to account for some 'effort' variables.

As seen from Eqs. (3) and (4), Roemer's view is that an equal-opportunity policy must respect individual effort, but this needs to be purged of any contamination coming from circumstances. As an example, for the case that 'Asian children generally work hard in school and thereby do well because parents press them to do so', he proposes that the extra efforts of the Asian student must not be rewarded because it is determined by a characteristic which lies behind his control, i.e., the circumstance of being in the 'Asian' cultural context of familial pressures. This idea is laid out in Eq. (4), where E, the efforts of the individual, is a linear function of C, the circumstances.

Compared to Roemer's idea that has been widely discussed and adopted in the literature, there are two other views on the relationship between effort and circumstances: Barry (2005) argues that 'effort' is independent of circumstances while Swift (2005) demonstrates that circumstances only include past variables and have to be cleaned from any correlation with descendant's effort. This

[6] The parametrically standardized distribution using γ and the mean values of each circumstance would be specified as: $\{\widehat{y}_{i2}\}, where \widehat{y}_{i2} = [\overline{C}\widehat{\gamma} + \widehat{\varepsilon}]$

[7] Parametrically standardized estimates are obtained as: $IOA_I = I(\{y_{i2}\}) - I(\{\widehat{y}_{i2}\})$, $IOR_I = I(\{y_{i2}\}) - I(\{\widehat{y}_{i2}\})/I(\{y_{i2}\})$

Table 1
Earnings and Circumstance, by 10-year Cohorts.

Birth Cohort	Share	Earnings				
		All	1951–1960	1961–1970	1971–1980	1981–1990
	100	27,838	18,097	27,053	33,815	33,227
Gender						
Female	46	19,724	10,938	16,925	25,783	26,170
Male	54	34,797	23,250	36,010	41,676	39,176
Father's Education						
Illiterate	45	19,639	14,231	20,662	25,925	23,823
Primary School	35	32,396	21,741	35,195	36,127	30,872
Junior High School	13	37,591	25,690	29,613	45,146	38,120
Senior High School	7	40,496	80,062	40,311	33,063	38,168
Father's Occupation						
Low-Status Job	89	25,189	17,256	24,302	32,694	28,438
High-Status Job	18	39,755	25,571	44,218	38,746	42,050
Ethnic Minority Status						
Non-Han Chinese	13	18,697	10,623	17,098	25,366	21,166
Han Chinese	87	29,150	18,994	28,537	35,027	35,183
Birth of Region						
East	40	31,422	23,143	29,324	37,740	38,060
Centre	26	29,700	16,529	27,030	43,099	35,393
West	34	22,040	12,235	24,264	24,070	25,399

study mainly builds on Roemer (1998), but the other two ideas are also important. Jusot, Tubeuf, and Trannoy (2013) provide more details on the debate between Roemer, Barry, and Swift. Further studies that derive the effort function from solving an optimization function may help resolve this debate,[8] but it is beyond the scope of this study.

Utilizing these two equations, Bourguignon et al. (2007) attempted to examine two channels through which circumstances impact on efforts: a direct channel, reflected in the coefficients on each circumstance in Eq. (3); and an indirect channel, working through one's effort, as shown in Eq. (4). They also use Monte-Carlo methods to explore the potential bias by considering a wide range of estimates. However, their corrigendum showed that it is not reliable (Bourguignon et al., 2013). While these two equations cannot be used in the way initially intended, they are still useful for examining how efforts are constrained by circumstances, and for predicting the counterfactual incomes in different scenarios where individuals from different 'types' exert higher levels of effort. This helps to shed light on whether those from a worse (or less 'lucky') type can overcome their poor circumstances by exerting more efforts. I attempt to do this here, by returning to Eqs. (3) and (4) and adding some important effort variables in the context of rural China to the framework in Section 5.

In addition, I follow Brunori et al. (2018) in adopting machine learning methods—conditional inference regression trees and forests—to estimate inequality of opportunity, providing a standardized way of trading off upward and downward biases in inequality of opportunity estimations. This method coud address the outlined shortcomings of current approaches. The basic idea is to divide the observations into M groups (types) by what is known as recursive binary splitting according to the value they take in one circumstance variable C^p. Further division is applied to one of the two groups, and the final results can be presented as an upside-down tree.[9] As for comparative purposes, these results would also contribute to robustness checks.

While the overall share of IO is the main focus, I am still interested in the contribution of each of the circumstances – even though any estimation of this kind is plagued by the biases noted above (for this reason, these results are treated with caution and intended to be suggestive, rather than conclusive). I use a Shapley-value decomposition (Shorrocks, 1982, 2012) to decompose inequality of opportunity into its sources by eliminating the relative importance of each circumstance one by one. Using this decomposition, I first need to generate the 'power set' of the K circumstances and estimate the inequality of measure for all possible permutations of these circumstances. I take every element of the power set that does not include a specific circumstance, and compare inequality in that set with the set that is otherwise identical but does include the circumstance. In a second step, the average marginal effect of each circumstance variable on the measure of inequality of opportunity is computed. There are two benefits in contrast to other decomposition methods. First, the decomposition is order independent and second, the sum of all contributions is the value of overall inequality (Juarez & Soloaga, 2014). In this case, this means that decomposed shares of each of the circumstance variables add up to the share of inequality of opportunity (IOR). It is noteworthy that this approach is not perfect; I have not attributed causality to any of these circumstances, and I concede that biases—arising mainly from omitted variables—are likely to exist in all of them.

[8] We thank the reviewer for raising this point.
[9] Due to space limit, details of this method are not shown here. Refer to Brunori et al. (2018) for detailed procedures.

Table 2
Regressions using circumstance variables across birth cohorts.

		Birth cohort			
	All	1951–60	1961–70	1971–80	1981–90
Male	0.559***	0.604***	0.620***	0.569***	0.392***
	(0.02)	(0.05)	(0.04)	(0.04)	(0.04)
Primary school	0.225***	0.140**	0.293***	0.183***	0.236***
	(0.03)	(0.06)	(0.04)	(0.05)	(0.07)
Junior high school	0.287***	0.291**	0.420***	0.229***	0.266***
	(0.04)	(0.14)	(0.07)	(0.07)	(0.07)
Senior high school	0.343***	0.372**	0.258**	0.229**	0.425***
	(0.05)	(0.17)	(0.11)	(0.10)	(0.08)
Father-High status job	0.213***	0.179**	0.154***	0.284***	0.261***
	(0.03)	(0.09)	(0.06)	(0.06)	(0.05)
Han Chinese	0.265***	0.395***	0.270***	0.193***	0.203***
	(0.04)	(0.09)	(0.06)	(0.07)	(0.07)
East	0.438***	0.494***	0.411***	0.471***	0.376***
	(0.03)	(0.06)	(0.05)	(0.05)	(0.06)
Centre	0.282***	0.205***	0.252***	0.346***	0.364***
	(0.03)	(0.07)	(0.05)	(0.06)	(0.06)
Birth cohort 61–70	0.479***				
	(0.03)				
Birth cohort 71–80	0.736***				
	(0.03)				
Birth cohort 81–90	0.758***				
	(0.04)				
Constant	8.188***	8.065***	8.626***	8.977***	9.069***
	(0.04)	(0.09)	(0.06)	(0.07)	(0.08)
N	9096	1993	3147	2207	1749
Adjusted R-square	0.203	0.123	0.148	0.141	0.148
F	211.588	35.812	69.175	46.165	38.997

Notes: * ** 1%,* * 5%,* 10%; Standard errors in parentheses.

4. Data and results

4.1. Data

The data is from the China Labor-force Dynamics Survey (CLDS) produced by the Center for Social Science Surveys of Sun Yat-sen University in China. CLDS is a nationally representative survey of Chinese communities, families, and individuals in contemporary China. The samples cover 29 provinces (excluding Hong Kong, Macau and Taiwan, Tibet and Hainan), with a focus on labour in the household. Using multistage cluster, stratified PPS sampling, CLDS did the trial survey in Guangdong in 2011, conducted the first formal investigation in 2012, and completed the first tracking survey in 2014. The 2014 data has been available to the public since January 2017.

I restrict the sample to individuals born in 1951–1990 (aged 24–63) in rural areas. We choose this sample in order to focus on rural individuals with the highest levels of labour market attachment. The complete CLDS 2014 sample size is 23,594 individuals. Excluding individuals with urban *hukou*, and those outside the 1951–1990 cohort range yields a sample of 15,035 individuals. The final sample is then divided into four 10-year birth cohorts: from individuals born between 1951 and 1960 through to those born between 1981 and 1990. This allows us to shed light not only on the role of circumstances in determining the observed earnings at one point in 2014, but also how this may vary across cohorts (although notably, not across time).

Turning to the circumstances, the survey contains abundant information. The circumstances I choose based on the above surveyed literature include (i) gender, which is a dummy with male—accounting for 54% of the respondents—taking value 1; (ii) ethnic minority status, which is a dummy as well, with Han taking value 1 and accounting for 87% in the full sample; (iii) father's education, which is re-coded into four categories—illiterate, primary school, junior high school and senior high school and above; (iv) father's occupation, which is re-coded into two categories. 'Lower status' includes farmers, fishermen, agricultural labourers, farm and forestry workers, hunters and related workers and 'higher status' mainly includes off-farm workers, such as administrative officials, corporation leaders, professionals, clerical jobs and transport and communication supervisors; (v) region of birth, which is categorized into three regions that have been widely adopted in China: East, Central and West. The East region includes Beijing, Tianjin, Hebei, Liaoning, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, Hainan. The Central region comprises the provinces of Shanxi, Jilin, Heilongjiang, Anhui. Jiangxi. Henan, Hubei, Hunan. The West region includes Inner Mongolia, Guangxi, Chongqing, Sichuan, Guizhou, Yunan, Xizang, Shan'xi, Gansu, Qinghai, Ningxia, Xinjiang. The West region is taken as the reference category.

Table 3
Inequality of opportunity in earnings in rural China.

	All	Birth Cohorts			
		1951–60	1961–70	1971–80	1981–90
Panel A					
Total inequality					
Gini	0.560	0.610	0.575	0.541	0.459
GE(0)	0.646	0.757	0.670	0.595	0.422
Inequality of Opportunity					
Absolute (IOA)	0.132	0.086	0.092	0.086	0.068
Relative Share (IOR)	**0.204**	**0.114**	**0.138**	**0.144**	**0.161**
Panel B					
Shapley decomposition (% of IOR; % of overall earnings inequality)					
Gender	27.25	48.06	51.09	46.03	28.92
	5.6	**5.4**	**7.1**	**6.6**	**4.7**
Father's education	15.23	7.69	16.37	8.07	15.68
	3.1	**0.9**	**2.2**	**1.2**	**2.5**
Father's occupation	8.33	4.09	4.57	11.17	19.17
	1.7	**0.5**	**0.6**	**1.6**	**3.1**
Ethnic minority status	5.38	11.14	8.42	7.01	10.15
	1.1	**1.3**	**1.2**	**1.0**	**1.6**
Birth region	14.27	29.02	19.55	27.72	26.08
	2.9	**3.3**	**2.7**	**4.0**	**4.2**
Birth cohort	29.54				
	6.0				

The dependent variable is individual annual earnings in 2014. Note that some individuals may work part-time, so the hours or days they work in one year may differ from full-time employees. The analysis could have been enriched if the survey included data for hourly or daily earnings – but it does not. However, it is still meaningful to understand what constrains or advances the annual earnings of individuals – for example, if women work fewer hours because of household duties, and their annual incomes are constrained accordingly, this is indicative of a lack of equal opportunity for annual – and hence lifetime – earnings. The preliminary statistics in Table 1 show that the average annual earnings for rural individuals is 27,838 yuan. This varies across birth cohorts, from 18,097 yuan for the 1951–1960 birth cohort to 33,815 yuan for 1971–1980 birth cohort. Not surprisingly, individuals born in the East, on average, earn a higher income than their counterparts born in the Centre and West, and men earn close to double that of women. Individual earnings also vary across people with other circumstances: for example, a male Han from the East, in age cohort 1971–80, with a more educated father in a high-status job, on average, earns more than any other 'types'.

4.2. Determinants of individual earnings

This section briefly reports on the estimation results. Table 2 presents the regression results of the earnings Eq. (6) estimated by OLS, for the whole sample and separately for each cohort.

All circumstance variables have the expected signs on individual earnings for the full sample (in Column 1). The gender and ethnic biases are clear: with the females and non-Han as the respective references, being male is associated with earnings that are 0.559 log points[10] higher and being Han is associated with earnings that are 0.265 log points higher. It is also clear that having a more educated father working in a higher-status job is beneficial to an individual's job performance, with primary school, junior high and senior high and above being associated with earnings that are 0.225, 0.287 and 0.343 log points higher respectively compared to their counterparts with no schooling, and with high-status job of the father being associated with a 0.213 log point income boost compared to those with low-status jobs. Geographical variables are highly significant and negative, indicating that being in the Central and Eastern regions is associated with earnings that are 0.282 and 0.438 log points higher than being in the west. Turning to birth cohorts, the results are consistent with the preliminary statistics that the two younger cohorts are associated with higher earnings.

The regressions for each birth cohort identify important variations in the magnitude and significance of key determinants, as seen in columns 2–5 in Table 2. Gender is a highly significant determinant of individual earnings for all cohorts. The coefficient of gender peaks at 0.620 log points for the cohort 1961–1970, and reaches a low of 0.392 log points the cohort 1981–1990. Other key points include the consistently positive coefficients on father's occupation and education, the significant advantage of being born in the centre and east compared to being born in the west, and of being Han in contrast to being an ethnic minority.

[10] For the log-linear form with dummy regressor, the exact percentage change in the dependent variable (i.e., semielasticity), is $(\exp(\beta) - 1) * 100$, where β is the coefficient of the dummy regressor. The case here amounts to 74.9%, given by $(\exp(0.559) - 1) * 100$.

Table 4
Earnings and Efforts, by 10-year Cohorts.

Birth Cohort	Share	Earnings				
		All	1951–1960	1961–1970	1971–1980	1981–1990
Education						
Below Junior High School	42	17,554	14,987	17,560	21,310	17,944
Junior High and Above	58	35,328	23,352	35,388	41,461	35,806
Off-farm employment						
No	51	15,591	11,999	16,334	17,966	20,125
Yes	49	40,626	35,698	41,292	44,965	37,657
Migration						
No	77	21,914	13,735	23,965	25,551	28,577
Yes	23	26,717	17,987	27,118	30,638	26,726
Marriage						
No	7	29,425	17,452	22,439	25,490	36,507
Yes	93	27,711	18,136	27,229	34,138	32,377

Table 5
The impacts of circumstances on efforts.

	Effort Variables			
	Education	Migration	Off-farm	Marriage
Gender	0.598***	0.455***	0.335***	-0.291***
	(0.03)	(0.04)	(0.03)	(0.04)
Primary school	0.391***	0.205***	0.287***	-0.091 *
	(0.03)	(0.04)	(0.03)	(0.05)
Junior high school	0.773***	0.146**	0.341***	-0.130 *
	(0.05)	(0.06)	(0.05)	(0.07)
Senior high school	0.659***	0.100	0.398***	-0.154 *
	(0.07)	(0.08)	(0.07)	(0.09)
Father-High status job	0.406***	0.176***	0.650***	-0.100 *
	(0.04)	(0.05)	(0.04)	(0.05)
Han Chinese	0.450***	-0.092 *	0.466***	0.005
	(0.05)	(0.05)	(0.05)	(0.07)
East	0.235***	-0.166***	0.602***	0.032
	(0.04)	(0.04)	(0.04)	(0.05)
Centre	0.126***	0.044	0.277***	0.038
	(0.04)	(0.05)	(0.04)	(0.06)
Birth cohort 61–70	0.412***	0.287***	0.504***	0.206***
	(0.04)	(0.05)	(0.04)	(0.06)
Birth cohort 71–80	0.549***	0.535***	0.899***	0.190***
	(0.04)	(0.05)	(0.04)	(0.07)
Birth cohort 81–90	1.158***	0.805***	1.211***	-0.727***
	(0.05)	(0.06)	(0.05)	(0.06)
Constant	-1.408***	-1.368***	-1.812***	1.822***
	(0.06)	(0.06)	(0.06)	(0.08)
N	9096	6966	9096	9096
Pseudo R2	0.179	0.073	0.187	0.112

Notes: * ** 1%,* * 5%,* 10%; Standard errors in parentheses.

4.3. Inequality of opportunity

Using the coefficient estimates from the reduced-form Eq. (6) reported in Table 2, I simulate the counterfactual distributions, following the method presented in Section 3. This enables the decomposition of earnings inequality for the whole sample and separately for each cohort into a component due to unequal circumstances over which individuals have no control (IOA) and a residual component.

Table 3 shows the IOA and IOR in total observed earnings inequality in rural China. The first two rows report the income inequality index, GE(0), and the Gini coefficient for comparative purposes. Total inequality in earnings is high in rural China, with the GE(0) and Gini coefficient being 0.560 and 0.646 respectively. Total income inequality for the younger cohorts is lower than that for the older cohorts.

Turning to the inequality due to unequal opportunities, the overall share for the full sample is 20.4%. The lower IORs for each birth

Table 6
The impacts of circumstances and efforts on earnings.

	Earnings
Gender	0.435***
	(0.02)
Primary school	0.111***
	(0.02)
Junior high school	0.135***
	(0.04)
Senior high school	0.187***
	(0.05)
Father-High status job	0.042
	(0.03)
Han Chinese	0.116***
	(0.03)
East	0.269***
	(0.03)
Centre	0.201***
	(0.03)
Birth cohort 61–70	0.311***
	(0.03)
Birth cohort 71–80	0.453***
	(0.03)
Birth cohort 81–90	0.404***
	(0.04)
Education	0.270***
	(0.02)
Off-farm employment	0.740***
	(0.02)
Marriage	0.220***
	(0.04)
Constant	8.038***
	(0.05)
N	9096
Adj. R-squared	0.306

Notes: * ** 1%,* * 5%,* 10%; Standard errors in parentheses.

cohort is due to the exclusion of birth cohort as a circumstance in estimating IORs across cohorts. The key point, however, is that IORs tend to be higher for the younger cohorts: with an IOR value of 16.1% for the 1981–1990 cohort (those born after the economic reform) compared with one of 11.4% for the 1951–1960 cohort (the one impacted directly by the Cultural Revolution).

To assess the contribution of each of the circumstances, I use the Shapley-value decomposition discussed in Section 3 to decompose the overall inequality of opportunity into circumstance-specific parts. Panel B in Table 3 presents two types of shares for each of the circumstances—the percentage contribution of a specific circumstance to the IOR (or equivalently the IOA) in the first row and to overall income inequality in the second row. The results suggest that age (birth cohorts)—which accounts for 6.0% of overall income inequality and 29.5% of the IOR—is the number one contributor for the full sample, followed closely by gender, which accounts for 5.6% of outcome inequality. The lack of equal opportunity for individuals with regard to their earnings also stems from father's education (3.1%), birth region (2.9%), father's occupation (1.7%) and ethnicity (1.1%).

Looking at each birth cohort, gender is the largest single contributor to the total income inequality, but the share is lower for the younger cohorts: 4.7% for the cohort 1981–1990 compared to 7.1% for the cohort 1961–1970. Region of birth consistently ranks in the second place. Father's education and occupation contribute more to income inequality for the younger cohorts, ranging from 0.9% to 2.5%, and from 0.5% to 3.1%, for the cohort 1951–60 and 1981–90 cohorts respectively. This implies that the combined contribution of these two circumstances – i.e., 'family background', or 'daddy's status' – is also larger for the younger cohorts, peaking at 5.6% of overall income inequality for the youngest, thus exceeding the contribution of gender (4.7%) for the youngest cohort. In other words, other than one's gender, it certainly seems that 'daddy is the key' for determining earnings among the younger cohorts in rural China. Ethnicity appears to be the smallest contributor, accounting for less than 2% of overall income inequality across the cohorts. While treating these partial findings with caution given the caveats noted above, these are interesting results nonetheless.

To explore the likely magnitude of the potential biases stemming from unobserved circumstances, we compare the overall inequality of opportunity from the parametric method above against trees and forests shown in Brunori et al. (2018). Table A1 in the appendix shows that the relative measure of overall inequality of opportunity using the parametric method (0.204) is larger than that using trees (0.164), but it is smaller than that using forests (0.254). Keeping the potential bias in mind, we find that the parametric method—as the most straightforward approach—presents seemingly average estimates of inequality of opportunity.

Table 7
Predicted income for the best type and worst type.

	(1) Worst Type	(2) Worst Type + Lowest Effort	(3) Worst Type+ Highest Effort	(4) Worst Type+ meanEffort
		Panel A		
Gender	Female	Female	Female	Female
Father's education	Illiterate	Illiterate	Illiterate	Illiterate
Region of birth	West	West	West	West
Father's occupation	Low status	Low status	Low status	Low status
Ethnicity	Non-Han	Non-Han	Non-Han	Non-Han
Own education		Below Junior High	Senior high and above	0.11
Off-farm employment		No	Yes	0.10
Marital status		Single	Married	0.97
FY	3598	3096	10,594	4259
No.observations	249	249	249	249
		Panel B		
	Best Type	Best Type + Lowest Effort	Best Type+ Highest Effort	Best Type+ mean Effort
Gender	Male	Male	Male	Male
Father's education	Senior high and above	Senior high and above	Senior high and above	Senior high and above
Region of birth	East	East	East	East
Father's occupation	High status	High status	High status	High status
Ethnicity	Han	Han	Han	Han
Own education		Below Junior High	Senior high and above	0.92
Off-farm employment		No	Yes	0.92
Marital status		Single	Married	0.77
FY	47,287	13,905	47,572	41,902
No.observations	79	79	79	79

5. The role of efforts in rural China

I now return to the method discussed in Section 3 to look more closely into the relationship between circumstances and effort. I first identify four effort variables for further study. The first one is off-farm employment, which indicates whether farmers are off-farm employed or not – an outcome that may be considered as requiring some degree of effort. The share of respondents who report themselves involved in non-agricultural activities is 49%.[11] The second one is migration. Note that the ratio of migrants is only 23%, indicating that in contrast to migration, some rural residents tend to work in the non-agricultural sector in the nearby village or towns. Local employment offers an important avenue for farmers to be off-farm employed.

The third effort variable I choose is one's own level of education,[12] which is well known as a key factor contributing to individual earnings. Given the rural context where schooling above junior high has never been compulsory, I treat it as a dummy that indicates whether an individual has completed junior high school and above. Table 4 shows that more educated individuals who choose to migrate or be off-farm employed have higher earnings. The last 'effort' variable included is marital status. The reason for including this variable is that couples, in contrast to single men or women, have interdependent preferences that affect their household income decisions, in ways that imply different employment choices for men and women (Zhang et al., 2008). Table 4 presents that married individuals are likely to earn more than those who are not married, with the only exception of the youngest cohort.[13]

In a first step, I show the impacts of circumstances on the four chosen effort variables using a Probit model shown in Eq. (4) (Table 5). Nearly all the coefficients take on their expected signs and relative magnitudes. For example, individuals with more educated fathers are more likely to be more educated themselves, and their decision of employment is determined by all the circumstances included: off-farm employment is more likely to be the choice of male and Han individuals with better father's background. People born in the east are more likely to be off-farm employed as well, but less likely to migrate (Chan, 2013). Men are less likely to be married than women in rural China, which reflects the highly male-biased adult sex ratios stemming from male-biased sex ratios at birth, combined with out-migration of women. The Pseudo R-square for the regressions of own-education and off-farm employment take on the highest values—0.18 and 0.19 respectively—while the migration regression takes on the lowest value (0.08), suggesting that the indirect effect of circumstances on earnings through migration is likely to be the smallest. Combined with the fact that there are too many missing values for migrants – resulting in the loss of more than 2000 observations – migration is not

[11] Note that the ratio of migrants is only 23%, indicating that in contrast to migration, some rural residents tend to work in the non-agricultural sector in the nearby village or towns. Local employment offers an important avenue for farmers to be off-farm employed.

[12] Whether or not it deserves to be classified as an 'effort' variable is a matter of extensive ongoing debate between Roemer and Barry, but I follow Bourguignon et al. (2007) in treating it as an 'effort' here. The reason is that although one's own education is partly explained by their circumstances (Wang et al., 2011), it is also a matter of choice.

[13] This is mainly due to the fact that the majority of those from the youngest cohort are unmarried. It could be argued that there is a reverse causality problem here, in that people may be married because they have high wages, and not the reverse. But for the older cohorts this is unlikely to be the case in a 2014 cross section of data: they were almost certainly married before their earnings were observed in 2014.

included in the following analysis.[14]

Table 6 presents the regressions of individual earnings on both circumstances and effort variables shown in Eq. (3). All the coefficients have expected values and signs: married individuals with better education and off-farm jobs are more likely to earn more. Specifically, working in a non-agricultural occupation is associated with earnings that are 0.740 log points higher than working in agriculture, while having junior high school education or above is associated with a 0.270 log point income boost. The results also show that inclusion of these three effort variables results in a considerable increase in adjusted R-squared values from 0.20 in Table 2 to 0.31 in Table 6.

The results above, together with those in Table 2 (Eq. 6), can be used to produce some simple predictions to examine whether individuals can overcome their poor circumstances by exerting more effort. I first identity the 'worst (or unluckiest) type' (ie. with the lowest average earnings) as women born in Western China with illiterate fathers who work in low-status industries, and the 'best (or most fortunate) type' as men born in Eastern China with fathers who have senior high school education and above and work in high-status industries. Based on the results in Table 2 where only circumstances are included, column 1 of Panel A and B Table 7 shows that predicted annual earnings for the 'worst type' (3598 yuan) is around ten times lower than the 'best type' (47,287 yuan).

Based on Eq. (3), columns 2 and 3 further show the predicted earnings from the regressions in which the three effort variables are also included. Column 2 shows the counterfactual earnings in the scenario where all individuals are assumed to exert the lowest levels of effort (that is, single, having below junior high school education and working in agriculture), while column 3 shows the counterfactual earnings in the scenario where individuals have the highest levels of effort (married, having junior high school education and above and working in non-agriculture)[15]. As shown in columns 2 and 3 of Panel A, the forecasted earnings for the worst type with lowest and highest level of effort are 3096 yuan and 10,594 yuan respectively, suggesting that exerting the lowest effort decreases the income slightly while having the highest effort increases the income dramatically compared to Column 1. In contrast, columns 2 and 3 of Panel B shows that, for the best type, exerting the lowest effort results in a considerable fall in income (13,905 yuan) while having highest effort increases the income slightly to 47,572 yuan compared to 47,287 yuan shown in column 1.

What this suggests is that, if individuals exert more effort, then the income of the 'worst type' would increase significantly, while the income of the 'best type' does not rise that much. But does this mean that inequality would fall if only the 'worst types' could exert more effort? The answer is not that simple. In reality, at least part of the reason why people from the 'worst type' have lower effort is because their effort is constrained by circumstances. To have a more straightforward understanding of this point, I calculate the mean value of the level of efforts for the 'worst type' and 'best type' respectively. Column 4 shows that 92% of the people from the 'best type' are better educated while only 11% of those from the 'worst type' have junior high school education and above. With regard to employment, the share of people working in non-agriculture for the 'best type' is 0.92%, considerably higher than that for the 'worst type' (10%). Coupled with the regression results in Table 5 (that show how efforts are determined by circumstances), this indicates that those who find themselves in the worst circumstances are likely to exert lower level of effort, not because they don't want to try harder, but because their circumstances prevent them from doing so.

6. Conclusions

This paper set out to quantify the role of inequality of opportunity – associated with the circumstances of family background, gender, ethnic minority status, region of origin and birth cohort – in generating individual earnings inequality in rural China. Using the China Labour-force Dynamics Survey (CLDS) for 2014, this is the first attempt in a Chinese rural context, to the best of my knowledge, to examine to what extent inequality of opportunity contributes to total income inequality.

The empirical results revealed that the share of inequality of opportunity in individual earnings for the full sample is 20.4% of the GE(0) coefficient. The adoption of machine learning methods provides a wide range of estimates between 16.4% (regression tress) and 25.4% (forests). The youngest cohort 1981–1990 faces the lowest total inequality in earnings, but it turns out to be the one with highest IOR value. The IOR in this study (20.4%) is smaller in contrast to those shown in Ferreira and Gignoux (2011) – 32.2% in Brazil, 33.5% in Guatemala, 30.1% in Panama, 27.9% in Perun, 25.9% in Colombia—and some others, such as around 27% in Egypt (Assaad et al., 2017), but larger than the case of some high welfare states, such as Sweden shown in Björklund et al. (2011). However, it is extremely close to the case of India where the overall inequality of opportunity share in total observed inequality is 20.8% for the rural sample (Singh, 2012). Moreover, the fact that only a rural sample was used would significantly reduce the measured IOR compared to a nationwide China study – because *hukou* status as a circumstance is a significant contributor to the rural-urban income gap and this would yield a substantially higher IOR (Li & Gibson, 2013; Sicular et al., 2007; Yang, 1999).

The Shapley-value decomposition showed that age (birth cohort) is the number one contributor for the full sample (6.0%), followed closely by gender, accounting for 5.6% of income inequality. This result varies across birth cohorts: (i) Gender is the largest single contributor to overall income inequality for all the cohorts, but becomes less important for the younger cohorts; (ii) Region of birth consistently ranks in the second place; (iii) Father's education and occupation is most important for the youngest cohorts, and if taken together as 'family background', they become the largest contributor for the youngest cohort; (iv) ethnicity consistently has the lowest

[14] This results in the adjusted R-squared value that is much lower when migration is included in the earnings equation (see Appendix). Table A2 also shows that migration turns out to have an insignificant impact on individual earnings.

[15] This does not indicate that being married is better than being single. It is, rather, just a reflection on earnings gap shown in our sample.

ranking. A further look into the role of effort shows that exerting more effort would increase income more dramatically for the 'worst type' compared to the 'best type'. But this does not imply that the 'worst type' simply need to work harder. The reality is that people from the 'best type' already have the highest level of effort; at least partially because their circumstances allow it.

In combination, the results presented in this paper suggest that inequality of opportunity is a serious problem in rural China. So what kind of equal opportunity policies could be used to mitigate this problem? Given that 'daddy seemed to be the key'for the youngest cohort and that one's socio-economic status clearly affects one's earning potential, both directly and indirectly through effort, policies that level out the educational and employment opportunities focusing especially on rural women from the lowest socio-economic family backgrounds would almost certainly reduce income inequality in the future. Although China has a strong meritocratic history, in recent years it appears that job opportunities are based more on who you know than what you know. Labour laws ensuring that women and rural *hukou* holders are not discriminated against in the workforce –both in terms of the job application process and promotions thereafter –would be steps in the right direction.

Measures to assist early career women, in particular, to shift them out of agriculture to find off-farm employment, would also improve their earnings potential in the future. This would need to happen in tandem with improved children and elderly care facilities in rural areas, given that the burden of caring for children and the elderly falls disproportionately on women. While investment in formal education is a long-term task, increased government funding for professional education (training)[16] for rural workers is another example of an equal opportunity policy that would help to level the playing field. Despite the substantial efforts made by the Chinese government to promote rural socio-economic development, rural people are still lagging behind their urban counterparts—and within rural China there are those who lag further behind still. Equality of opportunity is an ideal worthy of pursuit by any government, and is the key to the door of an equal and harmonious China in the future.

While the results presented in this paper has some policy implications for the Chinese government, there are several issues that have not be addressed beyond the scope of the thesis and could remain the task of future research. First, the framework of inequality of opportunity could also be applied to other outcomes, for instance, education and health, which have been done so in the international literature but barely shown in the context of (rural) China. The severity of inequality of opportunity in China could be further verified if circumstances also account for a large share of the total inequality in these outcome variables, which calls for more attentions on equal opportunity policies. Second, since gender and family background are the key contributors to total income inequality, it is also interesting to examine how inequality of opportunity differs between sub-groups, for instance, between male and female and between those from worse family background and those from better family background. Third, as have been emphasized, the partial contributions of each of the circumstances using decomposition methods are likely to be biased. More explorations on the method to address this issue is of great significance for future research.

Acknowledgments

This work is a revised version of the fifth chapter of my Ph.D. thesis. I am deeply grateful to Professor Jane Golley (my primary supervisor) for her valuable time, advice and corrections to this paper, and also for the insightful comments made by Professor Bruce Chapman. I would also like to thank Dr. Ligang Song and other participants in seminars given at Crawford School of Public Policy of The Australian National University. My thanks also go to all the participants in other seminars wherever the previous version of this work was presented. This work is supported by the by the National Natural Science Foundation of China (No. 72003170), the National Social Science Fund of China (No. 21&ZD091), the Fundamental Research Funds for the Central Universities in China, the think tank Construction Project in Zhejiang Province (No.2021-SKY-ZX-0227) and the project on the future rural high-quality development in Jiashan County: promoting the equalization of urban and rural public services through digital village construction (No.2021-SKY-523102-0030). I also appreciate the financial support from the Australian Centre on China in the World PhD Scholarship when I did my PhD at the Australian National University.

Appendix

See Tables A1 and A2 and Fig. A1.

Table A1
Comparison of overall inequality of opportunity.

	Parametric method	Trees	Forests
Absolute iop	0.132	0.106	0.164
Relative iop	0.204	0.164	0.254

[16] Yi et al. (2013) show the great significance of secondary vocational education in China.

Table A2
The impacts of circumstances and efforts on earnings (migration included).

	Earnings
Gender	0.443***
	(0.03)
Primary school	0.088***
	(0.03)
Junior high school	0.089**
	(0.04)
Senior high school	0.139**
	(0.06)
Father-High status job	-0.013
	(0.04)
Farther- Han Chinese	0.115***
	(0.04)
East	0.332***
	(0.03)
Centre	0.225***
	(0.03)
Birth cohort 61–70	0.341***
	(0.03)
Birth cohort 71–80	0.486***
	(0.04)
Birth cohort 81–90	0.442***
	(0.04)
Education	0.233***
	(0.03)
Off-farm employment	0.652***
	(0.03)
Marriage	0.319***
	(0.05)
Migration	0.034
	(0.03)
Constant	7.899***
	(0.06)
N	6966
Adj. R-squared	0.265
F	168.489

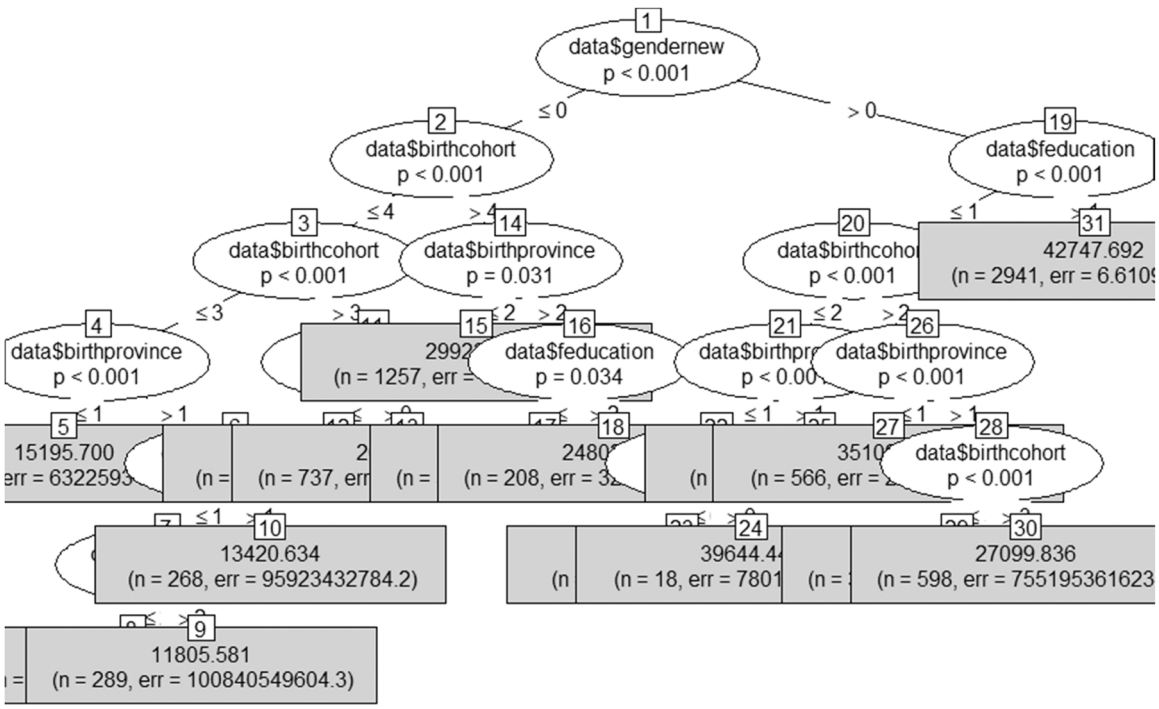

Fig. A1. An example of regression trees.

References

Arneson, R. J. (1989). Equality and equal-opportunity for welfare. *Philosophical Studies, 56*(1), 77–93.
Assaad, R., Krafft, C., Roemer, J., & Salehi-Isfahani, D. (2017). Inequality of opportunity in wages and consumption in egypt. *Review of Income and Wealth*.
Barry, B. X. (2005). *Why social justice matters*: Polity.
Benjamin, D., Brandt, L., & Giles, J. (2005). The evolution of income inequality in rural China. *Economic Development and Cultural Change, 53*(4), 769–824.
Björklund, A., Jäntti, M., & Roemer, J. E. (2011). Equality of opportunity and the distribution of long-run income in Sweden. *Social Choice and Welfare, 39*(2–3), 675–696.
Bonnefond, C., & Clément, M. (2012). An analysis of income polarisation in rural and urban China. *Post-Communist Economies, 24*(1), 15–37.
Bourguignon, F., Ferreira, F. H., & Menendez, M. (2007). Inequality of opportunity in Brazil. *Review of Income and Wealth, 53*(4), 585–618.
Bourguignon, F., Ferreira, F. H. G., & Menéndez, M. (2013). Inequality of Opportunity in Brazil: A Corrigendum. *Review of Income and Wealth, 59*(3), 551–555.
Bowles, S. (1972). Schooling and inequality from generation to generation. *Journal of Political Economy, 80*(3), S219–S251.
Brunori, P., Hufe, P., & Mahler, D. G. (2018). The roots of inequality: Estimating inequality of opportunity from regression trees. *World Bank Policy Research Working Paper*, (8349).
Chan, K. W. (2013). China: internal migration.
Checchi, D., & Peragine, V. (2010). Inequality of opportunity in Italy. *The Journal of Economic Inequality, 8*(4), 429–450.
Chen, Z., Ge, Y., Lai, H., & Wan, C. (2013). Globalization and gender wage inequality in China. *World Development, 44*, 256–266.
Clément, M. (2016). Income mobility and income inequality in rural China. *Frontiers of Economics in China, 11*(4), 1673–3444.
Cohen, G. A. (1989). On the currency of egalitarian justice. *Ethics, 99*(4), 906–944.
Dai, X., & Li, J. (2021). Inequality of opportunity in China: Evidence from pseudo panel data. *China Economic Review, 68*, Article 101637.
Dworkin, R. (1981). What is equality? Part 1: Equality of welfare. *Philosophy & Public affairs*, 185–246.
Fan, S., Kanbur, R., & Zhang, X. (2009). *Regional inequality in China: Trends, explanations and policy responses*. Routledge.
Ferreira, F. H. G., & Gignoux, J. (2011). The Measurement of Inequality of Opportunity: Theory and an Application to Latin America. *Review of Income and Wealth, 57*(4), 622–657.
Fleisher, B., Li, H., & Zhao, M. Q. (2010). Human capital, economic growth, and regional inequality in China. *Journal of Development Economics, 92*(2), 215–231.
Gao, J., Liu, Y., Chen, J., & Cai, Y. (2019). Demystifying the geography of income inequality in rural China: A transitional framework. *Journal of Rural Studies %@. 0743-0167*.
Golley, J., & Kong, S. T. (2018). Inequality of opportunity in China's educational outcomes. *China Economic Review, 51*, 116–128.
Golley, J., Zhou, Y., & Wang, M. (2019). Inequality of opportunity in China's labor earnings: The gender dimension. *China & World Economy, 27*(1), 28–50.
Gustafsson, B., & Li, S. (2002). Income inequality within and across counties in rural China 1988 and 1995. *Journal of Development Economics, 69*(1), 179–204.
Gustafsson, B., & Shi, L. (2003). The ethnic minority-majority income gap in rural China during transition. *Economic Development and Cultural Change, 51*(4), 805–822.
Ha, W., Yi, J., Yuan, Y., & Zhang, J. (2016). The dynamic effect of rural-to-urban migration on inequality in source villages: System GMM estimates from rural China. *China Economic Review, 37*(27–39), 1043–1951X (%@).
Hannum, E. (2005). Market transition: Educational disparities, and family strategies in Rural China: New evidence on gender stratification and development. *Demography, 42*(2), 275–299.
Hanoch, G. (1967). An economic analysis of earnings and schooling. *Journal of Human Resources*, 310–329.
Howell, A. (2017). Impacts of migration and remittances on ethnic income inequality in rural China. *World Development, 94(Migration on Income)*, 200–211.
Juarez, F. W. C., & Soloaga, I. (2014). iop: Estimating ex-ante inequality of opportunity. *Stata Journal, 14*(4), 830–846.
Jusot, F., Tubeuf, S., & Trannoy, A. (2013). Circumstances and efforts: how important is their correlation for the measurement of inequality of opportunity in health? *Health Economics, 22*(12), 1470–1495.
Kanbur, R., & Zhang, X. B. (2005). Fifty years of regional inequality in China: a journey through central planning, reform, and openness. *Review of Development Economics, 9*(1), 87–106.
Knight, J. (2017). Fair and unfair inequality in China. *Economic and Political Studies, 5*(3), 256–265.
Kung, J. K., & Lee, Y.-F. (2001). So what if there is income inequality? The distributive consequence of nonfarm employment in rural China. *Economic Development and Cultural Change, 50*(1), 19–46.
Lefranc, A., Pistolesi, N., & Trannoy, A. (2008). Inequality of opportunities vs. inequality of outcomes: Are Western societies all alike? *Review of Income and Wealth, 54*(4), 513–546.
Li, C., & Gibson, J. (2013). Rising regional inequality in China: Fact or artifact? *World Development, 47*, 16–29.
Liu, X., Shi, X., & Chen, K. (2022). Inequality of opportunity in children's nutritional outcomes in China. *Global Food Security, 33*, Article 100635.
Liu, X., & Sicular, T. (2009). Nonagricultural employment determinants and income inequality decomposition. *Chinese Economy, 42*(4), 29–43. %@ 1097-1475.
Matthews, R., & Nee, V. (2000). Gender inequality and economic growth in Rural China. *Social Science Research, 29*(4), 606–632.
McKinley, T. (2016). *The distribution of wealth in rural China*. Routledge,.
Morduch, J., & Sicular, T. (2000). Politics, growth, and inequality in rural China: Does it pay to join the Party? *Journal of Public Economics, 77*(3), 331–356.
Naughton, B. (2007). *The Chinese economy: Transitions and growth*. MIT press,.
Peragine, V., & Ferreira, F. (2015). Equality of opportunity: Theory and evidence. *World Bank Policy Research Paper*, 7217).
Qin, X., Wang, T., & Zhuang, C. C. (2016). Intergenerational transfer of human capital and its impact on income mobility: Evidence from China. *China Economic Review, 38*(306–321), 1043–1951X (%@).
Roemer, J. E. (1993). A pragmatic theory of responsibility for the egalitarian planner. *Philosophy & Public Affairs*, 146–166.
Roemer, J. E. (1998). *Equality of opportunity*. Cambridge, MA: Harvard University Press.
Scharf, M. M., & Rahut, D. B. (2014). Nonfarm employment and rural welfare: Evidence from the Himalayas. *American Journal of Agricultural Economics, 96*(4), 1183–1197.
Shi, X. (2019). Inequality of opportunity in energy consumption in China. *Energy policy, 124*, 371–382.
Shi, X. (2022). Moving out but not for the better: Health consequences of interprovincial rural-urban migration in China. *Health Economics, 31*(4), 555–573.
Shi, X. H., Liu, X. Y., Nuetah, A., & Xin, X. (2010). Determinants of household income mobility in rural China. *China & World Economy, 18*(2), 41–59.
Shorrocks, A. F. (1982). Inequality decomposition by factor components. *Econometrica: Journal of the Econometric Society*, 193–211.
Shorrocks, A. F. (2012). Decomposition procedures for distributional analysis: A unified framework based on the Shapley value. *The Journal of Economic Inequality, 11*(1), 99–126.
Sicular, T., Ximing, Y., Gustafsson, B., & Shi, L. (2007). The urban–rural income gap and inequality in China. *Review of Income and Wealth, 53*(1), 93–126.
Singh, A. (2012). Inequality of Opportunity In Earnings And Consumption Expenditure: The Case of Indian men. *Review of Income and Wealth, 58*(1), 79–106.
Song, J., Cavusgil, E., Li, J., & Luo, R. (2016). Social stratification and mobility among Chinese middle class households: An empirical investigation. *International Business Review, 25*(3), 646–656. %@ 0969-5931.
Song, Y., & Zhou, G. (2019). Inequality of opportunity and household education expenditures: Evidence from panel data in China. *China Economic Review, 55*, 85–98.
Swift, A. (2005). Justice, luck and the family: Normative aspects of the intergenerational transmission of economic status. *Unequal Chances: Family Background and Economic Success* (pp. 256–276). Princeton: Princeton University Press.
Walder, A. G. (2002). Markets and income inequality in rural China: Political advantage in an expanding economy. *American Sociological Review*, 231–253.
Wan, G. H., & Zhou, Z. Y. (2005). Income inequality in rural China: Regression-based decomposition using household data. *Review of Development Economics, 9*(1), 107–120.
Wang, X., Liu, C., Zhang, L., Luo, R., Glauben, T., Shi, Y., & Sharbono, B. (2011). What is keeping the poor out of college? Enrollment rates, educational barriers and college matriculation in China. *China Agricultural Economic Review, 3*(2), 131–149. %@ 1756-1137X.

Weiss, R. D. (1970). The effect of education on the earnings of blacks and whites. *The Review of Economics and Statistics*, 150–159.
Xie, Y., & Zhou, X. (2014). Income inequality in today's China. *Proceedings of the National Academy of Sciences United States A, 111*(19), 6928–6933.
Yang, D. T. (1999). Urban-biased policies and rising income inequality in China. *The American Economic Review, 89*(2), 306–310.
Yang, X., Gustafsson, B., & Sicular, T. (2021). Inequality of opportunity in household income, China 2002–2018. *China Economic Review, 69*, Article 101684.
Yi, H., Zhang, L., Liu, C., Chu, J., Loyalka, P., Maani, M., & Wei, J. (2013). How Are Secondary Vocational Schools in China Measuring up to Government Benchmarks? *China & World Economy*, *21*(3), 98–120.
Zhang, Y., & Eriksson, T. (2010). Inequality of opportunity and income inequality in nine Chinese provinces, 1989–2006. *China Economic Review, 21*(4), 607–616.
Zhang, Y., Hannum, E., & Wang, M. (2008). Gender-based employment and income differences in urban China: Considering the contributions of marriage and parenthood. *Social Forces, 86*(4), 1529–1560.
Zhang, Y., & Xie, Y. (2016). Family background, private tutoring, and children's educational performance in contemporary China. *Chinese sociological Review, 48*(1), 64–82. %@ 2162-0555.
Zhou, Y., Han, H., & Harrell, S. (2008). From labour to capital: intra-village inequality in Rural China, 1988–2006. *The China Quarterly*, 195.