Monograph:
Nan, J. and Popli, G. (2025) Labour market segmentation and urban-rural wage gap: the role of education. Working Paper. Sheffield Economic Research Paper Series (2025015). Department of Economics, University of Sheffield ISSN 1749-8368

The University Of Sheffield. | School Of Economics

Labour market segmentation and urban-rural wage gap: the role of education

Jiarui Nan and Gurleen Popli

Sheffield Economic Research Paper Series

SERPS no. 2025015

ISSN 1749-8368

17 December 2025

Labour market segmentation and urban-rural wage gap: the role of education*

Jiarui Nan[¶†], Gurleen Popli[ζ]

[¶]School of Economics, The University of Sheffield, Elmfield Building, Northumberland Road, Sheffield, S1 4DT

(email: j.nan@sheffield.ac.uk; telephone: +44 (0)7376465709)

[ζ] School of Economics, The University of Sheffield, Elmfield Building, Northumberland Road, Sheffield, S1 4DT

(email: g.popli@sheffield.ac.uk; telephone: +44 (0)114 222 3485)

Abstract

This paper examines the determinants of the urban-rural wage gap in China within the framework of segmented labour markets. Using nationally representative data from the China Family Panel Studies (2014–2022), we employ Oaxaca–Blinder decomposition at the mean, and Recentered Influence Function (RIF) regression across the wage distribution. A key contribution of this study is the use of alternative definitions of urban and rural status, based on hukou registration and geographic residence, allowing us to capture both institutional and spatial dimensions of inequality. The results show that mean wage disparities are largely explained by compositional differences, particularly in education and access to formal contracts, reflecting segmentation between distinct rural and urban labour markets. Yet rural workers also experience significant lower returns to education. Quantile decompositions reveal that the wage gap widens at higher percentiles, where unobserved or institutional disadvantages become more pronounced, for both men and women. Overall, the findings demonstrate that China's urban-rural wage inequality reflects both unequal endowments and structural segmentation. The definition of "urban" and "rural" critically shapes interpretation and policy implications.

Keywords: return to education, China, wage gap, regional differences, decomposition, recentered influence function, segmented labour markets

JEL classification: J42; J24; R23

---

* Acknowledgments: The data are from China Family Panel Studies (CFPS), funded by Peking University and the National Natural Science Foundation of China. The CFPS is maintained by the Institute of Social Science Survey of Peking University. We are also grateful to Andrew Bryce and Andrew Dickerson for comments on an earlier version of this analysis.

† Corresponding author.

1. Introduction

Large income and wage disparities between urban and rural populations in China have persisted despite rapid economic growth since the "Reform and Opening Up" policy.[1] According to the National Statistical Bureau of China (2020), the per capita disposable income of urban residents was twice that of rural residents. Similarly, Campos et al. (2016) found that annual labour market earnings for urban workers were 1.4 times higher than those for rural workers. These gaps remain a central feature of China's labour market.

While the existing literature often interprets the urban-rural wage gap through the lens of human capital theory (Becker 1964; Mincer 1974), emphasising differences in educational attainment and productivity (e.g., Zhang 2017), this perspective alone cannot fully capture the complexity of China's labour market. Education is important; urban residents consistently have more years of schooling, and returns to education differ across urban and rural settings (Asadullah and Xiao 2020; Guo et al. 2019).[2] However, the urban-rural differences in China are not shaped by human capital alone, but also by institutional arrangements, most notably the hukou system, a household registration system. This household registration system segments the labour market along both spatial and administrative lines.

China's institutional segmentation, resulting from the hukou system, complicates how we define "urban" and "rural." Literature on China has defined workers as urban and rural, either based on their hukou status or their geographical location. If status is determined by hukou registration, the wage gap reflects institutional barriers, where rural hukou holders, even when working in urban areas, often face discrimination and reduced access to secure, high-paying jobs. In contrast, if status is defined by physical residence, the wage gap reflects spatial inequalities tied to regional job opportunities, labour demand, and productivity dynamics. A stricter definition that combines both hukou and geographical location highlights compounded disadvantages faced by rural hukou

[1] China's Reform and Opening-Up policy (1978) transformed the economy from a centrally planned system to a socialist market economy, allowing market forces, private businesses, and foreign investment to play a major role alongside state ownership. During this period, China experienced significant industrialisation, with strong economic growth and a shift from public to private sector employment. For further discussions see Liu (2020) and Lu (2019).

[2] Figure A1 in appendix shows that on average, the years of education for urban residents are considerably higher than for rural residents.

holders who also reside in rural areas. These distinctions matter because they point to different underlying mechanisms of inequality, and therefore imply different policy solutions.

This paper investigates the determinants of the urban-rural wage gap in China using decomposition analyses that explicitly account for the different definitions of urban and rural status. Decomposition methods are widely used to explain earnings disparities between mutually exclusive groups of workers, as they help quantify the contribution of various factors (such as education, labour contracts, household characteristics, etc.) to observed differences and shed light on the underlying hypotheses and explanations (see Fortin et al., 2011, for a discussion of strengths and limitations of decomposition analyses). We apply decomposition at both the mean, using the Oaxaca-Blinder (OB) method, and across the wage distribution, using the Recentered Influence Function (RIF) regression approach. Our analysis relies on nationally representative data from the China Family Panel Study (CFPS), and we conduct separate analyses for men and women. By incorporating both institutional (hukou) and geographic (residence) definitions of urban-rural status, our study provides new insights into how education, labour contracts, sector of employment, and regional factors shape wage inequality in China.

Our contribution to the literature is threefold. First, we go beyond conventional human capital explanations by situating wage disparities within China's segmented labour market, explicitly comparing institutional and spatial dimensions of inequality. Second, we disentangle the roles of endowments and returns to endowments, offering a more nuanced understanding of how observed and unobserved factors drive wage differences. To do this, we go beyond the means, investigating the urban-rural gap at different quantiles of the wage distribution. Third, we provide robust evidence based on a large, nationally representative dataset, addressing common limitations of prior studies that relied on smaller samples or narrower definitions of urban-rural status.

The remainder of the paper is structured as follows. Section 2 discusses the segmented nature of China's labour markets and alternative definitions of urban and rural status. Section 3 reviews the literature on returns to education and wage disparities between urban and rural workers. Section 4 describes the data and methods. Section 5 presents the decomposition results, and Section 6 concludes.

2. Segmented labour markets and urban-rural definition

China's labour market is uniquely shaped by institutional factors that create a division between urban and rural populations, resulting in a dual economic structure enforced through policy. At the heart of this dual economic structure, resulting in labour market segmentation, is the household registration system, known as hukou, which has been in place since 1958. Under this system, every individual is assigned at birth an agricultural (rural) or non-agricultural (urban) hukou, indicating their original social and geographical background. This classification was originally tied to both occupation and place of residence, and was designed to restrict internal labour mobility. Although the link between hukou status and occupation has weakened over time, hukou status remains inherited, with children taking on their parents' hukou type.[3]

Reforms since the 1980s have partially decoupled hukou status from physical residence. As a result, many individuals now live in cities without an urban hukou, particularly rural-to-urban migrants who retain their rural hukou.[4] This mismatch contributes to the complexity of China's labour market. Although reforms have reduced some barriers to internal migration, significant institutional obstacles persist. Internal migrants (in this case, defined as individuals living in an area where they do not have hukou) retain their original hukou type, and crucial public services, including education, healthcare, housing subsidies, and social security, are only accessible in the location specified by one's hukou. For example, a person with a rural hukou living in a city, a rural migrant to an urban area, is typically ineligible for urban public services.

Substantial evidence shows that hukou status influences labour market outcomes. Migrants often face discrimination, lower wages, and limited access to secure well-paying jobs (Zhu, 2015). The hukou system, therefore, enforces both institutional and market-based constraints, contributing to a segmented and non-competitive labour market in which wages are only weakly linked to productivity. Much of the existing literature on China's urban-rural wage gap compares rural

---

[3] Hukou is classified not only by type (urban or rural), but also by specific geographic location (e.g., city or county). However, this paper focuses solely on hukou type, since changing location within a hukou type has historically been easier than changing hukou type, which remains extremely difficult or nearly impossible. Song (2014) provides details of the hukou system, how it has changed over time, and the impact it has on the Chinese economy.

[4] Flow of internal migration in China is predominantly from rural areas to urban areas, so it is much more common to see individuals with rural hukou living in urban areas, rather than the other way around.

migrants with urban hukou holders. Fewer studies examine the wage gap by geographic location, where individuals are classified based on where they live and work. These two approaches capture fundamentally different aspects of inequality.

When we define the urban-rural wage gap based on physical location, we are comparing people working in urban areas to those working in rural areas, regardless of hukou status. In this approach, rural migrants working in cities are classified as urban workers, and vice versa. What this captures is the geographic and economic differences between urban and rural regions, which can arise due to different kinds of jobs available in the two locations, and the skills required to do these jobs. In addition, there are likely to be different labour market dynamics in the two locations, i.e. differences in the demand and supply of workers with different levels of skills, one can expect higher demand and competition for skilled workers in the urban areas, relative to in the rural areas. The urban-rural wage gap here reflects spatial and economic disparities between urban and rural regions, not necessarily institutional barriers. When separating individuals based on their place of residence, we can explore how much of the urban-rural wage gap is due to differences in the characteristics of the workers, and how much is due to the wage structure, where, due to different labour market dynamics, similar workers in the two settings are paid differently.

When we define the urban-rural wage gap by hukou status, we are comparing workers with urban hukou to workers with rural hukou, regardless of where they live and work. For instance, a rural migrant working in a city is still considered a rural hukou holder. Much of the literature focuses on urban locations and compares the rural migrants to urban residents. The wage gap here reflects systemic inequality based on institutional classification, not necessarily geographic location. When separating workers based on their hukou status, we can explore how much of the observed wage gap between urban-rural hukou workers is due to their characteristics, someone with a rural hukou working in an urban area is more likely to work in private sector and have informal contract, and how much is due to the wage structure (returns to those characteristics) where similar individuals receive different pay due to hukou-based discrimination.

In this paper, we use both measures to define an individual's urban and rural status. The first measure is based on geographic divide, where urban status applies to those individuals who are residing in urban areas, irrespective of their hukou status; and the second measure is based on the

hukou status, irrespective of where they are physically located. Our main empirical analysis uses the geographic (location-based) definition as the baseline, classifying individuals by where they live. We include hukou status as a control variable to account for institutional factors.[5] We supplement our main analysis with three robustness checks. The first check is where we define urban-rural status based on hukou, and use geographic residence as a control. The second check is where we define urban and rural workers based on both their physical location and registration status, this definition excludes internal migrants (both rural to urban migrants and urban to rural migrants). The third, and final, check is to focus only on workers in urban locations, comparing rural-hukou holders (rural-to-urban migrants) with urban-hukou holders.

3. Literature Review

3.1. Returns to Education Urban vs. Rural Workers

Education consistently plays a crucial role in shaping labour market earnings, but the extent of returns varies widely across urban and rural contexts. The majority of cross-country and country-specific studies show that returns to education are significantly higher in urban areas.

Peet et al. (2015), using Living Standards and Measurement Surveys (1985–2012) across 25 developing countries, found that in most African and Eastern European nations, urban workers received higher wage returns to education compared to their rural counterparts. Similar evidence exists for Asia. For India, Rani (2014) reported urban wage premiums of 27.7%, 48.7%, and 72.7% for elementary, secondary, and tertiary education, compared with only 19.5%, 30.5%, and 35.4% in rural areas. Kanjilal-Bhaduri and Pastore (2018), using the 2011–2012 Employment and Unemployment Survey of India confirmed higher urban returns for both men and women, regardless of employment type (both permanent and casual workers).

For Thailand, Warunsiri and McNown (2010) initially found little difference between urban (11.5%) and rural (11.3%) returns under OLS estimation. However, using a pseudo-panel approach, they revealed a more substantial gap, 18.9% for urban workers versus 14.2% for rural workers,

[5] There are issues of selection as well, as internal migrants are likely to be different from those who do not migrate, and live and work in areas where they hold a hukou. Controlling for hukou status goes some way to address this selection.

arguing that higher urban returns incentivise both seasonal and permanent migration of rural workers to cities. In the Philippines, Choi (2021) showed that secondary education yielded higher urban returns, though additional vocational training boosted rural workers' earnings slightly more than urban workers.

In China, as well, literature has documented a persistent urban advantage. Weng et al. (2016), using China Health and Nutrition Survey data (1989-2011), found consistently higher returns in urban areas relative to rural areas since 2004. Further, they find that women have higher returns to education, relative to men, and the gap is bigger in the urban areas and much smaller in the rural areas. Asadullah and Xiao (2020) also compare the regional gap in returns to years of education based on individuals' residential status (their geographical location); they use data from the China General Social Survey of 2010 and 2015. The authors reported that rural returns were negligible in 2010 (0.1%) and rose modestly by 2015 (2.6%), while urban returns remained far higher (10.4% in 2010 and above 7% in 2015).

Fu and Ren (2010) conduct an analysis that combines the different methods, hukou status and geographical location, to define individuals' urban and rural status. They use data from the 2005 National Population Sample Survey and focus on southeast China. The sample is divided into three groups: urban residents, migrant workers in urban areas with rural hukou, and rural residents. They find a descending rank of return to different levels of education, for local urban, migrants, and rural workers, respectively. For example, the premiums to tertiary education compared with no schooling, are 115.4%, 89.7% and 73.8%, respectively, for urban, migrant and rural workers.

Gao and Li (2022) study the time variation in return to years of education between Chinese urban and rural wage earners, based on data from multiple social surveys.[6] They define the urban and rural workers based on their registration status without considering the residential location of the workers. They concluded that from 1993 to 2019, the difference in returns to education between workers holding urban hukou and rural hukou widened, increasing from 2.82% to 7.68%, reaching its peak in 2010. However, their analysis suffers from an unbalanced number of observations of

[6] Data sets used are China Health and Nutrition Survey (CHNS), Chinese Household Income Project (CHIP), China General Social Survey (CGSS), China Family Panel Studies (CFPS), China Household Finance Survey (CHFS) and Chinese Social Survey (CSS).

workers in each year, and a limited set of controls in their specifications, as a trade-off for including multiple datasets in the analysis.

Most of the above studies on China document the differential returns to education between urban and rural areas, defined either by registration or residential location. But they do not explore the impact of these differences in returns on the urban-rural wage gap.

3.2 Earnings Gap and the Decomposition Method

Beyond documenting returns, researchers have sought to explain the urban-rural wage gap by separating it into two components: the explained part (differences in endowments, such as education or experience), and the unexplained part (differences in returns, often interpreted as discrimination or labour market segmentation).

The OB decomposition (Oaxaca 1973; Blinder 1973) is the most widely used method to investigate wage differentials. For example, Pereiral and Galegol (2010) analysed regional wage gaps in Portugal; the authors study the wage differentials between Lisboa and four other districts. They found that education contributes positively to the endowment, reflecting the gap in education achievements between Lisboa and other districts. However, in terms of payoff, education shows a negative effect which eliminates the wage differentials between regions. Michaelsen and Haisken-DeNew (2015) applied the method to Mexico, showing that urban workers' higher education attainment explained much of the wage gap, while payoff differences were minimal.

For China, Sicular et al. (2007) examine the income differences at the household level between urban and rural areas. They find that in both years 1995 and 2002, the total income difference is driven more by the observable characteristics (explained part) than the return to characteristics (unexplained part). Further, in 2002, the returns show a negative effect in explaining the total wage difference, which means the gap generated by observable characteristics is offset by the wage structure. Turning to the education variable (average education of working-age adults in the household), the endowments and returns contribute 22.9% and 7.1% to the income difference, respectively, in 1995. However, in 2002, the effect of endowment increases to 30.3%, but the effect of returns becomes negative at -5.0%.

Also focusing on the case of China, Alam and He (2022) examine the decomposition of wage

gaps between rural-migrant and local urban workers, using data from the China Household Income Project. They provide evidence that discrimination against migrant workers is quite limited (i.e. the unexplained part of the wage gap is small). The decomposition of the benchmark model shows that 82.9% of the wage difference can be explained by individual endowments. Further, including occupation and industry variables, the extended model can explain 91.8% leaving only 8.2% as the source of the wage gap driven by differences in returns. Education contributes significantly to both explained and unexplained wage differences. However, given that the unexplained part of the wage gap is quite small, the overall wage gap is mainly driven by the differences in education endowments.

More recently, detailed decomposition methods such as RIF regressions have been employed to examine gaps across the wage distribution. Messinis (2013), for instance, studied wage differences between urban residents and rural migrants in urban China; their findings suggest that migrant workers suffer from wage disadvantage across all wage quantiles with most of the disadvantage being driven by differential endowments. Further, education explained about one-third of the gap for low-income groups and up to half for top earners.

Taken together, this body of research demonstrates that while endowments, particularly education, explain much of the urban-rural wage gap, differences in returns remain important in certain contexts, especially when institutional barriers such as hukou status are considered. However, there are very few studies which comprehensively define the urban and rural areas, which we do in our paper.

4. Data and methods

4.1 CFPS Dataset and Variables

This paper uses data from CFPS, a large nationally representative longitudinal survey in China designed by Peking University; it contains detailed information on education and labour market outcomes at the individual level. The survey has been administered every two years since 2010, with a target sample of 16,000 households (or 50,000 individuals) covering 25 (out of 34) Chinese provinces in the initial wave. We use five waves of the data, covering survey years 2014, 2016, 2018, 2020 and 2022, the period over which information on our key variables, including earnings, is

consistent.[7] The CFPS collects information aligned with established measures on labour market returns. It has several advantages in this context, including the information required to derive hourly earnings and the availability of information on other demographic, job-related and household characteristics. In addition, CFPS provides information on alternative measures to define an individual's urban and rural status, based on China's hukou system and the individual's place of residence and work.

Our sample consists of working-age individuals (16-60 for males and 16-55 for females) who are not currently in education and who are wage earners in their main job.[8] After excluding missing values of core variables used in the analysis and trimming 1% of the top and bottom earnings percentiles to avoid outliers, our maximum sample contains 33,947 observations (these are person-wave observations).[9] We look at the urban-rural difference in earnings separately for men and women. We have 20,010 person-wave observations on men, and 13,937 person-wave observations on women. Table 1A provides statistics on the overlap between the two measures of urban and rural status. Although more than half of both male and female workers reside in urban areas, it can be seen from the table that only about 50% of them hold an urban hukou. Specifically, 63% of men (12,613 of 20,010) reside in urban areas, of these, however, only 52% (6543 of 12,613) have an urban hukou; 48% of the men residing in urban areas have a rural hukou, these are classified in the literature as rural-to-urban migrants. We get similar figures for women. On the other hand, most of the workers residing in rural areas hold a rural hukou, for both men and women.

Education years

We focus on individuals' finished education levels, from which we obtain years of education, according to the transformation officially provided by CFPS (see Appendix Table A1).[10] There are in total seven levels of formal education in China, with the base being no schooling, which are all

[7] We pool multiple waves of the longitudinal dataset, so some individuals may appear more than once in the sample. To account for this, we use standard errors clustered at the individual level. We do not use panel fixed-effects analysis because our key independent variable, education, is time-invariant. Panel methods, particularly fixed-effects models, eliminate the effects of time-invariant variables; therefore, we do not incorporate a longitudinal design in our analysis.

[8] CFPS does not have income information for the self-employed workers.

[9] Missing values for the core variables appear to occur at random; roughly 18.6% of the observations are missing.

[10] Years of unfinished levels are also covered in CFPS, but this measurement suffers from a considerable number of missing values.

covered in our analysis, and the corresponding education years range from 0 to 22. Figure 1 illustrates the distribution of education levels for urban and rural workers, respectively. The predominant finished level of education in China remains lower-high school, for both men and women in urban and rural areas. In addition, the proportion obtaining tertiary education is considerably higher among urban workers, and the gap is larger for females. Table 2, gives the descriptive statistics for all variables, including mean years of education for urban and rural workers by gender. In Table 2 the workers are divided as urban and rural based on their geographical (residential) location, and not hukou status. It shows that male urban workers on average have 11.5 years of education, 2.3 years higher than male rural workers at 9.2 years; similarly, female urban workers have 11.8 years of education, 2.3 years more than female rural workers at 9.5 years.

Hourly earnings

Consistent with the established literature, our main dependent variable is (log) gross hourly pay (in Yuan RMB). All wages have been deflated according to each year's annual CPI, from 2014 to 2022, with 2014 as the baseline. It is constructed based on an annualised measure derived from average usual monthly earnings and weekly working hours. It includes net wage, subsidies, any kind of cash rewards and bonuses, and any payment in kind. The formula to calculate the hourly earnings is as follows:

hourly wage = [(monthly wage*12)/52]/hours worked in a week

Table 1B highlights systematic earnings differences by hukou status and place of residence. Individuals with urban hukou have higher mean log hourly earnings than those with rural hukou in both urban and rural areas, indicating a persistent hukou advantage. In addition, urban residents earn more than rural residents within each hukou category. These patterns suggest that both hukou status and place of residence are independently associated with higher earnings, with the highest earnings observed among urban-hukou individuals living in urban areas and the lowest among rural-hukou individuals living in rural areas. The patterns are similar for men and women.

In the Appendix, Table A1 presents mean log gross hourly earnings (hence forth referred to as log hourly earnings) for different education levels, separately by gender and region of residence. Overall, urban male workers earn 0.190 log points higher wages than rural workers, and women in

urban areas earn 0.248 log points higher wages than their rural counterparts; these differences are statistically significant. Urban workers, men and women, earn higher wages than rural workers at every education level, other than tertiary postgraduate education, where there is no statistical difference in the wages of the workers in rural and urban areas; however, note from Figure 1, there are very few workers with tertiary postgraduate education in rural areas. We also see no difference in wages among urban and rural women with no education. Largely, the regional wage gap increases with education, with the highest regional gap observed for those with a tertiary undergraduate degree.

In Figure 2, we provide a distribution of log hourly wages for men and women in rural and urban areas. For both men and women, the distribution of log wages in the urban area indicates more variation and is slightly right-skewed. The distribution of log wages in rural areas is narrower and peaks higher, indicating a more concentrated distribution around the mean. The peak of the urban wages is to the right of the peak of the rural wage distribution; however, there is an overlap between the two distributions over the range of observed wages.[11]

Other Covariates

The CFPS contains detailed information on personal and work-related characteristics which are well-established in the literature to affect earnings. Our controls for personal and household characteristics include age (and age squared), a dummy for ethnicity (minority), and 4 categories of province of residence (East, Northeast, Middle, West).[12] For our main analysis, we define urban and rural workers by their residential location, and hence include a dummy if the individual registration status is urban (urban hukou), this captures any (dis)advantage the individuals might face for being in a location for which they do have a hukou. Household characteristics are captured by a dummy for marital status (married), and dummies for children of different ages in the

[11] The overlap in distribution is important for decompositions as it indicates that the common support assumption is not violated.

[12] *Northeast*: Liaoning, Jilin, Heilongjiang; *East*: Beijing, Tianjing, Hebei, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, Hainan; *Middle*: Shanxi, Anhui, Henan, Hubei, Hunan; *West*: Neimenggu, Jiangxi, Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Shanxi, Gansu, Ningxia, Xinjiang.

household (ages 0 to 5, and 6 to 15). Work-related characteristics include dummy variables to capture having a formal contract and working in the public sector.

Table 2 gives the summary statistics for all covariates by region and gender. We can see several characteristic differences between the urban and rural regions. As noted above, urban workers are more likely to have more years of education, and are more likely to have an urban hukou, for both males and females. Urban workers are more likely to be older, and less likely to have minority status. While urban workers, both male and female, are marginally more likely to be married, a greater proportion of rural households have young children, regardless of whether the children are in the 0–5 or 6–15 age group. For work-related controls, a higher proportion of urban workers, both male and female, have formal contracts and are employed in the public sector.

We do not include occupation or industry controls in our wage regressions, as doing so would cause the estimated rural–urban wage differences to reflect only within-occupation or within-industry variation, rather than the overall wage gap. Moreover, since our focus is on understanding the role of education in explaining urban–rural wage differentials, including occupation and industry dummies would remove an important channel through which education influences wages. We do check for the urban and rural distribution of different occupations and industry, see Appendix Tabel A3, which shows that while urban employment is more diversified and service-oriented and rural employment is concentrated in manufacturing, construction, and production-related occupations, all occupations and industries nonetheless have a significant presence in both rural and urban areas.

4.2 Decomposition methods

We quantify the relationship between education and earnings using the Mincer wage equation (1974), widely adopted in the literature, where the natural log of gross hourly wages is regressed on a vector of covariates, as follows

$$y_{ig}^{M} = X_{ig}^{M}\beta_{g}^{M} + \varepsilon_{ig}^{M} \tag{1}$$

where g is the region indicator, r for rural residents and u for urban residents; M indicates male, y_{ig}^{M} is the log gross hourly wage of an individual $i \in g$; X_{ig}^{M} is the vector of covariates, including a

constant; β_g^M is the vector of coefficients for the returns to all variables in the covariate vector; and ε_{ig}^M is the error term. A wage equation, identical to equation (1) is estimated for women, W, as well.

By estimating the wage regression separately for urban and rural samples, for both men and women, we can decompose the observed urban-rural wage gap in average wages into the contribution of the characteristics (average covariates) and the returns to those characteristics (the estimated coefficients). This is done separately for men and women. Under the assumptions of linearity and zero-conditional mean errors, the OB decomposition of the urban-rural wage gap for men, can be expressed as:

$$\underbrace{\bar{y}_u^M - \bar{y}_r^M}_{\Delta_\mu} = \underbrace{(\bar{X}_u^M - \bar{X}_r^M)\hat{\beta}_u^M}_{\Delta_{\mu X}} + \underbrace{\bar{X}_r^M(\hat{\beta}_u^M - \hat{\beta}_r^M)}_{\Delta_{\mu S}} \quad (2)$$

where $\bar{y}_g^M$ is the mean log hourly wage for men in group $g \in u, r$; $\bar{X}_g^M$ is the set of covariate means for men in group g; and $\hat{\beta}_g^M$ are the parameters of the regression models estimated for urban and rural residents separately, including a constant.

The left-hand side of equation (2), Δ_μ, is the mean urban-rural wage gap for men. The first component on the right-hand side of equation (2), $\Delta_{\mu X}$, refers to the part of the wage gap explained by the differences in the composition, or endowments, of urban and rural workers, and the second component, $\Delta_{\mu S}$, reflects the differences in the wage structure, that is the differences in the estimated coefficients, often referred to as the unexplained part. However, we should be cautious in the interpretation of the unexplained part (see arguments in Nguyen et al., 2022) as it also includes the effects of unobservables. The reference group for the twofold decomposition used here is urban workers. Equation (2) give the aggregate compositional and structural parts of the regional wage gap, these can be expressed in terms of sums of the covariates, which give us the contribution of individual covariates to the total wage gap, as follows:

$$\bar{y}_u^M - \bar{y}_r^M = \sum_{k=1}^{K} (\bar{x}_{u,k}^M - \bar{x}_{r,k}^M)\hat{\beta}_{u,k}^M + \sum_{k=1}^{K} \bar{x}_{r,k}^M(\hat{\beta}_{u,k}^M - \hat{\beta}_{r,k}^M) \quad (3)$$

Where $\bar{x}_{g,k}^M$ is the mean of the k^{th} component of vector X_{ig}^M, and $\hat{\beta}_{g,k}^M$ is the corresponding coefficient. A similar decomposition is conducted for the urban-rural wage gap for females.[13]

[13] We acknowledge two issues with the OB decomposition. First is the reference group (index) problem: we take urban

While the OB decomposition serves as a useful starting point, allowing us to decompose the urban-rural gap in mean wages, we go beyond the decomposition of the mean wages and investigate how the urban-rural wage gap changes at the different quantiles for the wage distribution. To do this we estimate a RIF regression proposed by Firpo et al. (2009). The RIF for the θ-quantile (q_θ), $\theta \in (0,1)$, of y_{ig}^M is given as

$$RIF(y_g, q_\theta) = q_\theta + [\theta - d_{g,\theta}]/f_{Y_g}(q_\theta) \tag{4}$$

where to keep notation simple we supress the index i for the individual, and M for gender; $f_{Y_g}(q_\theta)$ is the density function of y_g computed at quantile q_θ and $d_{g,\theta}$ is the dummy variable taking value 1 if $y_g \leq q_\theta$, and zero otherwise. Following Fortin et al. (2011), the RIF is assumed to be a linear function of the covariate vector, such that,

$$RIF(\mathrm{Y}_g, q_\theta) = \mathrm{X}_g \beta_g^\theta + \nu_g^\theta \tag{5}$$

where β_g^θ is the vector of coefficients for the θ-quantile, and ν_g^θ is the error term. Under the properties of the RIF (for details see Fortin et al., 2009 and 2011) equation (5) is also referred to as the unconditional quantile regression. We estimate this separately for urban-male and rural-male workers, and for the urban-female and rural-female workers.

The difference in the θ-quantile of the wages for the urban and rural workers, $q_{\theta,u} - q_{\theta,r}$, can then be decomposed as:

$$\underbrace{q_{\theta,u} - q_{\theta,r}}_{\Delta_\theta} = \underbrace{(\bar{X}_u^M - \bar{X}_r^M)\hat{\beta}_u^\theta}_{\Delta_{\theta X}} + \underbrace{\bar{X}_r^M(\hat{\beta}_u^\theta - \hat{\beta}_r^\theta)}_{\Delta_{\theta S}} \tag{6}$$

The left-side of the equation (6), Δ_θ, yields the urban-rural wage gap at the θ^{th}-quantile; this regional wage gap at each quantile is estimated separately for men and women. The first term on the right-side, $\Delta_{\theta X}$, is the composition (or endowment) effect at the θ^{th}-quantile, and the second term, $\Delta_{\theta S}$, is the structural effect of the θ^{th}-quantile. Similar, to the OB decomposition, equation (6) can be used to obtain detailed decomposition identical to equation (3) by using the quantile specific

workers as the reference group, assuming their returns to characteristics more closely reflect true productivity-based returns. Second is the base group problem, which arises with categorical covariates (e.g., marital status, formal contract) since the omitted category affects the interpretation of the detailed, though not the aggregate, decomposition. Several approaches have been proposed to address this issue, but no consensus exists. Following Fortin et al. (2011), we select omitted categories based on substantive reasoning to ensure meaningful interpretation.

estimated coefficients, $\hat{\beta}_g^{\theta}$. The OB decomposition is a special case of this when the RIF is evaluated at the mean.

5. Empirical Results

5.1 Wage equations

We present the estimates from the wage regression by region and gender in Table 3. Returns to education are significantly higher in urban areas for both men and women. In urban regions, each additional year of education increases wages by 5% for men and 5.5% for women—nearly double the returns observed in rural areas at 2.5% and 2.6% respectively for men and women. Having an urban hukou is associated with higher wages for individuals (both men and women) residing in rural areas. Among men, those in urban areas with an urban hukou earn 4.3% more than their counterparts with rural hukou, while men in rural areas with an urban hukou earn 8.7% more than those with rural hukou. For women, the wage advantage is even greater for having an urban hukou, 9.1% in urban areas and 13.3% in rural areas, when compared to women with rural hukou. Holding a formal contract results in higher wages across both urban and rural regions, with a larger impact in rural areas. Men with formal contracts earn nearly 13% more in urban areas and 15% more in rural areas, compared to those without such contracts. For women, the wage premium is even higher, 22% in urban areas and 27.4% in rural areas. Employment in the public sector is linked to lower wages for urban men, but for women, it is associated with higher wages.

Age exhibits the typical quadratic relationship with wages, increasing with experience before tapering off. Belonging to a minority group is associated with higher wages for women only. Marital status impacts wages differently by gender: it has a positive effect for men but a negative effect for women. Having young children (ages 0–5) does not significantly affect wages. However, having older children (ages 6–15) is associated with lower wages for both men and women across urban and rural areas. The negative impact is stronger for rural men and urban women. Regional differences are also notable. Compared to workers in the western region, men in the eastern and central regions earn more. Among women, those in the northeast earn less, while those in the east earn more. There is a general upward trend in real wages over time.

Estimates from the wage regression indicate that there are differences in returns to different characteristics across rural and urban areas, and for men and women. We next turn to decomposition results, where we look at the urban-rural differences for men and women separately.

5.2 Decomposition of the wage gap between urban and rural workers

From the descriptive statistics (Table A2) we know that average wages by education differ significantly across urban and rural areas, with urban workers, in most education categories on an average earning more than the rural workers. Average years of education is also higher in the urban areas, relative to rural areas, and from the OLS regression, we find that the estimated returns to education for urban workers is higher than those for the rural workers. We next do a decomposition to see how much of the average urban-rural wage differential is a result of the differences in the endowments and how much is due to the returns to these endowments. We first present results for decomposition at the mean, and then at the different quantiles of the wage distribution using the RIF approach.

Decomposition at the mean

Results from the decomposition at the mean, using the Oaxaca-Blinder decomposition, are presented in Table 4. This decomposition is based on a counterfactual scenario in which rural wage earners retain their own characteristics (endowments), but face the wage structure of urban workers, i.e. urban workers serve as the reference group, and the estimated coefficients from their wage regression are used to construct the counterfactual wage outcomes for rural workers.

There is a statistically significant difference in the average urban and rural wages, for men the difference is 0.190 log points and for women it is 0.247 log points. The decomposition indicates that most of the urban-rural gap is explained by observable characteristics, rather than unexplained or residual factors. Specifically, 91.6% of the male income gap (0.174 of 0.190) and 91.5% of the female income gap (0.226 of 0.247) can be attributed to differences in the characteristics such as education, hukou status, and employment type. The unexplained component—potentially reflecting discrimination or unobserved heterogeneity—accounts for only about 8–9% of the gap for both genders.

The detailed breakdown shows that education is the most significant contributor to the explained component. For men, education accounts for 0.113, and for women, 0.127, underscoring the importance of educational attainment in the urban–rural wage differential. The unexplained component of education (0.226 for men and 0.277 for women) suggests that returns to education are higher in urban areas, particularly for female workers. Other personal characteristics, including age and ethnicity, have negligible contributions to the gap. In contrast, urban hukou status contributes significantly to the explained portion (0.017 for men and 0.037 for women), confirming the structural advantage associated with urban hukou status.

Regarding labour market characteristics, the presence of a formal contract contributes positively to the explained gap (0.025 for men and 0.035 for women), implying that urban workers are more likely to hold formal employment with higher returns. However, the unexplained portion is negative, particularly for women (-0.025), suggesting that even when rural women have similar contractual status, they earn less than comparable urban women. The public-sector variable shows gender heterogeneity: while it slightly reduces the explained gap for men (-0.006), it increases it for women (0.011), implying that public employment offers women more income stability and relative parity across regions.

For household characteristics, marriage and children have modest but noteworthy effects. Being married slightly increases the explained component for men but has minimal impact for women. The unexplained effect of marriage for women (0.056) may reflect gendered household labour divisions that disadvantage rural women's earning capacity. Having children contributes positively to the explained gap for both sexes but remains quantitatively small.

The year and province controls account for time-specific and province-specific effects, such as changes in the labour market conditions or policy shifts that affect wages independently of individual characteristics. Both the time-specific and province-specific indicate that over time and over provinces returns to rural workers and urban workers have changed at different pace (as can be seen in the OLS results).[14]

[14] For example, the year coefficients show a clear upward trend over time relative to the reference year, for both urban and rural areas. However, the magnitude of coefficients for urban areas are much larger relative to rural areas, indicating a different pace of growth overt time.

In summary, the urban–rural income gap is predominantly explained by differences in human capital, hukou status, and labour market structure, with educational disparities being the single most important factor. However, the persistence of small but significant unexplained effects—particularly in education and hukou status—suggests that structural and institutional biases continue to contribute to urban–rural inequality.

In the OB decomposition, the constant term, included in the unexplained component, is the gap in the urban-rural wages when all covariates are set to zero. This indicates baseline advantage/disadvantage, or group differences not explained by observed characteristics or their returns.[15] In our analysis we find that the contribution of the constant to be very high, negative, and significant, indicating that it narrows the unexplained wage gap (i.e., the unobserved characteristics favour the rural workers).

Quantile decomposition

We move beyond the mean urban-rural wage gap to examine how this gap varies across different points in the wage distribution. Using a quantile decomposition approach with the RIF method, we estimate the wage gap at the 10th, 50th (median), and 90th percentiles. Results for men and women are shown in Tables 5.1 and 5.2, respectively.

For both men and women, the urban-rural wage gap widens with the wage level. Among men, the gap grows from 0.086 log points at the 10th percentile to 0.175 at the median, and doubles to 0.349 at the 90th percentile. For women, the gap rises from 0.140 log points at the 10th percentile to 0.253 at the median, and 0.396 at the 90th percentile. Similar patterns are observed at the 25th and 75th percentiles (not shown), indicating a monotonic increase in urban-rural wage gap across the distribution.

The explained portion of the wage gap, attributable to observable characteristics, declines with wage level for both men and women. For men, 133.7% of the gap at the 10th percentile is explained by observable factors, implying that rural men would earn more than their urban counterparts if they

[15] The interpretation must be done carefully, as zero for all covariates may not be meaningful, such as age, and education. Further the size of the intercepts of the wage regression and hence their difference in the decomposition, will depend on how the categorical covariates are specified, i.e. the base category. For more discussion see Bryan et al. (2022).

had the same endowments; the unexplained part as a consequence is negative, however it is insignificant. At the 90th percentile, however, only 64.5% is explained, indicating increasing importance of unobserved or structural factors at higher wage levels. A similar trend is evident for women, with the explained share falling from 108.6% to 72.7% between the 10th and 90th percentiles.

At the lower end of the distribution, the unexplained component, capturing structural effects and differential returns, is negative, suggesting that rural workers may possess unobserved advantages that partially offset observable disadvantages. However, this component becomes positive, and its share increases at higher percentile, indicating growing structural disadvantages for rural workers. These structural effects are more pronounced among men, both at the bottom and the top of the distribution.

The contribution of specific factors also varies along the wage distribution. Education is a dominant determinant of the explained wage gap at all wage levels, with a higher share of explained component at lower percentiles and its share falls higher up. For men, education accounts for 85% (0.073 of the 0.086) of the wage gap at the 10th percentile, falling to 65.7% (0.115 of 0.175) at the median and 44.4% (0.115 of 0.349) at the 90th percentile. Formal contracts also have a significant impact at the bottom, with their contribution falling at the top of the distribution for men. Urban hukou status shows contrasting effects across the wage distribution: at the 10th percentile, its impact is insignificant, whereas at the 90th percentile, it significantly narrows the urban-rural wage underscoring the growing importance of urban hukou for higher wages. For women, education similarly plays a key role, although the decline in its explanatory power is concentrated in the upper half of the distribution. Formal contracts have similar effects as for men. However, urban hukou status consistently benefits rural women by narrowing the wage gap throughout the distribution.

Within the unexplained component, education remains significant. Equalizing returns to education would increase rural men's (women's) wages, more so at the 10th percentile than at the 90^{th} percentile. For both men and women, equalizing returns to urban hukou would reduce rural wages at the 10th percentile, while having little effect at the 90^{th} percentile. Returns to formal contracts also vary across the wage spectrum; for both men and women returns to formal contracts have an insignificant effect at the lower end of the wage distribution, whereas equalising the returns

to formal contracts in urban and rural areas would reduce the rural wages significantly, increasing the urban-rural wage gap.

Age, used as a proxy for labour market experience, has limited explanatory power throughout the distribution for men. Yet, in the unexplained component, age remains important and aligns directionally with education. For women, age counteracts education at the lower end, equalizing returns to education would raise rural wages, while equalizing returns to age would widen the wage gap. At the upper end, returns to age and education reinforce each other, boosting rural women's wages.

Other factors such as marital status and having children show varied effects across the distribution. From the unexplained component, we can see that marriage imposes a greater penalty on rural women across all percentiles. Among rural men, marriage is associated with a wage advantage at the bottom of the distribution but becomes a disadvantage at the top.

Overall, the findings from the quantile decomposition demonstrate that the urban–rural wage gap is largely attributable to differences in endowments at the lower end of the wage distribution, while disparities in the returns to these endowments play an increasingly important role at the upper end. This supports the existence of a segmented labour market in China. The degree of segmentation varies along the wage distribution: lower-wage earners in rural areas appear to fare better in relative terms, while higher-wage earners benefit more from being in urban areas. This highlights the nuanced and stratified nature of labour market inequality in China, with possible, low paying jobs concentrated in the rural areas and the higher paying jobs concentrated in the urban areas.

5.3 Examining the heterogeneity in the definition of urban and rural Status

As explained above, in China, individuals' urban-rural status can be defined using two alternative ways: geographical location and household registration. For our main analysis we used geographical location as the definition, we now do a series of robustness checks for the different ways of defining urban and rural. Differences in how the urban and rural samples are defined are likely to influence the estimated magnitude of returns to education across these groups (Fu and Ren, 2010; Zhu, 2015), thereby affecting both the size and composition of the explained and unexplained

components in the decomposition analysis.

First, we do the decomposition using the registration status as a way of defining rural and urban workers. In Appendix Table A4 we report the descriptive statistics for the sample when workers are divided based on their registration (hukou) status. A comparison of urban-rural classifications based on geographical location (Table 2) versus hukou registration status (Table A4) reveals systematic differences. Hukou-defined urban residents are older, more educated, and more likely to hold formal contracts and public-sector jobs than their geographically defined counterparts. They are also more likely to be married but less likely to have young or school-age children. These patterns indicate that hukou-based urban residents represent a more privileged segment of the population, while the geographically defined urban category captures a broader, more heterogeneous group.

In Table A5 we report the results from the OLS regression on registration based definition of urban-rural workers. Returns to education are higher for urban hukou holders relative to rural hukou holders. Having an urban residence has no significant impact on the wages of the urban hukou holders, whereas for rural hukou holders having an urban residence has a significant impact. For rural hukou holders, returns to working in urban areas are more than returns to an additional year of education, this potentially explains the large rural to urban migration in China.

The urban-rural average wage gap (see Table A6) increases when we use the registration based definition of urban and rural, the gap for men increases to 0.199 log points (compared to 0.190 log points earlier) and for women it increases to 0.310 log points (compared to 0.247 log points earlier). The explained portion of the gap for men now is above 100%, and the unexplained is negative, suggesting that if the observable characteristics were the same between the two groups, rural hukou holders would earn more than the urban hukou holders. However, when we look at the RIF decomposition (Tables A7.1 and A7.2), this advantage of male rural hukou holders is true only for the bottom half of the wage distribution, at the 90th percentile the explained part of the gap reduces to 78.4%, indicating that any advantage that rural hukou holders might have, is only in low paying jobs. The pattern for women is slightly different. At the mean the explained part of the wage gap is close to 95%, which increases to above 100% at the median, but at the 10th and 90th percentile of the wage distribution the explained part of the wage gap reduces.

We also do a robustness check where we define the workers using both the registration status

and their residence (see Appendix Tables A10, A11.1, and A11.2). This comparison removes all migrants from the sample, thus reducing the sample size. The urban-rural wage gap is the highest for this sample, for both men and women. Looking at decomposition the unexplained part is negative at the mean, median, and the lower part of the distribution, though only significant for males but not females, indicating that rural workers with rural hukou have an advantage in low paying jobs, and if the returns to characteristics were equivalised then they would earn much higher wages.

The final robustness check we do is to focus only on the urban residents and comparing the urban hukou holders with the rural hukou holders. In doing this we can focus only on the institutional dimension of the inequality. Results of the decomposition, both at the mean and RIF, over the sample of urban residents are presented in Appendix Tables A13, A14.1 and A14.2. The findings are broadly consistent with what we find in other robustness checks. At the mean and in the lower part of the wage distribution rural hukou holders have an advantage, which is significant for males but not females, whereas at the higher wages they are disadvantaged both in terms of their characteristics and the returns to these characteristics. Using this definition we also find that the role of the constant term in the decomposition for men reduces, it becomes small and insignificant; indicating that any advantage rural-hukou holders have in rural areas, disappears in the urban areas. For females the constant term remains large, negative and significant.

Education is the most consistent determinant, contributing to both the explained and the unexplained wage gap, across all definitions; with its importance decreasing as we move up the wage distribution. Labour market factors (especially having a formal contract) also consistently emerge as another important factor to explain the observed urban-rural wage gap across all definitions.

6. Conclusions

Using large scale nationally representative data from the CFPS 2014-2022 this paper investigates the determinants of the urban-rural wage gap in China using decomposition analysis, within the context of the segmented labour markets in China. In our analysis we account for the different definitions of urban and rural status. We use both the traditional Oaxaca-Blinder decomposition to

investigate the urban-rural wage gap at the mean, and the RIF based decomposition to investigate the determinants of the wage gap at different percentiles of the wage distribution.

In our empirical analysis we have four types of workers differentiated by hukou status and place of residence: rural hukou holders living in rural areas, rural hukou holders living in urban areas, urban hukou holders living in rural areas, and urban hukou holders living in urban areas. The regional distribution of workers largely aligns with their hukou status: most urban hukou holders reside in urban areas, and most rural hukou holders remain in rural areas. However, a significant share of rural hukou holders have migrated to urban areas (rural-to-urban migrants), while urban hukou holders living in rural areas are negligible. These spatial and institutional distributions provide the structural basis for a segmented labour market in which both geography and hukou status determine access to high-wage employment and social entitlements.

When comparing workers in urban and rural areas, the average wages in the urban areas are higher than those in the rural areas. The decompositions (both at the mean and across the distribution) reveal that the majority of the observed wage differences can be attributed to variations in endowments, such as education, and access to formal contracts; this is especially true at the low end of the wage distribution. Rural workers tend to have lower educational attainment, and are less likely to be employed with formal contracts, leading to lower earnings. Furthermore, the returns to education are significantly lower in rural areas, while holding an urban hukou yields considerable wage advantages. Interestingly, the constant term in the decomposition is negative and statistically significant, suggesting the presence of unobserved advantages for rural workers, possibly reflecting compensating wage differentials, informal income sources, or lower costs of living in rural regions.

The comparison between urban hukou and rural hukou workers captures institutional inequality embedded in the hukou system. Here too the decomposition analysis indicates that wage disparities between these two groups are primarily explained by differences in observable characteristics, especially at the lower end of the wage distribution, yet significant differences in returns to education persist, which become more important at the upper end of the wage distribution. Rural hukou holders receive lower returns on their human capital than urban hukou holders, and urban residence further amplifies wage advantages. These findings underscore the role of the hukou system as a form of institutional segmentation, which effectively operates as a mechanism of

restricting rural-hukou holders' access to urban employment opportunities.

When comparing the urban area workers with the rural area workers it is crucial to acknowledge the substantial number of rural hukou holders residing in urban areas. We do this by including an urban hukou dummy in our analysis. However, this does not address the general equilibrium effects of the rural-to-urban migrants. The inflow of rural hukou migrants into urban labour markets introduces a dual effect. On one hand, rural hukou migrants are typically likely to occupy low-wage jobs, depressing the average urban wages, and concentrating urban hukou holders in higher-paying occupations. On the other hand, rural hukou migrants represent a self-selected group, potentially more motivated and productive than the average rural population. To understand this further in our analysis we do a comparison between rural hukou holders living in urban areas and urban hukou holders living in urban areas, isolating institutional disparities within the same spatial setting. The decomposition results show that most of the wage gap between these groups is also explained by differences in observed characteristics, especially lower educational endowments and access to jobs with formal contracts. Further, rural hukou holders in urban setting continue to experience significantly lower returns to education, demonstrating that institutional segmentation persists even when spatial inequality is held constant; these differences are once again more evident at the upper end of the wage distribution. The constant term in this decomposition becomes negligible, implying that any unobserved advantages associated with being a rural hukou holder are no longer evident in the urban labour market. This situation exemplifies a dual structure within the urban economy, replicating spatial inequality within urban spaces.

Finally, we compare urban hukou holders in urban areas with the rural hukou holders in rural areas, each living within their hukou-designated area; the results reveal that most of the wage gap is again explained by observable characteristics, with the main unexplained component being the differential returns to education. This pattern suggests that both spatial and institutional inequalities reinforce one another: even when individuals remain within their designated domains, structural differences in human capital accumulation and its returns sustain persistent wage disparities.

Taken together, these findings provide compelling evidence of a multi-layered segmented labour market in China. Choice of the definition of urban and rural, allows us to demonstrate that the spatial segmentation, reflected in the urban–rural divide, coexists with institutional

segmentation embedded in the hukou system. Spatial segmentation produces disparities in education, employment structure, and wage levels, while institutional segmentation entrenches these inequalities through legal and social barriers that limit the mobility and economic advancement of rural hukou holders. The persistence of lower returns to education for rural hukou workers, even in urban settings, suggests that discrimination and institutional exclusion, rather than purely human capital differences, are an important mechanism reproducing inequality.

China's rural and urban economies diverge sharply in employment structure, labour market dynamics, and educational quality. Rural employment is concentrated in agriculture and informal sectors characterised by low productivity, unstable incomes, and limited non-farm opportunities, while urban economies are service- and technology-oriented, offering higher wages and more formal contracts (Knight & Song 2003; Meng 2012). Rural labour markets face surplus labour and underemployment in low-return agriculture (Golley & Meng 2011), whereas urban markets demand more skilled workers (Rozelle et al. 2020). Migration is driven by rural underemployment but constrained by hukou restrictions and migrants' limited skills (Démurger et al. 2009), reinforcing structural segmentation. Educational disparities intensify these divides: rural schools lag in facilities, teacher quality, attainment, and access to higher education (Xiang & Stillwell 2023), limiting rural youth's competitiveness (Loyalka et al. 2013) and depressing human capital returns (Gao & Li 2022).

Policy implications arising from these results are substantial. First, hukou reform remains essential for mitigating institutional barriers and ensuring equal access to social services and formal employment opportunities. Second, policies that enhance the quality of education and improve the returns to human capital in rural areas are critical for reducing the structural wage gap.[16] Finally, integrating rural hukou migrants into urban social insurance systems and formal labour markets would help weaken the urban–rural divide and promote inclusive growth. In sum, the analysis demonstrates that both spatial location and institutional status jointly determine labour market outcomes, perpetuating a system of stratified opportunities that reproduces inequality across generations.

[16] The lower returns to education in rural areas can influence the decision to invest in human capital in rural areas, thus perpetuating the existing gap in educational attainment. The lower returns to education are also linked with the quality of education in rural areas.

References

Alam, M. A. U., & He, X. (2022). Is the discrimination against migrant workers tending toward zero in urban China?. *International Studies of Economics*, 17(1), 65-81.

Asadullah, M. N. & Xiao, S. (2020). The changing pattern of wage returns to education in post-reform China. *Structural Change and Economic Dynamics*, *53*, 137-148.

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. New York: National Bureau of Economic Research.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources,* 8, 436–455.

Bryan, M., Bryce, A., Rice, N., Roberts, J., & Sechel, C. (2022). Exploring mental health disability gaps in the labour market: the UK experience during COVID-19. *Labour Economics*, 78, 102253.

Campos, B. C., Ren, Y., & Petrick, M. (2016). The impact of education on income inequality between ethnic minorities and Han in China. *China Economic Review*, *41*, 253-267.

Choi, S. (2021). Urban/rural disparities in the wage effect of additional vocational education after formal education: the case of the Philippines. *International Journal of Training Research*, 19(3), 229-241.

Démurger, S., Gurgand, M., Li, S., & Yue, X. (2009). Migrants as second-class workers in urban China? A decomposition analysis. Journal of Comparative Economics, 37(4), 610-628.

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica, 77*(3), 953–973.

Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition methods in economics. In O. Ashenfelter, & D. Card (Eds.), Handbook of labor economics (Vol. IV.A, pp. 1 – 102). North - Holland.

Fu, Q., & Ren, Q. (2010). Educational inequality under China's rural–urban divide: The hukou system and return to education. *Environment and Planning A*, 42(3), 592-610.

Gao, X., & Li, M. (2022). Differences between returns to education in Urban and rural China and its evolution from 1989 to 2019. *Plos one*, 17(10), e0274506.

Golley, J., & Meng, X. (2011). Has China run out of surplus labour? *China Economic Review*, 22(4), 555-572.

Guo, L., Huang, J., & Zhang, Y. (2019). Education development in China: Education return, quality, and equity. *Sustainability*, 11(13), 3750.

Kanjilal-Bhaduri, S., & Pastore, F. (2018). Returns to education and female participation nexus: Evidence from India. *The Indian Journal of Labour Economics*, 61(3), 515-536.

Knight, J., & Song, L. (2003). Increasing urban wage inequality in China: Extent, elements and evaluation. *Economics of transition*, 11(4), 597-619.

Li, H., Meng, L., Shi, X., & Wu, B. (2012). Does attending elite colleges pay in China? *Journal of*

Comparative Economics, 40(1), 78-88.

Liu, X. (2020). Structural changes and economic growth in China over the past 40 years of reform and opening-up. *China Political Economy*, 3(1), 19-38.

Loyalka, P., Song, Y., Wei, J., Zhong, W., & Rozelle, S. (2013). Information, college decisions and financial aid: Evidence from a cluster-randomized controlled trial in China. *Economics of Education Review*, 36, 26-40.

Lu, Y. (2019). Forty years of reform and opening up: China's progress and challenges. *China Economic Journal*, 12(2), 97-120.

Meng, X. (2012). Labor market outcomes and reforms in China. *Journal of Economic Perspectives*, 26(4), 75-102.

Messinis, G. (2013). Returns to education and urban-migrant wage differentials in China: IV quantile treatment effects. *China Economic Review*, 26, 39-55.

Mincer, J. (1974). *Schooling, Experience and Earnings*. New York: National Bureau of Economic Research.

Michaelsen, M. M., & Haisken-DeNew, J. P. (2015). Migration magnet: the role of work experience in rural–urban wage differentials. *IZA Journal of Migration*, 4, 1-25.

Nguyen, H., Parker, B. R., & Simpson, S. S. (2021). Equal pay for equal work? Considering the gender gap in illegal pay. *Journal of Quantitative Criminology*, 1-34.

Oaxaca, R. (1973). Male–female wage differentials in urban labor markets. *International Economic Review,* 14, 693–709.

Peet, E. D., Fink, G., & Fawzi, W. (2015). Returns to education in developing countries: Evidence from the living standards and measurement study surveys. *Economics of education review*, 49, 69-90.

Pereira, J., & Galego, A. (2011). Regional wage differentials in Portugal: Static and dynamic approaches. *Papers in Regional Science*, 90(3), 529-549.

Rani, G. P. (2014). Disparities in earnings and education in India. *Cogent Economics & Finance*, 2(1), 1-18.

Rozelle, S., Xia, Y., Friesen, D., Vanderjack, B., & Cohen, N. (2020). Moving beyond Lewis: Employment and wage trends in China's high-and low-skilled industries and the emergence of an era of polarization: Presidential address for the 2020 Association for comparative economic studies meetings. *Comparative Economic Studies*, 62(4), 555.

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.

Song, Y. (2014). What should economists know about the current Chinese hukou system?. *China Economic Review*, 29, 200-212.

Warunsiri, S., & McNown, R. (2010). The returns to education in Thailand: A pseudo-panel approach. *World Development,* 38(11), 1616-1625.

Weng, Y., Li, M., & Foo, C. T. (2016). Education inside China: a longitudinal (1989-2011) analysis. *Chinese Management Studies*, 10(2), 217-241.

Xiang, L., & Stillwell, J. (2023). Rural–urban educational inequalities and their spatial variations in China. *Applied Spatial Analysis and Policy*, 16(2), 873-896.

Zhang, H. (2017). Opportunity or new poverty trap: Rural-urban education disparity and internal migration in China. *China Economic Review*, 44, 112-124.

Zhu, R. (2015). Heterogeneity in the economic returns to schooling among Chinese rural–urban migrants, 2002–07. *Economics of Transition*, 23(1), 135-167.

Table 1A: Overlap between registration and residence status

	Male		Female	
	Urban residence	**Rural residence**	**Urban residence**	**Rural residence**
Urban hukou	6543 (52%)	876 (12%)	5209 (54%)	583 (14%)
Rural hukou	6070 (48%)	6521 (88%)	4447 (46%)	3698 (86%)
N	12,613 (63%)	7,397 (37%)	9,656 (69%)	4,281 (31%)
Total	20,010		13,937	

Table 1B: Mean log hourly earnings by place of residence and hukou status for males and females

	Male		Female	
	Urban residence	**Rural residence**	**Urban residence**	**Rural residence**
Urban hukou	3.034	2.933	2.859	2.731
Rural hukou	2.894	2.756	2.602	2.456

Table 2: Summary statistics by geographical region and gender

	Male				Female			
	Urban		**Rural**		**Urban**		**Rural**	
	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
Personal Characteristics								
Education years	11.455	3.867	9.197	3.874	11.837	4.109	9.532	4.525
Age	38.758	10.869	37.797	11.251	36.413	9.315	34.365	9.698
Minority	0.041	0.199	0.084	0.277	0.057	0.232	0.079	0.269
Urban hukou	0.519	0.500	0.118	0.323	0.539	0.498	0.136	0.343
Northeast	0.158	0.365	0.099	0.298	0.152	0.359	0.096	0.294
East	0.427	0.495	0.331	0.471	0.441	0.497	0.374	0.484
Middle	0.214	0.410	0.227	0.419	0.217	0.412	0.221	0.415
West	0.200	0.400	0.343	0.475	0.191	0.393	0.310	0.463
Work-related Characteristics								
Formal contract	0.627	0.484	0.429	0.495	0.627	0.484	0.466	0.499
Public sector	0.353	0.478	0.198	0.399	0.330	0.470	0.195	0.396
Household Characteristics								
Married	0.775	0.417	0.754	0.431	0.785	0.411	0.760	0.427
Children (0 - 5)	0.321	0.584	0.415	0.683	0.290	0.551	0.385	0.651
Children (6 - 15)	0.402	0.675	0.551	0.817	0.453	0.698	0.641	0.913
Year								
2014	0.175	0.380	0.172	0.378	0.179	0.383	0.176	0.381
2016	0.222	0.415	0.215	0.411	0.208	0.406	0.215	0.411
2018	0.222	0.416	0.218	0.413	0.220	0.414	0.211	0.408
2020	0.182	0.386	0.194	0.395	0.186	0.389	0.192	0.394
2022	0.199	0.400	0.201	0.400	0.207	0.405	0.206	0.405

Note: Urban/rural status is defined based on place of residence.
Education years are transformed from individuals' finished education levels.
Sd is standard deviation.

Table 3: OLS regression by gender and region; dependent variable: (log) gross hourly wage

	Male		Female	
	Urban	**Rural**	**Urban**	**Rural**
Education years	0.050***	0.025***	0.055***	0.026***
	(0.002)	(0.003)	(0.003)	(0.003)
Age	0.033***	0.032***	0.026***	0.023**
	(0.006)	(0.006)	(0.008)	(0.010)
Age squared	-0.000***	-0.001***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Minority	-0.026	-0.028	0.059*	0.099**
	(0.037)	(0.029)	(0.035)	(0.042)
Urban hukou	0.043***	0.087***	0.091***	0.133***
	(0.017)	(0.028)	(0.018)	(0.031)
Northeast	-0.022	-0.053	-0.088***	-0.075**
	(0.025)	(0.032)	(0.028)	(0.037)
East	0.241***	0.123***	0.246***	0.064**
	(0.020)	(0.020)	(0.022)	(0.026)
Middle	0.050**	0.078***	-0.028	0.015
	(0.023)	(0.023)	(0.025)	(0.032)
Formal contract	0.129***	0.147***	0.220***	0.274***
	(0.015)	(0.016)	(0.017)	(0.021)
Public sector	-0.041**	-0.021	0.080***	0.057*
	(0.017)	(0.023)	(0.019)	(0.030)
Married	0.050**	0.072***	-0.025	-0.099***
	(0.022)	(0.023)	(0.023)	(0.032)
Children (0 - 5)	-0.011	-0.005	-0.008	0.012
	(0.016)	(0.017)	(0.018)	(0.024)
Children (6 - 15)	-0.031**	-0.041**	-0.073***	-0.068***
	(0.015)	(0.017)	(0.017)	(0.022)
Year_2016	0.006	-0.108***	0.051***	-0.044
	(0.016)	(0.022)	(0.018)	(0.027)
Year_2018	0.132***	0.072***	0.168***	0.061**
	(0.017)	(0.021)	(0.018)	(0.029)
Year_2020	0.230***	0.139***	0.239***	0.121***
	(0.018)	(0.022)	(0.020)	(0.031)
Year_2022	0.325***	0.271***	0.309***	0.181***
	(0.018)	(0.023)	(0.020)	(0.029)
Constant	1.518***	1.903***	1.270***	1.798***
	(0.105)	(0.108)	(0.137)	(0.165)
N	12613	7397	9656	4281
adj. R^2	0.206	0.152	0.293	0.196

Note: Standard errors clustered at the individual level in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Reference categories: ethnicity: majority (Han); hukou: rural hukou; province: West; contract: no formal contract; sector: private sector; marriage: no living spouse; children: no children under 15 years old in household; year: 2014.

Table 4: Oaxaca-Blinder decomposition of urban-rural difference; dependent variable: (log) gross hourly wage

Controls	Male		Female	
Overall decomposition	Total		Total	
Urban	2.967		2.741	
Rural	2.777		2.493	
Difference	0.190***		0.247***	
Explained	0.174*** (91.6%)		0.226*** (91.5%)	
Unexplained	0.016 (8.45)		0.021 (8.5%)	
Detailed decomposition	Explained	Unexplained	Explained	Unexplained
Personal characteristics				
Education	0.113***	0.226***	0.127***	0.277***
Age	0.001	0.104	0.000	0.104
Ethnicity	0.001	0.000	-0.001	-0.003
Urban hukou	0.017***	-0.005	0.037***	-0.006
Province	0.021***	0.036**	0.012***	0.057***
Labour market factors				
Formal contract	0.025***	-0.008	0.035***	-0.025**
Public sector	-0.006**	-0.004	0.011**	0.005
Household Characteristics				
Married	0.001*	-0.016	-0.001	0.056*
Children	0.003*	0.002	0.006***	-0.008
Period controls				
Year	-0.002	0.067***	0.000	0.092***
constant		-0.385***		-0.527**

Note: Urban/rural status is defined based on place of residence.
Decomposition uses the urban group as the reference category.

Table 5.1: RIf decomposition of urban-rural difference at different quantiles; dependent variable: (log) gross hourly wage: MALE

Controls	10^{th}		50^{th}		90^{th}	
Overall decomposition	Tot		Tot		Tot	
Urban	2.129		2.933		3.899	
Rural	2.043		2.759		3.551	
Difference	0.086***		0.175***		0.349***	
Explained	0.115*** (133.7%)		0.177*** (101.1%)		0.225*** (64.5%)	
Unexplained	-0.029 (-33.7%)		-0.002 (-1.1%)		0.124*** (35.5%)	
Detailed decomposition	Exp	Unexp	Exp	Unexp	Exp	Unexp
Personal characteristics						
Education	0.073***	0.108*	0.115***	0.227***	0.155***	0.268***
Age	-0.002	0.153	0.001	0.073	0.007***	0.667***
Ethnicity	0.003	-0.007	0.000	0.003	-0.001	0.005
Urban hukou	-0.009	-0.015***	0.011	-0.005	0.047***	0.003
Province	0.020***	0.023	0.022***	0.037*	0.027***	0.080**
Labour market factors						
Formal contract	0.030***	0.017	0.025***	-0.015	0.010	-0.070***
Public sector	-0.001	0.029***	0.000	-0.003	-0.021***	-0.035***
Household Characteristics						
Married	0.002**	-0.084*	0.001*	-0.001	-0.001	-0.040
Children	0.001	0.022	0.004**	-0.018	0.002	0.031
Period controls						
Year	-0.003	0.083**	-0.002	0.061***	-0.002	0.069*
constant		-0.358		-0.361**		-0.853***

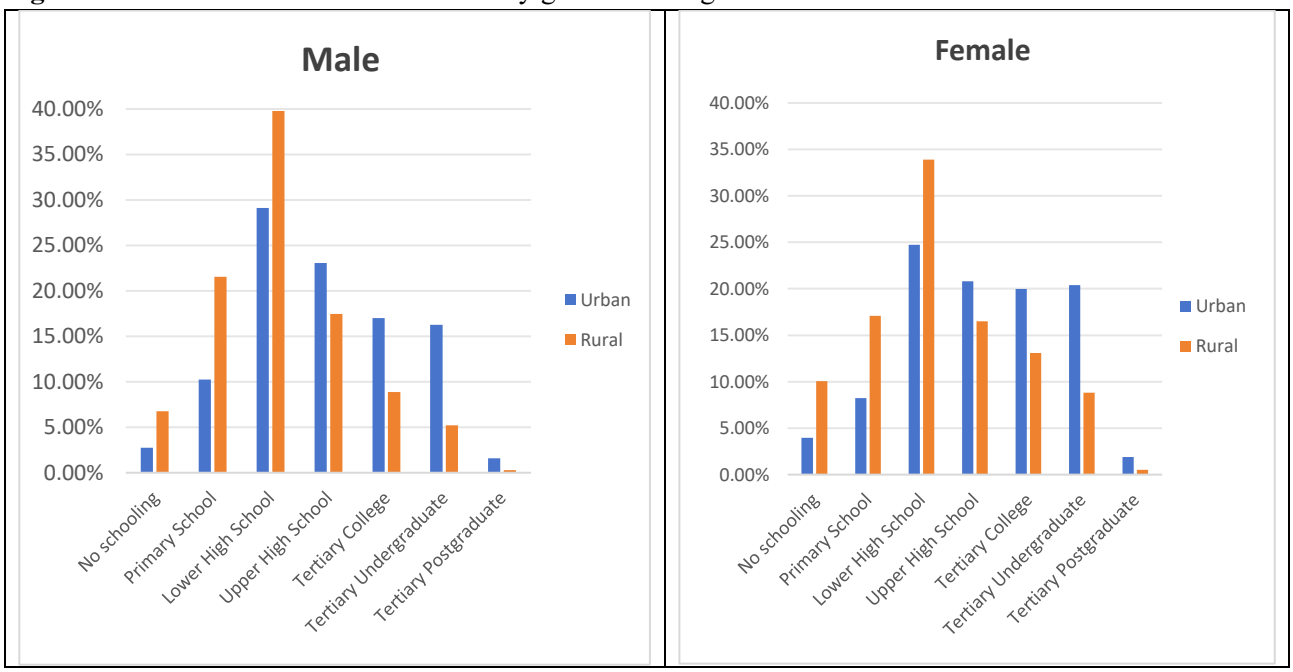
Note: Urban/rural status is defined based on place of residence.
Decomposition uses the urban group as the reference category.

Table 5.2: Rif decomposition of urban-rural difference at different quantiles; dependent variable: (log) gross hourly wage: FEMALE

Controls	10^{th}		50^{th}		90th	
Overall decomposition	Tot		Tot		Tot	
Urban	1.865		2.693		3.707	
Rural	1.726		2.440		3.311	
Difference	0.140***		0.253***		0.396***	
Explained	0.152*** (108.6%)		0.243*** (96.0%)		0.288*** (72.7%)	
Unexplained	-0.012 (-8.6%)		0.010 (4.0%)		0.108*** (27.3%)	
Detailed decomposition	Exp	Unexp	Exp	Unexp	Exp	Unexp
Personal characteristics						
Education	0.083***	0.287***	0.150***	0.344***	0.155***	0.173*
Age	-0.015	-0.416***	-0.002	0.192	0.015**	0.047
Ethnicity	0.001	-0.001	-0.001	-0.003	-0.001	-0.013
Urban hukou	0.024**	-0.015**	0.025**	-0.011*	0.062***	-0.007
Region	0.004	0.049	0.013**	0.043	0.021**	0.123***
Labour market factors						
Formal contract	0.048***	0.002	0.038***	-0.030*	0.014**	-0.093***
Public sector	0.002	0.002	0.017***	0.013*	0.014**	0.012
Household Characteristics						
Married	0.000	0.044	-0.001	0.038	0.000	0.091
Children	0.005	0.015	0.004	0.016	0.009*	-0.019
Period controls						
Year	0.000	0.117**	0.000	0.093	0.000	0.107**
constant		-0.097		-0.686***		-0.314

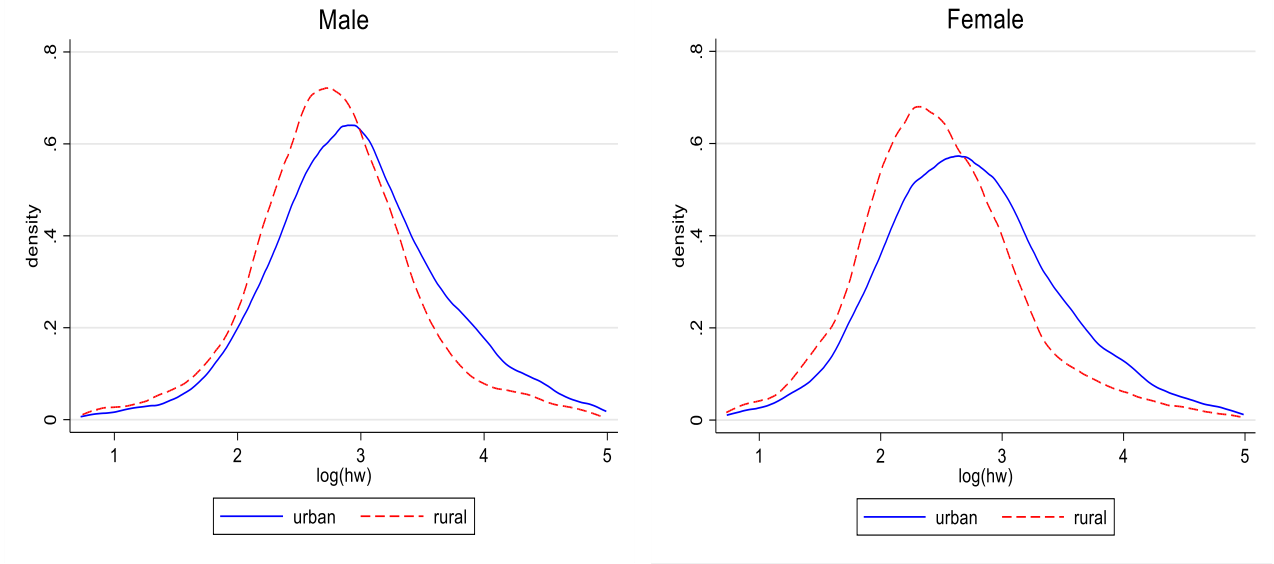
Note: Urban/rural status is defined based on place of residence.
Decomposition uses the urban group as the reference category.

Figure 1: Distribution of education levels by gender and region



Note: Urban/rural status is defined based on place of residence. The levels of education are for the pooled sample covering all five waves of CFPS from 2014 to 2022.

Figure 2: Distribution of log hourly wage by gender and region



Note: Urban/rural status is defined based on place of residence.

Appendix

Table A1: Education years for each finished education level

Education level	Education years
No Schooling	0
Primary School	6
Lower High School	9
Upper High School	12
Tertiary College	15
Tertiary Undergraduate	16
Tertiary Postgraduate (master's degree)	19
Tertiary Postgraduate (PhD)	22

Source: CFPS users' manual

Table A2: Mean log hourly earnings by education levels and gender in urban and rural areas

	Male			Female		
	Urban	Rural	Difference in Mean	Urban	Rural	Difference in Mean
All	2.967	2.777	0.190***	2.741	2.493	0.248***
No schooling	2.611	2.506	0.105**	2.234	2.208	0.026
Primary School	2.662	2.612	0.050**	2.322	2.245	0.077***
Lower High School	2.766	2.764	0.002	2.419	2.411	0.008
Upper High School	2.889	2.827	0.062***	2.636	2.527	0.109***
Tertiary College	3.146	3.034	0.112***	2.944	2.784	0.160***
Tertiary Undergraduate	3.426	3.243	0.183***	3.230	3.060	0.170***
Tertiary Postgraduate	3.729	3.785	-0.056	3.573	3.543	0.030

Notes: Urban/rural status is defined based on place of residence.
Tertiary Postgraduate level includes both master's and doctoral degrees.
Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Occupation and industry distribution by region

	Urban areas		Rural areas	
	N	**%**	**N**	**%**
Occupations				
1. Heads of government, Party, and enterprises	1,496	6.26	486	4.18
2. Professional and technical personnel	4,490	18.80	1,400	12.04
3. Clerical and related personnel	3,190	13.37	882	7.59
4. Commercial and service workers	5,441	22.81	2,481	21.35
5. Agricultural and related workers	232	0.97	250	2.15
6. Production and transportation equipment operators	7,015	29.40	5,881	50.56
7. Others	326	1.36	262	2.26
Missing	79	0.33	36	0.31
Total	**22,269**	**100.00**	**11,678**	**100.00**
Industries				
1. Agriculture, forestry, animal husbandry, fishery	191	0.86	257	2.20
2. Mining	357	1.61	319	2.73
3. Manufacturing	6,106	27.42	3,641	31.17
4. Electricity, gas, and water supply	394	1.77	230	1.97
5. Construction	1,814	8.14	2,154	18.44
6. Transport, storage, and postal services	1,319	5.93	537	4.60
7. Information and computer services	547	2.45	149	1.28
8. Wholesale and retail trade	2,419	10.86	945	8.09
9. Accommodation and catering	950	4.26	458	3.92
10. Financial industry	713	3.20	128	1.10
11. Real estate	632	2.84	187	1.60
12. Leasing and business services	753	3.38	234	2.00
13. Scientific research and technical services	210	0.94	57	0.49
14. Water conservancy and public facilities	372	1.67	99	0.85
15. Resident and other services	736	3.31	393	3.37
16. Education	1,717	7.70	684	5.86
17. Health and social welfare	870	3.90	313	2.68
18. Culture, sports, and entertainment	309	1.39	132	1.13
19. Public administration and social organizations	1,649	7.40	489	4.19
20. Others	13	0.06	5	0.04
Missing	141	0.63	82	0.70
Total	**22,269**	**100.00**	**11,678**	**100.00**

Table A4: Summary statistics by registration status and gender

	Male				Female			
	Urban		**Rural**		**Urban**		**Rural**	
	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
Personal Characteristics								
Education years	12.614	3.502	9.446	3.838	12.941	3.600	9.841	4.416
Age	40.600	10.538	37.108	11.095	37.557	8.997	34.523	9.614
Minority	0.042	0.201	0.066	0.248	0.054	0.225	0.071	0.256
Urban residence	0.882	0.323	0.482	0.500	0.899	0.301	0.546	0.498
Northeast	0.205	0.404	0.096	0.294	0.186	0.390	0.097	0.296
East	0.381	0.486	0.398	0.489	0.410	0.492	0.428	0.495
Middle	0.219	0.414	0.219	0.414	0.225	0.418	0.213	0.409
West	0.195	0.396	0.288	0.453	0.179	0.383	0.262	0.440
Work-related Characteristics								
Formal contract	0.702	0.457	0.466	0.499	0.696	0.460	0.493	0.500
Public sector	0.497	0.500	0.177	0.382	0.448	0.497	0.175	0.380
Household Characteristics								
Married	0.809	0.393	0.743	0.437	0.797	0.403	0.764	0.425
Children (0 - 5)	0.246	0.431	0.312	0.463	0.245	0.430	0.282	0.450
Children (6 - 15)	0.308	0.462	0.369	0.483	0.340	0.474	0.416	0.493
Year								
2014	0.212	0.409	0.151	0.358	0.202	0.402	0.161	0.368
2016	0.224	0.417	0.217	0.412	0.209	0.406	0.211	0.408
2018	0.210	0.407	0.227	0.419	0.211	0.408	0.221	0.415
2020	0.167	0.373	0.197	0.398	0.184	0.388	0.190	0.392
2022	0.187	0.390	0.208	0.406	0.194	0.396	0.216	0.412

Notes: Urban/rural status is defined based on registration status.
Education years are transformed from individuals' finished education levels.
Sd is standard deviation.

Table A5: OLS regression by gender and registration status; dependent variable: (log) gross hourly wage

	Male		Female	
	Urban	**Rural**	**Urban**	**Rural**
Education years	0.071***	0.025***	0.074***	0.030***
	(0.003)	(0.002)	(0.004)	(0.002)
Age	0.008	0.043***	0.027**	0.028***
	(0.008)	(0.005)	(0.011)	(0.008)
Age squared	-0.000	-0.001***	-0.000**	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Minority	0.007	-0.047*	0.043	0.079**
	(0.047)	(0.027)	(0.048)	(0.032)
Urban residence	0.009	0.051***	0.008	0.062***
	(0.027)	(0.013)	(0.029)	(0.016)
Northeast	-0.018	-0.013	-0.068**	-0.069**
	(0.031)	(0.026)	(0.034)	(0.030)
East	0.324***	0.143***	0.355***	0.104***
	(0.029)	(0.016)	(0.029)	(0.021)
Middle	0.038	0.078***	0.004	-0.020
	(0.031)	(0.018)	(0.033)	(0.024)
Formal contract	0.131***	0.150***	0.194***	0.266***
	(0.020)	(0.013)	(0.022)	(0.016)
Public sector	-0.023	-0.060***	0.103***	0.027
	(0.021)	(0.018)	(0.023)	(0.023)
Married	0.090***	0.038**	0.016	-0.094***
	(0.031)	(0.018)	(0.029)	(0.025)
Children (0 - 5)	-0.041*	0.006	-0.021	0.011
	(0.022)	(0.014)	(0.024)	(0.018)
Children (6 - 15)	-0.037*	-0.037***	-0.061***	-0.084***
	(0.021)	(0.013)	(0.022)	(0.017)
Year_2016	-0.005	-0.065***	0.039*	0.004
	(0.020)	(0.017)	(0.023)	(0.020)
Year_2018	0.147***	0.083***	0.173***	0.110***
	(0.020)	(0.017)	(0.022)	(0.021)
Year_2020	0.257***	0.159***	0.278***	0.154***
	(0.023)	(0.018)	(0.024)	(0.023)
Year_2022	0.314***	0.295***	0.324***	0.243***
	(0.022)	(0.018)	(0.024)	(0.022)
Constant	1.627***	1.692***	0.909***	1.679***
	(0.152)	(0.087)	(0.184)	(0.130)
N	7419	12591	5792	8145
adj. R^2	0.247	0.164	0.310	0.222

Standard errors clustered at the individual level in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Reference Categories: ethnicity: majority (Han); residence: urban residence; province: West; contract: no formal contract; sector: private sector; marriage: no living spouse; children: no children under 15 years old in household; year: 2014.

Table A6: Oaxaca-Blinder decomposition of urban-rural difference; dependent variable: (log) gross hourly wage.

Controls	Male		Female	
Overall decomposition	Total		Total	
Urban	3.022		2.846	
Rural	2.823		2.536	
Difference	0.199***		0.310***	
Explained	0.232*** (116.6%)		0.294*** (94.8%)	
Unexplained	-0.033* (-16.6%)		0.016 (5.2%)	
Detailed decomposition	Exp	Unexp	Exp	Unexp
Personal characteristics				
Education	0.224***	0.432***	0.230***	0.438***
Age	-0.005	-0.519***	0.011**	0.142
Ethnicity	0.000	0.004	-0.001	-0.003
Urban residence	0.004	-0.020	0.003	-0.030
Province	-0.007	0.063***	-0.012**	0.113***
Labour market factors				
Formal contract	0.031***	-0.009	0.039***	-0.035***
Public sector	-0.007	0.007	0.028***	0.013**
Household Characteristics				
Married	0.006***	0.039	0.001	0.084***
Children	0.005**	-0.015	0.006**	0.001
Period controls				
Year 2016	-0.017***	0.050***	-0.011***	0.063***
constant		-0.065		-0.769***

Note: Urban/rural status is defined based on registration status.
Decomposition uses the urban group as the reference category.

Table A7.1: Rif decomposition of urban-rural difference at different quantiles; dependent variable: (log) gross hourly wage. MALE.

Controls	10th		50th		90th	
Overall decomposition	Tot		Tot		Tot	
Urban	2.144		2.995		3.966	
Rural	2.069		2.803		3.637	
Difference	0.075***		0.193***		0.328***	
Explained	0.154*** (205.3%)		0.239*** (123.8%)		0.257*** (78.4%)	
Unexplained	-0.078** (-105.3%)		-0.046** (-23.8%)		0.071** (21.6%)	
Detailed decomposition	Exp	Unexp	Exp	Unexp	Exp	Unexp
Personal characteristics						
Education	0.173***	0.384***	0.223***	0.434***	0.261***	0.332***
Age	-0.02**	-0.423	-0.004	-0.6***	0.016*	-0.113
Ethnicity	-0.001	0.007	-0.001	0.006	0.001	-0.002
Urban residence	-0.037**	-0.049**	-0.009	-0.030	0.028	-0.023
Province	-0.001	0.028	-0.006	0.03	-0.009	0.145***
Labour market factors						
Formal contract	0.032***	0.004	0.036***	-0.006	0.020***	-0.033
Public sector	0.008	0.030***	0.007	0.008	-0.043***	-0.016
Household Characteristics						
Married	0.012***	0.046	0.005**	0.031	-0.004	-0.028
Children	0.007*	-0.028	0.004	-0.011	-0.001	0.038
Period controls						
Year	-0.019***	0.069	-0.017***	0.037**	-0.013***	0.061**
constant		-0.146		0.056		-0.291

Note: Urban/rural status is defined based on registration status.

Decomposition uses the urban group as the reference category.

Table A7.2: Rif decomposition of urban-rural difference at different quantiles; dependent variable: (log) gross hourly wage. FEMALE.

Controls	10th		50th		90th	
Overall decomposition	Tot		Tot		Tot	
Urban	1.954		2.799		3.834	
Rural	1.749		2.481		3.393	
Difference	0.205***		0.318***		0.442***	
Explained	0.178*** (86.8%)		0.344*** (108.2%)		0.302*** (68.3%)	
Unexplained	0.027 (13.2%)		-0.026 (-8.2%)		0.140*** (31.7%)	
Detailed decomposition	Exp	Unexp	Exp	Unexp	Exp	Unexp
Personal characteristics						
Education	0.176***	0.440***	0.274***	0.556***	0.235***	0.217**
Age	-0.012	-0.325	0.011*	0.095	0.027**	-0.024
Ethnicity	0.000	-0.002	-0.002	0.003	0.000	-0.010
Urban residence	-0.041**	-0.081***	0.004	-0.034	0.014	-0.024
Province	-0.02**	0.07*	-0.011*	0.095***	-0.008	0.207***
Labour market factors						
Formal contract	0.065***	0.022	0.034***	-0.072***	0.020**	-0.054**
Public sector	0.018**	0.020***	0.040***	0.020***	0.018	-0.002
Household Characteristics						
Married	0.001	0.025	0.001	0.094***	0.001	0.152**
Children	0.005*	0.006	0.005	0.005	0.006	-0.003
Period controls						
Year	-0.014***	0.145***	-0.011***	0.038	-0.01***	0.094**
constant		-0.292		-0.826***		-0.413

Note: Urban/rural status is defined based on registration status.
Decomposition uses the urban group as the reference category.

Table A8: Summary statistics by gender and region defined using both residence and registration status.

	Male				Female			
	Urban		**Rural**		**Urban**		**Rural**	
	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
Personal Characteristics								
Education years	12.746	3.424	8.870	3.752	13.060	3.505	9.164	4.462
Age	40.755	10.469	37.576	11.270	37.928	8.928	34.384	9.812
Minority	0.039	0.194	0.086	0.280	0.052	0.221	0.080	0.271
Urban hukou	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
Northeast	0.214	0.410	0.094	0.291	0.194	0.395	0.092	0.289
East	0.384	0.486	0.327	0.469	0.410	0.492	0.369	0.482
Middle	0.223	0.416	0.232	0.422	0.232	0.422	0.230	0.421
West	0.179	0.383	0.347	0.476	0.164	0.370	0.310	0.463
Work-related Characteristics								
Formal contract	0.713	0.452	0.403	0.490	0.702	0.457	0.438	0.496
Public sector	0.505	0.500	0.167	0.373	0.456	0.498	0.165	0.372
Household Characteristics								
Married	0.810	0.393	0.747	0.435	0.798	0.401	0.756	0.430
Children (0 - 5)	0.235	0.424	0.321	0.467	0.231	0.422	0.298	0.457
Children (6 - 15)	0.297	0.457	0.396	0.489	0.338	0.473	0.452	0.498
Year								
2014	0.214	0.410	0.169	0.375	0.203	0.402	0.173	0.378
2016	0.225	0.417	0.215	0.411	0.206	0.404	0.212	0.409
2018	0.214	0.410	0.223	0.416	0.213	0.409	0.213	0.410
2020	0.166	0.372	0.196	0.397	0.184	0.388	0.193	0.395
2022	0.182	0.386	0.198	0.398	0.195	0.396	0.209	0.407

Notes: Education years are transformed from individuals' finished education levels.
Sd is standard deviation.

Table A9: OLS regressions by gender and region, with region defined based on both place of residence and registration status; dependent variable: (log) gross hourly wage

	Male		Female	
	Urban	**Rural**	**Urban**	**Rural**
Education years	0.073***	0.021***	0.077***	0.022***
	(0.004)	(0.003)	(0.004)	(0.003)
Age	0.011	0.036***	0.029**	0.019*
	(0.009)	(0.006)	(0.012)	(0.011)
Age squared	-0.000	-0.001***	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
Minority	-0.007	-0.035	0.054	0.118***
	(0.052)	(0.030)	(0.051)	(0.044)
Northeast	0.005	-0.026	-0.045	-0.051
	(0.034)	(0.034)	(0.036)	(0.040)
East	0.350***	0.120***	0.388***	0.061**
	(0.032)	(0.021)	(0.032)	(0.028)
Middle	0.050	0.092***	0.019	0.020
	(0.035)	(0.024)	(0.036)	(0.033)
Formal contract	0.116***	0.134***	0.196***	0.286***
	(0.022)	(0.017)	(0.024)	(0.022)
Public sector	-0.017	-0.023	0.106***	0.041
	(0.022)	(0.025)	(0.024)	(0.034)
Married	0.082**	0.058**	0.010	-0.117***
	(0.034)	(0.024)	(0.031)	(0.037)
Children (0 - 5)	-0.033	0.002	-0.015	0.017
	(0.024)	(0.018)	(0.026)	(0.026)
Children (6 - 15)	-0.028	-0.034*	-0.054**	-0.056**
	(0.023)	(0.018)	(0.023)	(0.024)
Year_2016	0.003	-0.116***	0.042*	-0.047
	(0.021)	(0.023)	(0.024)	(0.030)
Year_2018	0.139***	0.058**	0.173***	0.056*
	(0.022)	(0.023)	(0.024)	(0.032)
Year_2020	0.255***	0.125***	0.280***	0.120***
	(0.024)	(0.024)	(0.026)	(0.034)
Year_2022	0.303***	0.265***	0.332***	0.195***
	(0.024)	(0.025)	(0.026)	(0.032)
Constant	1.544***	1.893***	0.809***	1.903***
	(0.174)	(0.115)	(0.207)	(0.182)
N	6543	6521	5209	3698
adj. R^2	0.244	0.136	0.319	0.177

Standard errors in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Reference Categories: ethnicity: majority (Han); province: West; contract: no formal contract; sector: private sector; marriage: no living spouse; children: no children under 15 years old in household; year: 2014.

Table A10: Oaxaca decomposition with urban-rural difference; dependent variable: (log) gross hourly wage.

Controls	Male		Female	
Overall decomposition	Total		Total	
Urban	3.034		2.859	
Rural	2.756		2.456	
Difference	0.278***		0.403***	
Explained	0.327*** (117.6%)		0.408*** (101.2%)	
Unexplained	-0.050** (-17.6%)		-0.005 (-1.2%)	
Detailed decomposition	Exp	Unexp	Exp	Unexp
Personal characteristics				
Education	0.283***	0.459***	0.299***	0.497***
Age	-0.003	-0.309	0.017**	0.316
Ethnicity	0.000	0.002	-0.002	-0.005
Province	0.020	0.068***	0.012	0.121***
Labour market factors				
Formal contract	0.036***	-0.007	0.052***	-0.040***
Public sector	-0.006	0.001	0.031***	0.011
Household Characteristics				
Married	0.005**	0.017	0.000	0.096***
Children	0.006*	-0.009	0.007**	-0.008
Period controls				
Year	-0.014***	0.076***	-0.008***	0.103***
constant		-0.349*		-1.095

Note: Urban/rural status is defined based on both place of residence and registration status.
Decomposition uses the urban group as the reference category.

Table A11.1: Rif decomposition with urban-rural difference at different quantiles; dependent variable: (log) gross hourly wage. MALE.

Controls	10^{th}		50^{th}		90^{th}	
Overall decomposition	Tot		Tot		Tot	
Urban	2.143		3.001		3.990	
Rural	2.021		2.734		3.521	
Difference	0.121***		0.267***		0.469***	
Explained	0.275*** (227.3%)		0.347*** (130.0%)		0.345*** (73.6%)	
Unexplained	-0.154*** (-127.3%)		-0.080*** (-30.0%)		0.124*** (26.4%)	
Detailed decomposition	Exp	Unexp	Exp	Unexp	Exp	Unexp
Personal characteristics						
Education	0.212***	0.334***	0.292***	0.469***	0.332***	0.458***
Age	-0.021**	-0.332	-0.003	-0.337*	0.021**	0.332
Ethnicity	0.000	-0.002	-0.002	0.006	0.002	-0.002
Province	0.025**	0.046	0.017***	0.036	0.027***	0.142***
Labour market factors						
Formal contract	0.035***	0.008	0.041***	-0.010	-0.042*	-0.058***
Public sector	0.017	0.037***	0.006	0.003	-0.042***	-0.029***
Household Characteristics						
Married	0.012***	-0.012	0.005**	0.017	-0.005	-0.046
Children	0.01	-0.009	0.003	-0.016	-0.003	0.048
Period controls						
Year	-0.015***	0.076*	-0.015***	0.063***	-0.01***	0.076*
constant		-0.300		-0.310		-0.798**

Note: Urban/rural status is defined based on both place of residence and registration status.
Decomposition uses the urban group as the reference category.

Table A11.2: Rif decomposition with urban-rural difference at different quantiles; dependent variable: (log) gross hourly wage. FEMALE.

Controls	10^{th}		50^{th}		90^{th}	
Overall decomposition	Tot		Tot		Tot	
Urban	1.947		2.818		3.853	
Rural	1.683		2.404		3.253	
Difference	0.264***		0.415***		0.600***	
Explained	0.324*** (122.7%)		0.470*** (113.3%)		0.400*** (66.7%)	
Unexplained	-0.060 (-22.7%)		-0.055 (-13.3%)		0.199*** (33.2%)	
Detailed decomposition	Exp	Unexp	Exp	Unexp	Exp	Unexp
Personal characteristics						
Education	0.237***	0.522***	0.360***	0.619***	0.297***	0.293***
Age	-0.015	-0.353	0.014*	0.475	0.031**	0.175
Ethnicity	-0.001	0.003	-0.004	0.000	0.000	-0.015
Province	-0.004	0.079*	0.012	0.103***	0.026**	0.238***
Labour market factors						
Formal contract	0.088***	0.016	0.044***	-0.065***	0.021*	-0.092***
Public sector	0.022**	0.016*	0.045***	0.018**	0.019	0.006
Household Characteristics						
Married	0.002	0.052	0.001	0.091***	0.002	0.171*
Children	0.005	0.014	0.005	0.004	0.012	-0.027
Period controls						
Year	-0.009***	0.202***	-0.007***	0.073**	-0.007***	0.126**
constant		-0.609		-1.371***		-0.676

Note: Urban/rural status is defined based on both place of residence and registration status.
Decomposition uses the urban group as the reference category.

Table A12: OLS regression by gender and registration status for only urban residents; dependent variable: (log) gross hourly wage

	Male		Female	
	Urban	**Rural**	**Urban**	**Rural**
Education years	0.073***	0.029***	0.077***	0.036***
	(0.004)	(0.003)	(0.004)	(0.003)
Age	0.011	0.051***	0.029**	0.036***
	(0.009)	(0.007)	(0.012)	(0.011)
Age squared	-0.000	-0.001***	-0.000**	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Minority	-0.007	-0.066	0.054	0.039
	(0.052)	(0.051)	(0.051)	(0.044)
Northeast	0.005	0.008	-0.045	-0.074*
	(0.034)	(0.038)	(0.036)	(0.043)
East	0.350***	0.166***	0.388***	0.135***
	(0.032)	(0.024)	(0.032)	(0.029)
Middle	0.050	0.067**	0.019	-0.050
	(0.035)	(0.028)	(0.036)	(0.033)
Formal contract	0.116***	0.163***	0.196***	0.245***
	(0.022)	(0.019)	(0.024)	(0.023)
Public sector	-0.017	-0.097***	0.106***	0.015
	(0.022)	(0.024)	(0.024)	(0.031)
Married	0.082**	0.011	0.010	-0.080**
	(0.034)	(0.026)	(0.031)	(0.032)
Children (0 - 5)	-0.033	0.012	-0.015	0.004
	(0.024)	(0.020)	(0.026)	(0.025)
Children (6 - 15)	-0.028	-0.038*	-0.054**	-0.102***
	(0.023)	(0.020)	(0.023)	(0.024)
Year_2016	0.003	-0.002	0.042*	0.050*
	(0.021)	(0.026)	(0.024)	(0.028)
Year_2018	0.139***	0.118***	0.173***	0.155***
	(0.022)	(0.026)	(0.024)	(0.029)
Year_2020	0.255***	0.203***	0.280***	0.185***
	(0.024)	(0.027)	(0.026)	(0.031)
Year_2022	0.303***	0.337***	0.332***	0.281***
	(0.024)	(0.028)	(0.026)	(0.030)
Constant	1.544***	1.497***	0.809***	1.532***
	(0.174)	(0.130)	(0.207)	(0.178)
N	6543	6070	5209	4447
adj. R^2	0.244	0.176	0.319	0.248

Standard errors in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Reference Categories: ethnicity: majority (Han); province: West; contract: no formal contract; sector: private sector; marriage: no living spouse; children: no children under 15 years old in household; year: 2014.

Table A13: Oaxaca decomposition with urban-rural difference; dependent variable: (log) gross hourly wage.

Controls	Male		Female	
Overall decomposition	Total		Total	
Urban	3.034***		2.859***	
Rural	2.894***		2.602***	
Difference	0.139***		0.257***	
Explained	0.163*** (117.3%)		0.238*** (92.6%)	
Unexplained	-0.024 (-17.3%)		0.019 (7.4%)	
Detailed decomposition	Exp	Unexp	Exp	Unexp
Personal characteristics				
Education	0.196***	0.441***	0.204***	0.422***
Age	-0.006	-0.638***	0.013***	0.077
Ethnicity	0.000	0.003	-0.001	0.001
Province	-0.029***	0.083***	-0.029***	0.137***
Labour market factors				
Formal contract	0.021***	-0.025	0.032***	-0.026
Public sector	-0.005	0.015**	0.029***	0.017**
Household Characteristics				
Married	0.006**	0.052*	0.000	0.070**
Children	0.003*	-0.009	0.004*	0.014
Period controls				
Year	-0.022***	0.009	-0.013***	0.031
constant		0.047		-0.724***

Note: Urban/rural status is defined based on registration status and only unban residents are included. Decomposition uses the urban group as the reference category.

Table A14.1: Rif decomposition with urban-rural difference at different quantiles; dependent variable: (log) gross hourly wage: MALE.

Controls	10th		50th		90th	
Overall decomposition	Tot		Tot		Tot	
Urban	2.143***		3.001***		3.990***	
Rural	2.112***		2.864***		3.763***	
Difference	0.031		0.136***		0.226***	
Explained	0.132*** (425.8%)		0.186*** (136.8%)		0.163*** (72.1%)	
Unexplained	-0.101*** (-325.8%)		-0.049** (-36.8%)		0.063* (27.9%)	
Detailed decomposition	Exp	Unexp	Exp	Unexp	Exp	Unexp
Personal characteristics						
Education	0.146***	0.438***	0.202***	0.447***	0.230***	0.282**
Age	-0.029***	-0.724*	-0.003	-0.717***	0.024**	-0.367
Ethnicity	0.000	0.005	0.000	0.003	0.000	-0.004
Province	-0.017**	0.069	-0.027***	0.04	-0.041***	0.122**
Labour market factors						
Formal contract	0.020***	-0.043*	0.024***	-0.019	0.013	-0.031
Public sector	0.016	0.029***	0.006	0.015**	-0.039**	0.000
Household Characteristics						
Married	0.013***	0.127**	0.005*	0.033	-0.005	-0.010
Children	0.006*	-0.043*	0.002	0.01	-0.002	0.021
Period controls						
Year	-0.024***	-0.045	-0.023***	-0.011	-0.016***	0.000
constant		0.086		0.150		0.049

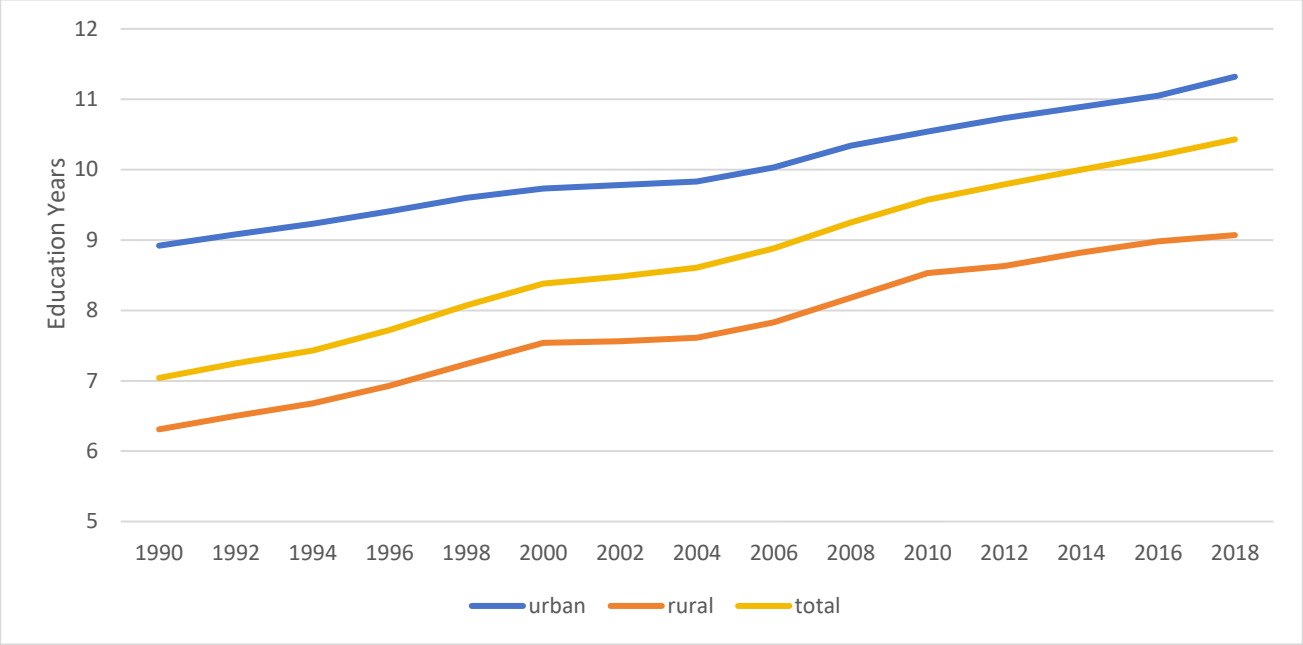
Note: Urban/rural is defined based on registration status and only unban residents are included. Decomposition uses the urban group as the reference category.

Table A14.2: Rif decomposition with urban-rural difference at different quantiles; dependent variable: (log) gross hourly wage: FEMALE.

Controls	10th		50th		90th	
Overall decomposition	Tot		Tot		Tot	
Urban	1.947		2.818		3.853	
Rural	1.793		2.564		3.465	
Difference	0.154***		0.254***		0.387***	
Explained	0.177*** (114.9%)		0.287*** (113.0%)		0.223*** (57.6%)	
Unexplained	-0.023 (-14.9%)		-0.032 (-12.6%)		0.164*** (42.4%)	
Detailed decomposition	Exp	Unexp	Exp	Unexp	Exp	Unexp
Personal characteristics						
Education	0.162***	0.410***	0.246***	0.549***	0.202***	0.198*
Age	-0.015*	-0.264	0.012*	-0.063	0.026***	-0.176
Ethnicity	0.000	0.002	-0.001	0.007	0.000	0.002
Province	-0.029***	0.082*	-0.029***	0.122***	-0.031***	0.219***
Labour market factors						
Formal contract	0.054***	0.034	0.027***	-0.067***	0.013*	-0.030
Public sector	0.021*	0.030***	0.042***	0.023**	0.018	-0.004
Household Characteristics						
Married	0.002	0.031	0.001	0.101***	0.001	0.125*
Children	0.002	0.018	0.002	0.018	0.005	-0.005
Period controls						
Year	-0.018***	0.099**	-0.013***	-0.005*	-0.012***	0.1*
constant		-0.464		-0.717**		-0.265

Note: Urban/rural status is defined based on registration status and only unban residents are included. Decomposition uses the urban group as the reference category.

Figure A1: Average education years for urban and rural working age individuals



Source: China Human Capital Report, 1990-2018.
Urban/rural is defined based on place of residence.