I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11022

The Role of China's Household Registration System in the Urban-Rural Income Differential

Ernest Boffy-Ramirez
Soojae Moon

SEPTEMBER 2017

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11022

The Role of China's Household Registration System in the Urban-Rural Income Differential

Ernest Boffy-Ramirez
University of Colorado Denver and IZA

Soojae Moon
University of Colorado Denver

SEPTEMBER 2017

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Role of China's Household Registration System in the Urban-Rural Income Differential

Together with the rapid growth of the Chinese economy, there has been a growing divide in the earnings of urban and rural residents. In this paper we focus on China's household registration system, or "hukou", as a potential source of the earnings gap. Using multiple waves of data from the Chinese Health and Nutrition Survey from 1993 through 2011, we take advantage of variation in hukou status generated by individual-level changes over time. Unlike previous studies, we are able to control for fixed individual-specific characteristics that determine earnings and focus specifically on estimating an urban hukou "premium". For estimates that do not account for time-invariant individual characteristics, urban hukou holders earn almost 30% more than rural hukou holders. After we account for individual-level fixed characteristics, the urban hukou premium drops to 6–8%. Our empirical evidence indicates that the hukou system is a notable component of the urban-rural earnings differential, but its importance should not be overstated. Given long-standing differences in access to government funding and social services between rural and urban populations, relaxing residency restrictions may not be a panacea for curbing rising income inequality.

JEL Classification: J30, J80, O15, R23

Keywords: hukou, migration, China, urban-rural income gap, inequality, labor market frictions

Corresponding author:
Ernest Boffy-Ramirez
Department of Economics
University of Colorado Denver
1380 Lawrence Street, Suite 470
Denver, CO 80204
USA

E-mail: ernest.boffy-ramirez@ucdenver.edu

1 Introduction

China has experienced unprecedented changes to its economic demography over the past 30 years. One striking feature has been the fast pace of urbanization– the population living in urban areas rose from 26.4% in 1990 to 56.8% in 2016 (United Nations Population Division's World Urbanization Prospects, 2017).[1] This shift coincides with a widening disparity between the incomes of rural and urban populations. Xie and Zhou (2014) report that the GINI coefficient based on family income rose from 0.3 in 1980 to 0.55 in 2012. Since the early 1960s, a household registry known as "hukou", has attempted to regulate labor mobility and population growth in larger cities (Sicular, et al., 2007; Wang and Zuo, 1999). This registry requires every resident to be accounted for in a household and report identifying information such as name, spouse, date of birth, and educational attainment. Importantly, each registrant is designated as an urban or rural resident depending on where the individual's parents or grandparents were registered and not necessarily where the resident lives or where they were born.[2]

Beginning in 1978, the development of labor-intensive industries created relatively higher paying employment in industrial urban areas. This has translated to mass migration and geographic-specific income differentials (Meng, 2012; Bosker, Brakman, Garretson, and Schramm, 2012; Wang and Cai, 2008; Zhu, 2002; Zhao 1999b). Cai (2011) estimated that rural to urban migrants exceeded 155 million in 2010, while Melander and Pelikanova (2013) had the number of migrant workers in cities exceeding 260 million in 2012– a third of the total urban population.

Historically, in granting a rural resident with urban status, governing bodies permitted access to entitlement programs reserved for urban residents, such as pension benefits, subsidized housing, medical insurance, workplace provided meals, and childcare.[3] Individuals holding local urban hukou are eligible to work in local state-owned enterprises, civil administration, public service, and in businesses which employ more than 7 employees, while non-locals are often unable (Zhao, 2005). According to Meng (2012), more than 50% of urban workers are professional and official workers, whereas this number is only 6% for migrant workers. Home

[1] Sicular, Ximing, Gustafsson, and Shi (2007) document that the urban population in China increased from 26% in 1990 to 48% in 2006.

[2] During the period studied, only urban and rural classifications exist. For additional historical context see Meng (2012) and Zhao (2005).

[3] See Wang and Zuo (1999) for more information on the institutions separating rural and urban populations.

buying often requires a local hukou, and in addition to these differences, a child without the appropriate status may not be able to enroll in local urban schools.[4]

An urban hukou can be difficult to acquire and is commonly understood as a barrier to upward economic mobility.[5] The Chinese government has long recognized the need for hukou reform and recent 5-year plans have explicitly mentioned the importance of "social fairness and justice" in official statements on income inequality.[6]

The hukou system inhibits labor mobility and segregates the labor force. Removing restrictions on labor market mobility may allow wages and salaries to equalize, and begin to address growing inequality (Meng and Zhang, 2001; Harris and Todaro, 1970). Zhu (2002) provides some empirical evidence of this possibility, finding the size of the urban-rural income gap has an important positive relationship with the probability of migrating for individuals originating in Hubei province.

In this paper, we use longitudinal China Health and Nutrition Survey (CHNS) data spanning 1993 to 2011 to estimate the income differential attributable to official hukou status. The primary aim is to quantify the size of the hukou urban "premium". Our analysis thoroughly addresses many of the concerns associated with using CHNS data and uses panel data methods to estimate the urban hukou premium.

Previous research has estimated rural vs. urban or migrant vs. non-migrant earnings differentials using aggregated data or controlling for personal and geographic characteristics. Yet, endogeneity is a concern, either due to the omission of unobserved individual income-related characteristics, selection bias, or non-random changes in location of residence.[7] If obtaining an official urban hukou is positively associated with unobserved income augmenting characteristics, comparisons of urban to rural hukou holders will overstate the urban hukou premium due to this positive selection. An alternate avenue for identifying the urban hukou premium would be to use variation across provinces or prefectures generated by policy changes. That said, this source of variation is limited and not the result of a natural experiment.

[4] Differences in benefit generosity and coverage are popular topics in the news media. For examples, see The Economist (2017), The Associated Press (2015), Branigan (2014, 2012), and Schiavenza (2013).

[5] Knight, Song, and Huaibin (1999) discuss the use of migrant quotas, employer taxes, and employment permit fees, to control migration at the city level in Beijing, Shenzhen, Wuhan, and Suzhou.

[6] In a speech during the Fifth Plenum of the 18th Communist Party of China Central Committee meetings, President Xi Jinping stressed the importance of allowing migrant workers to become urban residents (Xinhua News, 2015). For a list of hukou policy reforms, see Melander and Pelikanova (2013).

[7] Selection concerns are an important subject in many studies. For example, see Appleton, Song, and Xia (2014), Zhu (2002), and Zhao (1999a).

This paper differs from previous research because it uses variation generated from individual hukou status changes and estimates the hukou premium directly from actual reported status. To our knowledge, this is the only analysis to take advantage of within-individual variation to net out individual-specific characteristics that are unchanging over time.[8] For example, we are able to control for characteristics such as fixed ability or productivity which bias estimates of the urban hukou premium.

On average, urban hukou holders have 6% to 8% higher incomes than rural hukou holders after controlling for individual-fixed effects and other important covariates. This is in stark contrast to the almost 30% premium estimated with just observable demographic controls. Looking deeper, the premium is being driven by men while estimates for women are not statistically different from zero. There also appears to be a larger premium for China's ethnic minorities compared to Han Chinese.

Domestic migration patterns have and will continue to respond to changes in government policy (Bosker, Brakman, Garretson, and Schramm, 2012; Wang and Zuo, 1999). The empirical evidence presented in this study indicates that the hukou system is an important component of the urban-rural income differential but the barriers to labor mobility caused by official residency laws constitute only one of many factors that play a role in China's urban-rural income gap. The hukou system is tied to long-standing disparities in social services and infrastructure development that are likely to persist long after the hukou system is officially ended. Its elimination will not address the underlying differences in access or quality in the short-run.

The remainder of this paper is organized as follows: section 2 briefly contrasts relevant literature with our approach, section 3 is a description of the CHNS sample, including our sample construction, descriptive statistics on hukou attainment, as well as income trends between 1993 and 2011, section 4 presents our empirical strategy and section 5 presents our estimation results for various subsamples including men, women, Han Chinese, and ethnic minorities. In section 6, we summarize the results and conclude by commenting on the policy relevance of our findings.

[8] Zhao (1999a, 1999b) and Zhu (2002) use a single wave of data from Sichuan and Hubei provinces, respectively. Appleton, Song, and Xia (2014) use repeated cross-sections from the China Household Income Project.

2 Hukou in the literature

Zhao (2005) lists a number of articles that study the determinants of migration and the urban-rural income inequality in China. One branch of that work has focused on smaller geographic areas. For example, comparing urban residents and rural migrant workers in Shanghai in 1995, Meng and Zhang (2001) find that 82% of the earnings gap is explained by occupation and other individual characteristics. They note that discrimination may play a role in the determination of occupation and within-occupation wage differentials. In a survey of 4 villages, Yao (2001) finds that most of the wage gap between locals and migrants can be explained by firm type, village type, age, education, job experience, marital status, and political affiliation.

Other studies have used more representative data sets. Using the Chinese Urban Household Survey from 1988 to 2009, Meng (2012) tracks the returns to education, work experience, occupation, gender, province, and working in a state-owned enterprise over time. Together, these characteristics account for 43% of variation in annual earnings in 1988 and 30% in 2009. Using the data from the 2002 China Household Income Project, Demurger, Gurgand, Li, and Yue (2009) find that a large share of the earnings difference is due to differences in population characteristics and educational opportunities.

Sicular et al. (2007) take a closer look at location of residence and analyze the importance of inhibiting labor mobility on the wage differential between urban and rural areas. After controlling for household and individual characteristics, location of residence is the most important factor– accounting for 75% of the remaining wage gap.

We build upon the previous literature in three ways. First, we use the CHNS which tracks individuals over time– recording hukou status separately from current location. Second, the panel covers a long time period and a more representative sample of individuals from a greater number of provinces. Third, we directly focus on identifying the hukou premium using within-individual variation after controlling for various determinants of earned income.

A limitation of our study is our final sample focus is a subset of the CHNS. To reduce the possibility of measurement error in recorded hukou status, we modify our sample to avoid questionable status changes. While this increases confidence in the internal validity of our study, we run the risk of reducing its external validity. We detail our approach in the data section.

3 The CHNS sample

The CHNS is an ongoing longitudinal data set initiated in 1989 with individual, household, geographic, and community characteristics. Household and individual surveys were conducted inside the participant's home. Hukou status was collected starting in 1993.

The retirement age for women is 55, but can be as early as 50 in certain occupations. For men, the retirement age is as early as 60. For the purposes of our analysis, we limit the sample to individuals ages 23 to 49– individuals likely to be finished with their education and not retired.[9] Our sample includes a diverse set of provinces including: Guangxi, Guizhou, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Liaoning, and Shandong. These provinces are shown in Figure 1.[10]

[Figure 1]

Table 1 shows the number of individuals in each survey wave. Individuals observed in one wave or in more than 7 waves are removed. The remaining 13,597 unique individuals account for 50,682 total observations.[11]

[Table 1]

Table 2 reports how many times the same individual is observed across the 7 waves. An advantage of the CHNS is that it contains a large number of individuals observed in many waves, allowing us to estimate individual fixed-effects.

[Table 2]

The number of individuals observed in at least 3 waves is large. For example, 2,849 individuals are observed in 4 waves.

[9] College students receive temporary residency hukou for the duration of their studies. Though temporary status restricts official concurrent employment in regular universities, we still check the data for individuals who are both in school and earning an income. We find that only 146 individuals ever report simulataneously earning an income and attending school.

[10] The CHNS samples mega cities Beijing, Shanghai, and Chongqing but not enough waves of data are available yet.

[11] For more information on the survey's representativeness, see the general notes in the appendix and the CHNS website at http://www.cpc.unc.edu/projects/china

3.1 Measuring hukou

Our primary variable of interest is hukou. The CHNS is one of the few data sets that records it, yet using the hukou variable is problematic.[12] According to CHNS representatives, some misclassification of hukou may be present due to city expansion and redistricting, surveyor error, and data entry error. While we have no indication that errors are systematic, misclassification of hukou status magnifies attenuation bias in individual fixed-effects regression estimates compared to estimates without individual fixed-effects (Freeman, 1984).

Individual fixed-effects and an extensive set of controls will mitigate geographically and temporally systemic errors, but to directly address this concern we use two different measures of hukou status. The baseline measure of hukou status is the unaltered version provided by the CHNS. With the understanding that hukou status is difficult to change and that misclassification errors are present in the data, we recode individuals who change status two or more times in the 7 waves as missing values.[13] We believe our alternative measure is more accurate in reflecting genuine hukou status changes. The alternative measure, removes questionable observations and allows us to be more confident in the validity of our results (albeit for a smaller sample). That said, the alternate measure may discard legitimate status changes, so the measurement error issue is not completely resolved.

Table 3 reports urban and rural hukou counts for the two measurements beginning in 1993 when hukou status was first recorded. We name the unaltered measure of hukou status as "hukou A".

[Table 3]

Table 3 shows that our alternative hukou variable decreases the size of the sample in every year. The lack of a valid status for a large number of individuals in later waves raises additional selection concerns, so our analysis leaves individuals with missing hukou values in the dataset and they are analyzed separately from individuals with a reported hukou. If we do not know why certain individuals do not have a hukou status reported in the data or why the percentage of non-reporters changes over time, then we cannot be

[12] The Longitudinal Survey on Rural Urban Migration in China and the Chinese Health and Retirement Longitudinal Survey record hukou status but have a very limited numbers of waves available. The Chinese General Social Survey also records hukou but is composed of repeated cross-sections.

[13] For example, hukou histories of RRURR, URRU, RRUR_RR, RUR, RURURR, URRUU, RUURU, etc. are replaced with missing values for their entire history. The remaining individuals have hukou histories like UUUR, RRUU, RUUUU, etc. depending on the number of waves they are observed in.

sure that the sample with a reported hukou is representative. This issue highlights the need to use within-individual variation instead of across-individual variation.[14] That said, there are only 82 observations where an individual with a missing hukou measure A status reports wage and bonus income.

Hukou status varies separately from location. Table 4 shows that urban hukou holders tend to be in urban areas and rural holders in rural areas, but there is substantial variation in location conditional on hukou designation.

[Table 4]

Based on modified hukou variable, 20% of the rural hukou observations are recorded as not being in a rural village. 33% of urban hukou observation are recorded as not being in an urban neighborhood.

3.2 Identifying variation

Identifying variation comes from individuals who change their hukou. To get a sense of the available variation, Table 5 lists the number of people who switched hukou and the type of switch. Switchers are identified as individuals who at some point over the sample period recorded two different hukou registration statuses.

Official switches may occur through marriage, business investment, military service, employer sponsorship, and city expansion or redistricting. According to CHNS representatives, city expansion or redistricting can change an individual's hukou status without the individual moving. This type of variation is the result of a natural experiment.

Of the subsample with a reported hukou, 8.63% (937 unique individuals) have changed their status from rural to urban and 6.86% have changed it changed it from rural to urban at some point between 1993-2011.[15]

[Table 5]

These numbers are quite large, and indeed, some individuals report changing status more than once so they may be double counted. Using the alternate measure of hukou, the number of rural to urban switches falls to 4.40% of the sample.[16]

[14] On average, individuals with missing values of hukou are more similar to rural hukou holders with respect to age, education, sex, marital, education, and minority status. Nonetheless, they are different enough to justify categorizing them them separately. A comparison of individuals with a reported hukou status to those without one is detailed in the appendix.

[15] Both types of switches are evenly spread across waves. There is no evidence of bunching in the timing of switches.

[16] The remaining switches are solely from individuals who switched once over the sample period and have a complete hukou

3.3 Income differences by hukou

To understand differences in income by hukou, we focus "annual wage, bonus, and other income" from the primary job. Hukou is most relevant in formalized sectors where individuals earn a salary or are paid wages and receive benefits (e.g. white-collar work vs. farming or self-employment).[17] The CHNS measures income from other sources including individual farming, fishing, business, gardening, livestock, and retirement payments but we do not focus on these other sources of income.[18]

Figure 2 plots the average real annual wage and bonus income over time using individuals included in hukou measure B. Displayed is a local linear smoothing function with a 95% confidence interval.[19] Annual wage and bonus incomes are not top-coded and are chained to 2010 yuan using the World Bank's GDP deflator. The urban hukou group is reported in red.

[Figure 2]

The difference in average real annual wage and bonus income between the two hukou groups is not statistically significant until after 1997 when the separation between urban and rural hukou holders continually increases. These trends mirror the overall trend in GDP per capita which shows a slower growth period between 1993-2000 and rapid growth period post-2000 (World Bank Development Indicators).[20]

Table 6 reports median values by hukou and year.[21] In both hukou measures, the difference in real wage and bonus income between the two group begins in earnest after 1997. A non-parametric test of median equality at each year results in statistically significant differences in every year, but p-values are larger in 1993 and 1997. The difference in 1997 is almost not significant at the 5% level using hukou measure B.

history. For example, the individual RRR_UU is a hukou switcher but the timing is not known so they are not included in Table 5. This is not problem for the estimation.

[17] Non-retirement wage or salary income is calculated by taking the average monthly income times the number of months worked. Bonuses, cash transfers, and other in-kind payments from the employer are included.

[18] If all sources are included, total income can be negative in a given year. 61% of households who raised livestock reported higher expenses than revenues during at least one of the waves of data collection. This is due to the cyclic nature of raising livestock. Other activities like farming and gardening are subject to weather and market prices that can vary from year to year. As the project director notes, "It's normal that a person invested big money in livestock or other businesses and earned nothing in one year, earned a lot in the second year, and again earned nothing in the third year. I heard such stories several times when I participated in data collection or supervised the fieldwork."

[19] The epanechnikov kernel smoothing function is estimated each year from 1993-2011. The degree of the function is inconsequential.

[20] This may be related to the central government's encouraging of limited migration between 1992-1995 or the tightening of controls in many major cities beginning in 1995 (Zhao, 2005).

[21] "No hukou status" medians are not reported since sample sizes are extremely small. By year, in the subsample with reported income values, the number of individuals without a reported hukou measure A are 0, 8, 66, 4, 1, 1, and 2, respectively. These numbers are larger for hukou measure B due to our recoding. By year, in the subsample with reported income values, the number of individuals without a reported hukou measure B are 127, 197, 262, 134, 151, 145, and 136, respectively.

[Table 6]

Interestingly, in 1993, rural hukou holders earn more than urban hukou holders in terms of the median and the average, but this quickly reverses in the following wave. Focusing on hukou measure B, by 2000, the urban hukou median is about 27% larger than the rural hukou median. This grows to over 75% by 2009. Figure 2 and Table 6 are consistent with Meng's (2012) finding using the Chinese Urban Household Survey that between 1988 and 2009, urban hukou workers saw fivefold increases in real annual earnings.

4 Estimation strategy

The previous summary measures indicate the potential for a large urban hukou premium, yet uncontrolled differentials are subject to biases arising from unobserved differences between groups. If switching is costly in terms of direct and information costs, switchers may be positively selected. For example, obtaining urban hukou in a desirable city can depend on education, financial resources, and individual characteristics.[22] A straight comparison of rural and urban hukou holders will identify the urban hukou premium and the positive selection associated with having the resources to change status.

We take advantage of the longitudinal data to estimate the impact of individual hukou status changes using individual-fixed effects. Using *within-individual* hukou variation provides a cleaner estimate of the hukou premium and addresses selection concerns. The CHNS provides the rare opportunity to difference out time invariant factors specific to the individual that are correlated with changing hukou or reporting a hukou status.

This empirical strategy allows us to ask whether earnings differ for the same individual across different statuses. A key advantage of this strategy is we do not need to find a separate control group that is observably similar. Without a true natural experiment, we argue that the hukou premium estimates we present are more reliable than previous estimates.[23]

Our main specification is as follows:

[22] Beijing and Chongqing have experimented with systems that grant points towards a local urban hukou based on educational attainment, occupation, a history of paying local taxes, a clean criminal record, and other socially desirable characteristics.

[23] Our empirical strategy does raise questions of external validity. The estimates are being derived from the subsample of individuals who switched. Nonetheless, our extensive controls will mitigate geographically or temporally systemic biases.

$$\ln(inc_{it}) = \alpha + \beta \cdot urban_{it} + X'\delta + \sum_{t=2}^{7} \varphi_t \cdot wave_t + \sum_{i=2}^{n} \theta_i \cdot individual_i + \varepsilon_{it}$$

where the coefficient of interest is β, the estimate on the indicator measuring the average difference in the natural log of real yearly wage and bonus income between urban and rural hukou holders (a separate indicator is included to represent individuals with no reported hukou). The vector X contains individual-level indicator variables for high school completion (at least "upper middle" school) and marital status (married or not married), and a cubic function in age. We stratify our estimates by sex and minority status in a separate set of results. φ are year-fixed effects which control for income shocks that effect all individuals' earnings across all provinces. θ represent the individual-fixed effects.

For comparability to previous studies, we also run specifications without individual-fixed effects but with the individual-level controls, wave-fixed effects, and province-fixed effects. This specification is as follows:

$$\ln(inc_{its}) = \alpha + \beta \cdot urban_{its} + X'\delta + \sum_{t=2}^{7} \varphi_t \cdot wave_t + \sum_{s=2}^{9} \omega_s \cdot province_s + \varepsilon_{its}$$

The same vector of controls are included, but now province-fixed effects ω are included.[24] In all specifications, standard errors are clustered on province.

5 Estimation results

5.1 Primary results

The first column in Table 7 reports the coefficient on β without any controls pooling all waves of data. We report the results using both hukou measures but the two measures are extremely similar so the discussion focuses on our preferred measure– hukou B.

[Table 7]

Relative to rural hukou holders, urban hukou holders earn approximately 39% more when no controls are included in column iv. When year, province, and individual demographic controls are included, the difference decreases to approximately 30%.

[24] Province-fixed effects are perfectly predicted by the individual-fixed effects because individual ID numbers are generated by concatenating province, community, city, and household codes. Thus, individual-fixed effects will capture various income predictors. Nonetheless, a comparison of specifications show individual-specific factors other than province-fixed effects are responsible for the dramatic drop in the urban hukou premium. The CHNS says it tried to find individuals who have a rural hukou, but were working in other city.

The importance of individual fixed-effects can be observed in column vi. After adding the individual fixed-effects, on average a switch from rural to urban hukou represents almost 8% more in earnings. This difference is not significant ($p = 0.101$). A much larger portion of the variation in logged wage and bonus income is explained.[25] Unsurprisingly, the education indicator is consistently large and positive.

For the sample of individuals who changed their hukou status, Table 7 suggests the gap attributable to hukou is much less pronounced. Unobserved but fixed individual specific characteristics help explain a large portion of the overall gap between urban and rural hukou holders. This conclusion is not affected by the hukou measure used.[26]

Individuals who do not have a reported status have consistently higher average wage and bonus income– sometimes higher than urban hukou holders. Estimates vary across the two measures due to hukou measure B containing more recoded observations, but if we focus on the alternative measure, it appears that the difference between rural hukou holders and those without a reported status is unchanging. Individuals without a reported hukou are younger and are slightly more likely to be female, not Han Chinese, and married. They are less educated than urban hukou holders but are more educated than rural hukou holders. Without further information on the composition of this group, we will withhold drawing conclusions regarding this small group and further investigate the gaps by stratifying the sample by gender and minority status.

5.2 Differences by gender

The individual fixed-effects in Table 7 capture both gender and minority status, so Tables 8 and 9 report the hukou premium separately for each group.[27]

[Table 8]

Relative to rural hukou holders, urban hukou holders earn approximately 36% and 50% more when no controls are included for men and women, respectively. While the premium decreases as controls are added,

[25] The decrease is sample size for specifications with individual fixed-effects reflects the fact that the hukou coefficient is estimated from a subsample of hukou changers. The change in the coefficients and R^2 between column iv and v is consistent with Meng's (2012) conclusion about the relative importance of education and province in explaining logged earnings, though in our study age is also found to be a significant factor.

[26] Region-specific linear trends may account for differences in income due to different economic growth trajectories between central, coastal, and northern provinces. Including these trends has such a minor effect on our main results that we do not include them in our tables.

[27] Despite the potential for measurement error in reported hukou, Table 7 shows hukou measures A and B produce similar results in urban-rural difference estimates. Going forward, the estimates are based on hukou variable B only. Estimates based on hukou measure A can be provided upon request.

there are notable differences between men and women. When the full set of controls are included there appears to be no urban hukou premium for women, indicating that the estimates in Table 7 are being driven by men.

In the female specification with individual fixed-effects, the education indicator is insignificant, as is the indicator for no reported hukou. After controlling for individual fixed-effects there are no statistical differences between hukou groups. Whether a result of occupational selection, differential application of employment laws, or simply that hukou is less relevant for women in the labor force, the observed difference between men and women is an area worth further investigation.

5.3 Differences by minority status

The CHNS oversamples ethnic minorities. Minorities account for 14% of the sample.[28] Higher minority representation in rural areas reflect the sample design and the large number of Miao, Man, and Buyi in the CHNS. This is not surprising given the sampled provinces include Guizhou, Liaoning, and Heilongjiang.

Table 9 reports difference estimates by minority status. An individual is a minority if they identify as Mongolian, Hui, Tibetian, Vaguer, Miao, Yi, Zhuang, Buyi, Korean, Man, Dong, Yao, and Tujia peoples.

[Table 9]

A familiar pattern can be seen in Table 9. While the size of the minority sample that experienced hukou changes is relatively small, the urban hukou premium is substantially larger for minorities compared to Han Chinese. On average, for minorities, a change from rural to urban hukou is associated with approximately 16% more wage and bonus income.[29]

6 Conclusion

The primary aim of this study is to describe the urban hukou premium. Our estimates of the urban hukou premium help inform the debate as to the role residency restrictions play in the urban-rural income differential. On average, urban hukou holders earn about 6% to 8% more than rural hukou holders after

[28] According to the 2010 Chinese census, 8.49% on the population were of various national minority populations (National Bureau of Statistics of People's Republic of China, 2011).

[29] The coefficient on having a missing hukou value in the sample of minorities in column vi is dropped due to perfect collinearity.

controlling for individual-fixed effects and other important covariates. Additional analysis strongly suggests that the urban hukou income premium is being driven by the men in the sample. Interestingly, estimates for women are not statistically different from zero. This gender difference merits further investigation.

China's recent history has shown that domestic migration patterns will respond to changes in government policy. In theory, removing this friction in the labor market would allow wages and salaries across areas to equalize, and thus begin to address growing income inequality in China. The empirical evidence presented in this study indicates that hukou is an important component of the urban-rural income differential but its role should not be overstated. It is important to acknowledge that the hukou system is tied to disparities in social services and infrastructure development that are likely to persist long after the hukou system is officially ended. Its elimination will not immediately address the underlying differences in access or quality.

Frictions to labor mobility, like hukou, constitute only one factor among many. Strong geographic wealth disparities, uneven fixed capital investments, differential access to human capital, and local government institutions all play a role in the growth of income inequality in China. As more data is collected, future research should continue to focus on understanding the differences in the composition of urban-to-rural and rural-to-urban switchers. Specifically, how the type of switch depends on the return to skills in urban and rural areas in the spirit of Zhu (2002) and Roy(1954).

References

[1] Appleton, Simon, Lina Song, and Qingjie Xia. 2014. "Understanding Urban Wage Inequality in China 1988-2008: Evidence from Quantile Analysis", *World Development,* 62(C): 1-13

[2] Associated Press, The. 2015. "China to ease restrictions on living in cities for millions", *The Guardian Online*, December 12, retrieved from http://www.theguardian.com/world/2015/dec/12/china-to-give-more-people-access-to-basic-public-services-in-urban-areas

[3] Branigan, Tania. 2012. "China's 'golden decade' brings some relief to rural poor", *The Guardian*, November 5, retrieved from http://www.theguardian.com/world/2012/nov/05/china-golden-decade-rural-poor

[4] Branigan, Tania. 2014. "China reforms hukou system to improve migrant workers' right", *The Guardian*, July 31, retrieved from http://www.theguardian.com/world/2014/jul/31/china-reform-hukou-migrant-workers

[5] Bosker, Maarten, Steven Brakman, Harry Garretsen, and Marc Schramm. 2012. "Relaxing Hukou: Increased labor mobility and China's economic geography", *Journal of Urban Economics,* 72(2): 252-266

[6] Cai, Fang. 2011. "Hukou Reform and Unification of Rural-Urban Social Welfare", *China and World Economy*, 19(2): 33-48

[7] Demurger, Sylvie, Marc Gurgand, Shi Li, and Ximing Yue. 2009. "Migrants as second-class workers in urban China? A decomposition analysis", *Journal of Comparative Economics*, 37(4): 610-628

[8] Economist, The. 2017. "How Chinese schools discriminate against 65% of the population", *The Economist*, April 12, retrieved from https://www.economist.com/news/leaders/21720602-rural-pupils-are-shut-out-city-schools-and-neglected-their-villages-cruel-and

[9] Freeman, Richard B. 1984. "Longitudinal Analysis of the Effects of Trade Unions", *Journal of Labor Economics*, 2(1): 1-26

[10] Harris, John R., and Michael P. Todaro. 1970. "Migration, Unemployment and Development: A Two-Sector Analysis", *American Economic Review*, 60(1): 126-142

[11] Knight, John, Lina Song, and Jia Huaibin. 1999. "Chinese Rural Migrants in Urban Enterprises: Three Perspectives", *Journal of Development Studies*, 35(3): 73-104

[12] Kuhn, Peter J., and Kailing Shen. 2015. "Do Employers Prefer Migrant Workers? Evidence from a Chinese Job Board", I*ZA Journal of Labor Economics,* 4(22)

[13] Melander, Annika, and Kristyna Pelikanova. 2013. "Reform of the Hukou System: A Litmus Test of the New Leadership", *ECFIN: Economic Briefs*, 26: 1-16, retrieved from http://ec.europa.eu/economy_finance/publications/economic_briefs/2013/eb26_en.htm

[14] Meng, Xin. 2012. "Labor Market Outcomes and Reforms in China", *Journal of Economic Perspectives*, 26(4): 75-102

[15] Meng, Xin, and Junsen Zhang. 2001. "The Two-Tier Labor Market in Urban China", *Journal of Comparative Economics*, 29(3): 485-504

[16] National Bureau of Statistics of People's Republic of China. 2011. "Communique of the National Bureau of Statistics of People's Republic of China on Major Figures of the 2010 Population Census No.1", *National Bureau of Statistics of China,* April 28, retrieved from http://www.stats.gov.cn/english/NewsEvents/201104/t20110428_26449.html

[17] Popkin, Barry M., Shufa Du, Fengying Zhai, and Bing Zhang. 2009. "Cohort Profile: The China Health and Nurtrition Survey–monitoring and understanding socio-economic and health change in China, 1989-2011", *International Journal of Epidemiology*, 39(6): 1435-1440

[18] Roy, Arthur. 1954. "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers,* 3(2): 135-146

[19] Schiavenza, Matt. 2013. "Mapping China's Income Inequality", *The Atlantic*, September 13, retrieved from http://www.theatlantic.com/china/archive/2013/09/mapping-chinas-income-inequality/279637/

[20] Sicular, Terry, Yue Ximing, Björn Gustafsson, and Li Shi. 2007. "The Urban-Rural Income Gap and Inequality in China", *Review of Income and Wealth*, 53(1): 93-126

[21] United Nations Department of Economic and Social Affairs Population Division. 2017. "World Urbanization Prospects", *United Nations Department of Economic and Social Affairs Population Division*, retreived from World Bank Development Indicators Data Bank http://databank.worldbank.org/data/reports.aspx?source=2&country=CHN

[22] Wang, Feng and Xuejin Zuo. 1999. "Inside China's Cities: Institutional Barriers and Opportunities for Urban Migrants", *The American Economic Review*, 89(2): 276–280

[23] Wang, Meiyan and Fang Cai. 2008. "Gender Earnings Differential in Urban China", *Review of Development Economics*, 12(2): 442-454

[24] Xie, Yu and Xiang Zhou. 2014. "Income inequality in today's China", *Proceedings of the National Academy of Sciences*, 111(19): 6928-6933

[25] Xinhua News. 2015. "Xi expounds on guideline for 13th Five-year plan", *Xinhuanet*, November 3, retrieved from http://news.xinhuanet.com/english/2015-11/03/c_134780297.htm

[26] Yao, Yang. 2001. "Social Exclusion and Economic Discrimination: The Status of Migrants in China's Coastal Rural Area", Working Paper No. E2001005, China Center for Economic Research, Peking University

[27] Zhao, Yaohui. 1999a. "Labor Migration and Earnings Differences: The Case of Rural China", *Economic Development and Cultural Change*, 47(4): 767-782

[28] Zhao, Yaohui. 1999b. "Leaving the Countryside: Rural-to-urban Migration Decisions in China", *American Economic Review*, 89(2): 281-286

[29] Zhao, Zhong. 2005. "Migration, Labor Market Flexibility, and wage determination in China: a review," *The Developing Economies*, 43(2): 285-312

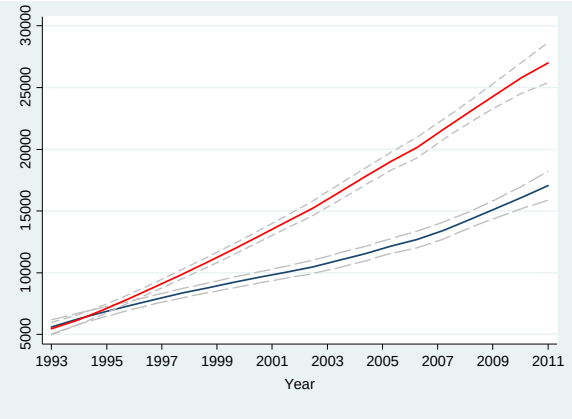
[30] Zhu, Nong. 2002. "The Impacts of Income Gaps on Migration Decisions in China", *China Economic Review*, 13(2-3): 213-30

Figure 1. Distribution of sampled provinces in the CHNS

Source: China Health and Nutrition Survey website.

Note: Yunnan, Zhejiang, and Shaanxi are not in the master survey so they cannot be included.

Figure 2. Real average wage and bonus income by hukou



Notes: The average real wage and bonus income for urban hukou holders is in red. The dotted lines represents the 95% confidence bounds. Only individuals with a reported hukou status B and are in the sample. Income values are in 2010 RMB.

Table 1. Number of observations in each wave

Survey year	1993	1997	2000	2004	2006	2009	2011
n	4,943	6,500	7,410	6,750	8,616	8,653	7,810

Note: Individuals are ages 23-49 from 9 provinces. Individuals that are reported in only 1 wave or in more than 7 waves are removed.

Table 2. Number of times individuals are observed

Times observed	2	3	4	5	6	7
n	3,643	3,192	2,849	1,850	1,267	796

Note: The same sample criteria from Table 1 is used. Individuals cannot be counted in more than one column.

Table 3. Urban & rural hukou counts

Hukou A

Survey year	1993	1997	2000	2004	2006	2009	2011
Urban	1,477	2,041	2,320	1,928	1,787	1,649	1,407
Rural	3,062	3,699	4,032	2,939	2,780	2,736	2,229
No status	404	760	1,058	1,883	4,049	4,268	4,174

Hukou B

Survey year	1993	1997	2000	2004	2006	2009	2011
Urban	1,367	1,857	2,154	1,715	1,620	1,498	1,299
Rural	2,877	3,505	3,775	2,751	2,570	2,582	2,077
No status/recoded	699	1,138	1,481	2,284	4,426	4,573	4,434

Note: The same sample criteria from Table 1 is applied.

Table 4. Location surveyed by hukou classification

	Hukou A		
	Urban hukou	Rural hukou	No status
Urban neighborhood	3,845	203	1,126
Suburban village	2,849	3,518	1,790
County town neighborhood	4,471	963	2,515
Rural village	1,444	16,793	11,165

	Hukou B		
	Urban hukou	Rural hukou	No status/recoded
Urban neighborhood	3,744	106	1,324
Suburban village	2,512	3,126	2,519
County town neighborhood	4,173	732	3,044
Rural village	1,081	16,173	12,148

Note: The same sample criteria from Table 1 is applied.

Table 5. Number of hukou switchers

	Hukou A	Hukou B
Rural to urban	937	457
Urban to rural	745	264

Notes: The same sample criteria from Table 1 is applied. Of the sample who have a recorded hukou status, a switcher is an individual who reported a different status in the following wave.

Table 6. Wage & bonus income medians by year & hukou

Hukou A							
	1993	1997	2000	2004	2006	2009	2011
Urban	4,251	7,308	9,743	13,914	14,906	19,251	22,120
Rural	4,700	7,071	7,696	9,651	9,938	10,941	13,272
Hukou B							
	1993	1997	2000	2004	2006	2009	2011
Urban	4,295	7,240	9,789	14,477	14,907	19,251	22,120
Rural	4,652	7,071	7,696	8,847	9,689	11,064	13,272

Note: The same sample criteria from Table 1 is applied. Annual wage and bonus incomes are not top-coded and are chained to 2010 RMB using a World Bank provided GDP deflator.

Table 7. Estimated real wage & bonus income differences

	Hukou A			Hukou B		
	i.	iii.	iii.	iv.	v.	vi.
Urban hukou	0.3655***	0.2806***	0.0609*	0.3875***	0.2976***	0.0781
	(0.0384)	(0.0358)	(0.0327)	(0.0385)	(0.0373)	(0.0476)
No status	0.3028**	0.4343***	0.1718	0.2621***	0.2420***	0.2464*
	(0.0929)	(0.1017)	(0.1181)	(0.0550)	(0.0463)	(0.1370)
High school		0.3008***	0.1192***		0.2962***	0.1196***
		(0.0316)	(0.0384)		(0.0326)	(0.0384)
Married		0.0015	0.0700**		-0.0028	0.0700**
		(0.0409)	(0.0333)		(0.0414)	(0.0333)
$f(age)$		✓	✓		✓	✓
Individual fe			✓			✓
Wave fe		✓	✓		✓	✓
Province fe		✓			✓	
R^2	0.0352	0.2570	0.6956	0.0362	0.2574	0.6956
Obs.	13,734	13,548	10,705	13,734	13,548	10,705

Notes: The outcome variable in all specifications is the natural log of annual real wage, bonus, and other income. Hukou, married, and high school, are all indicators equal to 1 if the condition is true. The relationship between age and income is modeled as a cubic function. Standard errors all in parenthesis and are clustered by province. *** indicates p-values less than 0.01, ** less than 0.05, and * less than 0.10.

Table 8. Estimated real wage & bonus income differences for men & women

	Men			Women		
	i.	iii.	iii.	iv.	v.	vi.
Urban hukou	0.3550***	0.2920***	0.1456***	0.4972***	0.3899***	-0.0539
	(0.0373)	(0.0448)	(0.0427)	(0.0501)	(0.0383)	(0.0501)
No status	0.2389***	0.2368***	0.3471**	0.3604***	0.3375***	-0.0754
	(0.0645)	(0.592)	(0.1713)	(0.0663)	(0.0541)	(0.1394)
High school		0.2155***	0.1395***		0.3777***	0.0770
		(0.0343)	(0.0444)		(0.0302)	(0.0728)
Married		0.1116**	0.1395*		-0.1116**	0.0630
		(0.0435)	(0.0440)		(0.0362)	(0.0479)
$f(age)$		✓	✓		✓	✓
Individual fe			✓			✓
Wave fe		✓	✓		✓	✓
Province fe		✓			✓	
R^2	0.0328	0.2860		0.0550	0.2684	0.5605
$Obs.$	7,846	7,744		5,888	5,804	4,436

Notes: The outcome variable in all specifications is the natural log of annual real wage, bonus, and other income. Hukou, married, and high school, are all indicators equal to 1 if the condition is true. The relationship between age and income is modeled as a cubic function. Standard errors all in parenthesis and are clustered by province. *** indicates p-values less than 0.01, ** less than 0.05, and * less than 0.10.

Table 9. Estimated real wage & bonus income differences for Han & minorities

	Han			Minority		
	i.	iii.	iii.	iv.	v.	vi.
Urban hukou	0.3960***	0.3004***	0.0531	0.2873***	0.2899***	0.1624
	(0.0460)	(0.0413)	(0.0336)	(0.0635)	(0.0315)	(0.1376)
No status	0.2708***	0.2417***	0.1588	0.1187	0.2232*	
	(0.0578)	(0.0447)	(0.1185)	(0.1142)	(0.1099)	
High school		0.2906***	0.1360***		0.3857***	-0.0800
		(0.0354)	(0.0403)		(0.0384)	(0.1454)
Married		0.0026	0.0726**		-0.0906	0.01568
		(0.0367)	(0.0344)		(0.1253)	(0.1273)
$f(age)$		✓	✓		✓	✓
Individual fe			✓			✓
Wave fe		✓	✓		✓	✓
Province fe		✓			✓	
R^2	0.0376	0.2576	0.6931	0.0195	0.2926	0.7229
Obs.	12,416	12,245	9,755	1,265	1,250	913

Notes: The outcome variable in all specifications is the natural log of annual real wage, bonus, and other income. Minority indicates individuals who are not Han Chinese. Urban hukou, no status, married, and high school, are all indicators equal to 1 if the condition is true. The relationship between age and income is modeled as a cubic function. Standard errors all in parenthesis and are clustered by province. *** indicates p-values less than 0.01, ** less than 0.05, and * less than 0.10.

7 Appendix

7.1 A few notes on CHNS representativeness

Across all waves and without any restrictions there are a total of 31,208 unique individuals in the CHNS. 85% of households have been surveyed in at least five rounds between 1989 and 2006 (Popkin, Du, Zhai, and Zhang 2009). The baseline sample is representative of each province sampled, but is not representative over time due to sample attrition (despite efforts to maintain the original representativeness). The CHNS asserts that their survey methodology creates a random and representative sample of counties, cities, and individuals conditional on the province but it is not possible to assess the representativeness across time due restricted access to Chinese government population statistics. Following the recommendation of the CHNS, no sample weights are used in our study.

7.2 Analysis of individuals without a reported hukou status

In this appendix we compare individuals who do not have a reported hukou status with those that do. One reason for the growth in non-reporting over time is the increased use of illegal or make-shift housing and residence in unincorporated urban areas. Continued migration to cities from the countryside implies that the number of unofficial residents are increasing over time (Wang and Zuo, 1999). Individuals who are living and working without local hukou may be apprehensive of reporting their hukou status. In addition, since hukou is tied to a household, individuals who are away from their household may not have their official physical documentation.

Table A shows mean differences between groups and the statistical significance of the difference. Individuals without a reported hukou are younger. They have a slightly larger proportion of women, minorities (not Han Chinese), and married individuals.

We find interesting differences in the education differences and location. When compared to individuals without a hukou, rural hukou holders have a smaller proportion of upper middle school graduates. A much larger proportion of urban hukou holders are upper middle school graduates. Compared to rural hukou holders, individuals without a reported hukou are also more likely to not be living in rural villages. If

individuals who do not report a hukou are predominantly urban migrants, this pattern is consistent with the possibility of positive selection of non-locals into migration as discussed in Kuhn and Shen (2015).

The differences in observable characteristics suggest differences in unobservable characteristics as well, but it is also clear that individuals without a reported hukou are more similar to rural hukou holders if we look at age, proportion female, proportion non-Han Chinese, proportion married, and proportion with an upper middle school education.

Table A: Average differences by hukou reporting status

	$\overline{x}_{rural\ hukou} - \overline{x}_{no\ hukou}$	$\overline{x}_{urban\ hukou} - \overline{x}_{no\ hukou}$
Age	3.45***	3.78***
Female	-0.05***	-0.06***
Minority	-0.01***	-0.08***
Rural village	0.16***	-0.54**
Upper middle school	-0.16***	0.26***
Married	-0.03***	-0.09***

Notes: The same sample criteria from Table 1 is applied. The authors' alternative measure of hukou is used is all calculations (hukou B). All differences are based on two-sample t-tests comparing individuals without a reported hukou to individuals with a valid modified hukou variable. * indicates a statistically significant difference at the 10% level, ** at the 5% level, and *** and the 1% level.