ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Xing, Chunbing

Working Paper

Migration, self-selection, and income distributions: Evidence from rural and urban China

IZA Discussion Papers, No. 4979

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Xing, Chunbing (2010) : Migration, self-selection, and income distributions: Evidence from rural and urban China, IZA Discussion Papers, No. 4979, Institute for the Study of Labor (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/36923

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

DISCUSSION PAPER SERIES

IZA DP No. 4979

Migration, Self-Selection, and Income Distributions: Evidence from Rural and Urban China

Chunbing Xing

May 2010

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Migration, Self-Selection, and Income Distributions: Evidence from Rural and Urban China

Chunbing Xing
Beijing Normal University
and IZA

Discussion Paper No. 4979
May 2010

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 4979
May 2010

ABSTRACT

Migration, Self-Selection, and Income Distributions: Evidence from Rural and Urban China*

As massive rural residents leave their home countryside for better employment, migration has profound effects on income distributions such as rural-urban income gap and inequalities within rural or urban areas. The nature of the effects depends crucially on who are migrating and their migrating patterns. In this paper, we emphasize two facts. First, rural residents are not homogeneous, they self-select to migrate or not. Second, there are significant differences between migrants who successfully transformed their hukou status (permanent migrants) and those did not (temporary migrants). Using three coordinated CHIP data sets in 2002, we find that permanent migrants are positively selected from rural population especially in terms of education. As permanent migration takes more mass from the upper half of rural income density, both rural income level and inequalities decrease, the urban-rural income ratio increases at the same time. On the contrary, the selection effect of temporary migrants is almost negligible. It does not have obvious effect on rural income level and inequalities.

JEL Classification: O15

Keywords: migration, self-selection, income distribution, China

Corresponding author:

Chunbing Xing
Beijing Normal University
in 2010 on leave at:
University of Western Ontario
Room 4012, Social Science Centre
London, Ontario, N6A 5C2
Canada
E-mail: xingchunbing@gmail.com

* I am grateful for the suggestions and comments of Li Shi from Beijing Normal University, Aldo Colussi from University of West Ontario, Chen Yuyu from Peking University, and many other participants in the seminars at University of Western Ontario, Beijing Normal University, and OECD Development Center. The author also acknowledges the financial supports from the Ministry of Education of the People's Republic of China (No. 08JC790008) and the IDRC/CIGI Young China Scholars Poverty Research Network.

Migration, Self-selection, and Income Distributions: Evidence from Rural and Urban China

1. Introduction

The widening rural-urban income gap and massive migration from rural to urban areas are of utmost concerns for researchers and policy makers nowadays.[1] Migration is obviously the consequence of the income gap, and is often regarded as a way to reduce it as rural population leave for a better paid job in urban areas. The reverse direction of causality, however, is neglected. Migration can also serve as reason of widening rural-urban income gap. The reason is simple: the rural population is not homogeneous, they ***self-select*** to migrate or not. Whether migration reduces the gap or not, to large extent, depends on who are migrating and their migrating patterns. In particular, if the rural migrants are mainly composed of those who are more able and more educated ("positively" selected), and if they are well integrated in the urban labor market, the gap will probably augment. The reverse will probably be true if the migrants are randomly chosen from the rural population.

The role of self-selection in migration is not only at the heart of the income gap, but also critical for us to understand the income distributions ***within*** these two regions. Consider the rural areas first. If the migrants are randomly chosen from the population, the income distribution of the stayers will remain unchanged under the assumption that the skill price is unchanged. But the shape of it will probably change if the migrants are not randomly chosen under the same assumption. [2] The characteristics of the migrants also have strong influences on the income distributions of urban residents as well. Inequality will increase if the migrants are "negatively" selected.[3] In this article, however, we put more emphasize on the distribution of income in rural

---

[1] The ratio of per captia annual income between those of urban and rural areas is 2.2:1 in 1990, and by 2004, the ratio increased sharply to 3.2:1 (NBS, 2005), and see Li and Yue (2004) for reference. During this period, rural to urban migration also increased. According to MOA (the Ministry of Agriculture) source, the number of rural migrants soared from 2 million in 1982 to 102 million in 2004. (Cai and Wang, 2007)

[2] This is a strong assumption of course. When some rural residents migrate out, the skill price may also change in equilibrium, especially when the skill distribution changed in case of positive selection. But we need this assumption to construct our counterfactual income distribution. This kind of assumption is also used by Dinardo, Fortin, and Lemieux's (1996) and Chiquiar and Hanson (2002).

[3] See Borjas (1987, 1999) for a discussion of self-selection in an international migration context, and for a comprehensive survey.

areas.

Although much of the research have been done to investigate the role of migration on inequality (Li, 1999) and poverty reduction (Wang and Cai, 2006), one basic question has yet to be answered: “how much would a migrant have been paid in her/his home villages were he/she stayed home?” This question is fundamentally very important for our purpose. For example, when we study the influence of migration on income distribution *within* rural areas, we are in essence carrying out a *factual-counterfactual* comparison. The *factual* is just what we observed for rural stayers. But in order to construct the *counterfactual*, we must include migrant workers in our rural sample because they would be working in rural areas if they don’t migrate. And what we are interested in is their counterfactual income instead of their actual income in urban areas. The same logic applies when we study rural-urban income gap. The *factual* is just what we observed for rural stayers and urban workers including migrants, especially permanent migrants. But in order to construct *counterfactual*, we should know the migrants’ income in rural areas if they don’t migrate.

We modify Dinardo, Fortin, and Lemieux’s (1996) framework to construct counterfactual income densities. We begin with rural residents’ actual income density, which corresponds to rural areas’ income structure integrated over the distribution of rural-worker characteristics. Then we re-weight this distribution according to the observed characteristics of all migrants. This simulates the wage distribution that would prevail given migrant characteristics and rural skill prices.

In this paper, we use three coordinated data sets from the 2002 China Household Income Project (henceforth CHIP, see the data section for a brief introduction) to examine who in rural areas migrates to the urban areas and how their observable skills and incomes compare to those who remain at home. Another feature of this article is that we make distinction between two types of migrants: temporary migrants and permanent migrants. We refer permanent migrants to those who successfully got urban hukou and temporary the ones did not. Both descriptive statistics and formal analytical results indicate that this distinction is essential for us to understand the effect of migration on income distribution both within and between rural and urban areas. The permanent migrants are rather selective than temporary migrants. They are more educated, and they migrate to urban areas for both higher level of wages and for higher skill prices. Were they stayed in rural areas, they will fall disproportionately on the medium-to-upper part of the income distributions. These results are consistent with the findings on the changing patterns of China’s income distribution by Li and

Yue (2004) and many others, which indicate that the rural-urban income gap enlarged and that the income distribution within rural areas decreased slightly. We also find an obvious composition change of the permanent migrants.

The paper is organized as follows. The next section provides some institutional backgrounds and a brief literature review. Section 3 introduces and describes our data. Section 4 first gives the methodology we use to construct the counterfactual income distribution and then gives the empirical results. Section 5 investigates the effect of migration on rural income distributions and rural-urban income gap. The last section concludes.

2. Background and Brief Literature Review

Several articles give excellent introduction and analysis on the institutional background of China's rural-urban migration, in particular the *Hukou system* (see for example Zhao (2005), Deng and Gustafsson (2006), Wang and Cai (2006)). Basically, rural migrants can be divided into two groups, those obtained an urban hukou and those did not. Only those obtained an urban hukou are registered officially as urban residents, and the urban hukou is a prerequisite for them to be covered by the urban social security system and to have various forms of welfare and subsidies.[4] Moreover, once they are registered as urban residents, they are no longer rural residents, their land in sending regions will no longer belong to them, and officially they are no longer villagers and have no voting right any more on village affairs. Both casual observation and academic research (Deng and Gustafsson, 2006 for example) indicate that rural migrants who successfully obtain an urban hukou are well integrated in urban society. On the other hand, many rural migrants retain their rural hukou, they still have land and have political right in their village affairs. Although they may spend quite a long time in urban areas, they are not covered by the urban social security system and are not entitled to various subsidies.

Many researches examine the migrant characteristics and compare them with those in the destination and origins. Basically speaking, the rural migrants are more educated than the non-migrants and tend to be younger. Majority of them have junior middle school or primary school education. There are fewer female migrants. Minority nationalities are less likely to migrate. (see Zhao (2005) for a more comprehensive survey). However, most of the existing literature is on temporary

[4] We'll give more detailed institutional background information in the data section.

migrants, those who migrate to urban labor market without obtaining the urban hukou. The others, who successfully obtained the urban hukou, are largely neglected (Deng and Gustafsson (2006) is among a few exceptions). We'll see that these two groups of migrants are different in many aspects from individual characteristics to labor market outcomes. Therefore, neglecting permanent migrants will bias our understanding of the role of migration on income distributions. In this article, we contribute to the literature by looking on a more representative sample of migrants.

Second, due to the methodology and/or the data limitation, most of the above researches focus on means or a particular position (the poor whose incomes are below a certain level, for example) of the income distribution. Many researchers examine the effects of migration on poverty, emphasizing the role of remittance (Li (1999), and Du, et. al (2005) for example).[5] However, the evidence on the effects of migration on the entire income distribution, not just on poverty, is still limited. We still don't know what the shape of income distribution in rural areas would be like if the migrant workers stayed home. Moreover, although we have the attributes of those migrants, we don't know the position of the income distribution where they would fall in if they stayed home. To accomplish the above tasks, we use DFL approach. This approach was introduced by Dinardo, Fortin, and Lemieux (1996), and has been widely used.[6]

It should be noted that the DFL is not the only approach to construct the counterfactuals. Researchers using Oaxaca (1973)'s approach to decompose the income differentials must also be familiar with the "counterfactual" concept. Oaxaca decompositions are based on simple counterfactuals such as "how much would a migrant with the mean characteristics of all migrants have been paid in rural area". What distinguish DFL approach to Oaxaca approach is that the latter focus on means alone and the former works with the entire distribution. Only focusing on means will conceal much information, and it is possible that the means remain unchanged while the distribution changed. DFL approach can give us information exactly on where the income distribution was influenced most by migration. Another notable alternative is the method proposed by Machado and Mata (2005), which can also be used to

[5] Du, et. al (2005), find that households with migrant(s) have higher household income per capita by about 8.5 to 13.1 percent on average compared with those without migrants, but the impact on poverty is small because most of the poor do not migrate. They also find that migrants remit a large share of incomes and are somewhat responsive to the needs of other family members. In a somewhat earlier research, Li (1999) finds that remittance play a significant role in the per capita income of migrant households.

[6] For example, Lemieux (2006) extended the DFL to decompose the change of residual inequality into price effect and composition effect. Chiquiar and Hanson (2002) analyzed the self-select effect of migration from Mexico to the U.S.

construct the whole counterfactual distribution. This method is based on "parametric model for the quantiles of the conditional distribution." To construct the counterfactuals, they "resort to resampling procedures to obtain a marginal distribution consistent with both the conditional model and the covariate densities." Clearly, because of resorting to a parametric model, the MM approach is necessarily restrictive. The resampling procedures may also be quite cumbersome. In contrast, the DFL approach is based on non-parametric kernel-density method. It's not as restrictive, and it's easy to implement.

In fact, these two approaches are conceptually similar. Researches using these two approaches haven't found that either one is superior to the other. One notable research is Autor, Katz, and Kearny (2005). They extended the MM approach to investigate composition effect and price effect in the change of wage inequality. To compare their results to Lemieux (2005), they find its substantive differences with the latter are *not* consequential for their conclusions (They do draw different conclusions, but not because of different methods). But in our context, the DFL re-weighting method is more readily used to calculate counterfactual inequality, and it's easy to modify this method to account for labor participation differential. Therefore, we use the DFL approach only.

Finally, self-selection in the process of migration and its effects on income distributions are well studied (especially for the US). The foci are on the effect of immigration on inequality within US (see Borjas, 1987, 1999). The analytical framework of self-selection largely based on Roy's (1951) model developed in the international migration literature is very instructive for Chinese studies. As pointed out by Foster and Rosenzweig (2007), "the selectivity of the process in terms of the human capital of those who leave the agricultural sector" is still less studied for developing countries. This paper adds empirical results to this literature.

3. Data and summary statistics

The data comes from the 2002 China Household Income Survey (CHIP) which was conducted by the Chinese Academy of Social Sciences in joint work with the National Bureau of Statistics (NBS) in the beginning of 2003. There are three coordinated datasets, namely, urban household survey, rural household survey, and urban migrants survey.

The urban household survey covers 2 municipalities and 10 provinces. It collects information for 6835 households in 77 cities. The rural survey covers rural

households in 22 provincial level units comprising 122 counties or towns from which 9200 households were selected. It should be noted that the sample frame for both the rural and urban sample is based on registers of people possessing a local *hukou*. Meanwhile, both the urban and rural samples were drawn from larger samples regularly used by NBS to produce official statistics for China. Therefore, the data we use is representative of China (see Deng and Gustafsson, 2006, and Li, et al, 2006). It should also be noted that most of the information we use is at individual level. There are 20632 urban individuals and 37969 rural individuals, 35.21% and 64.79% of the whole sample respectively. The percentage is very close to the figures given by NBS (2005), 39.09% and 60.91%. [7]

3.1 Two samples of temporary migrants

For the study of rural-to-urban migration, a migrant survey was also conducted. It covers 5327 individuals in 2000 migrant households living in the same municipalities and provinces as the survey of urban residents. However, this data set has at least two shortcomings for our study. First, it doesn't include permanent migrants as all the individuals have rural hukou. Second, the sample may be more representative of temporary migrants with longer migration duration. This representativeness bias is due to the sampling process of the CHIP migrant survey. As Deng and Gustafsson (2006) commented on this data set, "Short term rural migrants living in dormitories and/or at urban work places are rather difficult to identify, while rural migrants with housing conditions similar to the registered urban population can be simpler to identify and sample, our sample of temporary migrants has this character".

To overcome the second shortcoming, we identify another set of temporary migrants using the Rural Household Survey. In the 2002 rural survey, there're around 3800 rural residents earn wages outside their home countryside out of 37969 rural residents. These migrants are also called temporary migrants in this paper. However, this temporary migrants sample also has representativeness problems. First, it's difficult to tell whether these migrants are to other rural areas or to urban areas. Second, the migrants sample from the rural household survey may be under

[7] The fact that the rural and urban samples are representative of the national population is import for our analysis when constructing the counterfactual wage densities. We are lucky that our sample seems representative. To be assure that our results are robust to this problem, we also do exercises to adjust our sample size. We drop observations randomly from the urban survey, so that the shares are exactly as given by NBS. The results didn't change much.

representative of the migrants who move out for a longer period of time and those who migrate out with the whole family.

Fortunately, although both these two temporary migrant samples have representativeness problems, the biases are in opposite directions. Therefore, they can complement each other. As these two datasets are not necessarily mutually exclusive, we consider these two temporary migrants samples separately. And for convenience, we sometimes called the temporary migrants from the urban migrants survey *type I migrants*, and those from the rural household survey *type II migrants*. We'll compare the characteristics and income distributions of the temporary migrants with other groups later on. Next, we construct our sample of permanent migrants (*type III migrants*).

3.2 Identifying permanent migrants

In the 2002 urban survey, the urban residents were asked "when did you get the urban hukou?" Those who answered this question with a specific year were former rural residents, and successfully obtained urban hukou. Therefore, this piece of information helps us to identify the permanent migrants. From Table 1, we can see that around 20% urban residents were not born with urban *hukou*, among which over one half obtained their urban *hukou* within the last two decades, which is consistent with the fact that the hukou system has become less restrictive in the post-reform period.

(Insert table 1 around here)

It's also helpful for us to investigate the ways by which the former rural residents obtained their urban hukou. We can categorize the different ways into two groups: those obtaining urban hukou through formal ways (through education, through being a cadre,[8] or through joining PLA), and those through informal ways (through buying house or through "losing land"[9]). Rural residents, whose lands are occupied by urban construction projects are often given urban hukou. Rural households who can afford buying houses in urban areas are also sometimes given urban hukou. It should be noted that, the CHIP data are poor at collecting this information. Nearly half of the permanent migrants didn't report the specific ways how they obtained the urban

[8] The "cadre" here refers to those who are in charge of some administrative affairs in villages or towns. They are generally rural residents. But these people can be promoted. Once they are promoted, they have chances of getting urban hukou.

[9] Those obtained their urban hukou simply because their land are occupied do not necessarily move. We categorize them as permanent migrants in the sense of hukou status transformation. We'll come back to this problem later.

hukou (*other* or *missing*, see Table 1). Consider only those who did report. It can be shown that majority of them got urban hukou through formal education. Joining the PLA is the second largest channel. The last two informal channels take up a relatively small share of the permanent migrant population.

The above different channels played different roles in different time of periods. Three aspects are noteworthy. First, higher education plays a predominant role in most of the time periods. For those obtained urban hukou in the 1980s in particular, above 30% of the them were through the way of education. Even during the 1960s, when joining PLA is the major channel of getting urban hukou, there're still more than 18% of the permanent migrants obtaining their hukou by education. Second, in the post-reform periods, neither being a cadre nor joining PLA are playing a significant role in the permanent migration process. Third, there are some new trends in the 1990s. It seems that more and more rural residents buying houses in urban areas. Many local governments offer them urban hukou in order to encourage rural residents to buy houses and therefore raise government revenue. There also seems to be more and more rural residents get their urban hukou after their lands are occupied. These two groups enlarged tremendously in the 1990s. In the 1970s, only less than 5% of the migrants get their urban hukou through these two channels, in the 1990s however, the share increased to more than 20%.

The characteristics of the permanent migrants are closely related to how and when they obtained urban hukou (see Table A 1 in the appendix). There's no doubt that those who obtained their urban hukou through formal education have the highest years of schooling. Those through being a cadre or through joining PLA have relatively fewer years of schooling. Migrants with the lowest level of education are those through informal ways. Meanwhile, the average years of schooling increased steadily from 7.4 for the "-1950" migration cohort to 11.4 for the "1980-90" migration cohort. This increasing trend reflects partly the fact of education expansion. Interestingly, this trend stopped for the "1990-" cohort. This was probably caused by the composition change of permanent migrants as mentioned above.

Consider age structure. Those who obtained hukou through education, through being a cadre, and through joining PLA were relatively young when they migrated. The ages of those who obtain hukou through losing land, or through buying house are not only higher but also with more dispersions. In 2002 (when the survey was conducted), the subgroups with the highest average ages are "cadre" and "army" (53.7 and 53.0 respectively), and that with the lowest average age is "land". This contrast

reflects the fact that most of those belonging to "cadre" and "army" groups obtained their hukou before 1980, while "losing land" is a rather recent phenomenon.

These simple description statistics reminds us that, even within the group, permanent migrants themselves are not homogenous. It's obvious that permanent migration is driven by different forces and under different institutional constraints. This largely reflects China's feature of development and transition. Therefore, it's important to investigate the cohort effects in the following analysis. We thus divide our samples into different groups in the main empirical part according to the year when the migrants obtained their urban hukou. To account for the effects of some historical events in the transitional process, we do not divide the sample by decades as in this section but by some critical years. Those who obtained their hukou after 1992 are termed as recent migrants.[10] The recent migrants are of more concern because they are more likely to predict the consequences of permanent migration. But it should be kept in mind that, even the recent migrants are not alike. We will come back to this later.

It's also worth noting that, the processes of permanent migration are selective. Those with high level of education, high income level, high ability or those locating in the suburban area are more likely to be selected (or self select) to get urban hukou. This process of selection can happen either at rural origins or at urban destination. It's possible that people first choose to migrate without urban hukou, and successfully get urban hukou later. Therefore, the distinction between permanent migrants and temporary migrants are not clear-cut, especially when there're a large proportion of permanent migrants who didn't report how they get urban hukou. This shortcoming is not fatal however. First, a large proportion of permanent migrants get urban hukou through education, joining army, etc. In these cases, the decision is probably made before they migrate. Second, it is extremely difficult for temporary migrants to get urban hukou until very recently. For most temporary migrants, if not through particular official channels (mentioned above) to go to urban areas, they are doomed to keep their rural hukou. Finally, our focus in this paper is to see the self-selection effect of permanent migrants and temporary migrants. Our conclusions are not conditional on when and where did the migrants get urban hukou.[11]

---

[10] We choose 1992 because the 14th CCPC take place in this year. "Establishing socialist market economy" is set up the goal of China's economic reform.

[11] For example, people may only select to migrant and some exogenous factors (such as luck) decide whether they can successful get an urban hukou. If this is the case, permanent migrants' counterfactual income distribution will be identical to that of temporary migrants (see the methodological part). We thank the referee for raising this point.

3.3 Rural residents, urban residents, and migrants

Characteristics

In the following analysis, we keep individuals aged 18 to 60. Those in school are also dropped. The shares of rural and urban sample remain almost unchanged. In order to avoid the endogeneity problem of education, we also dropped permanent migrants who obtained their urban hukou under the age of 16. Next, we compare four subsamples, with the first two (the urban natives and permanent migrants) having urban hukou and the last two (rural residents and temporary migrants in the urban survey)[12] having rural hukou.

Table 2 reports the simple statistics for men. We compare the urban natives and the rural residents at first. These two groups are distinct in almost every aspects we considered, from education level to labor participation. First, the urban natives are more educated than rural residents, with the average years of schooling being 11.1 and 7.8, respectively. As for education levels, more than half of the rural residents only have middle school education, and nearly one quarter of the population have only primary (or below) education. In contrast, nearly 70% of the urban residents have above high school (inclusive) education. The fraction of population with below primary (inclusive) education is only 3%. In terms of age, ethnicity, and political status, the urban natives are older than the rural residents, and they are less likely to be minorities, and more likely to have party membership. Finally, we calculate the shares of people who have wage income in 2002, and we use these as a crude indicator of labor participation. For urban natives, the share is around 81%, while for rural residents, it is only 54%.

Although permanent migrants are similar to urban natives, and type I migrants are similar to rural residents, there are dissimilarities. As for type I migrants, they are slightly more educated than rural residents, and they are younger. They have a lower share of party memberships (5% as opposed to 13% for rural residents), and a much higher labor participation rate (95% as opposed to 54% for rural residents). For the permanent migrants, the fraction of the sample with above college education is more than 40%, which is even much higher than that of urban natives (29%). Similarly, the share of party members for the permanent migrant sample is also much higher than

[12] The rural residents also include some temporary migrants (type II migrants). We'll separate it out in the following analysis.

that for the native sample. The dissimilarities are to a larger extent when compare the permanent migrants with rural residents. All these indicate that permanent migrants are positively selected, and the selection effect is larger than type I migrants.

(Insert table 2a and 2b around here)

Table 2b reports the statistics for female. The contrasts between rural residents and urban residents are similar. Urban female residents have higher levels of education, they are less likely to be minorities, and more likely to be party members and wage earners. When compare to men, the statistics are consistent with most of the literature and with our casual observation. Female not only have lower level of education than men, but also have lower labor participation rate both in political and economic activities. This is true for both rural and urban residents, but varies in extent. The gender gap seems to be larger in rural areas than in urban areas. Within the urban residents, however, the differential patterns between natives and migrants female are not similar to those of males. Both the migrants' education level and labor participation rate are slightly lower than their urban native counterparts, and their likelihood of being party members are almost the same as natives.

It's obvious that labor participation rates as defined by the shares of wage earners vary considerably among different samples. In the following part, we rely heavily on the sample of wage earners. When we come to rural wage earners in the rural household survey, we can divide them further into two subgroups, local wage earners and temporary migrants (type II migrants). In Table 3a and Table 3b, we give summary statistics for wage earners for male and female respectively. Generally speaking, wage earners have slightly higher level of education than the whole sample, and this is especially true for female. And the contrasts between urban residents and rural residents, between urban natives and permanent migrants, and between male and female are similar to what we have discussed for all residents. As the labor participation rate is relatively high for urban residents and type I migrants, the statistics for wage earners are very similar to those in Table 2a and Table 2b. Therefore, we compare mainly two subgroups of wage earners, namely type II temporary migrants and rural local wage earners.

(Insert table 3a and 3b around here)

As for education, it's clear that type II migrants have more years of schooling on average for both male and female, but only marginally. When we break down the years of schooling into four education levels, majority of the wage earners are junior middle school graduates. The fraction, however, is much higher for type II migrant

workers than for local wage earners (0.63 and 0.54 respectively for men, 0.60 and 0.45 respectively for female). As for other levels of education, rural local wage earners have both larger share of primary education and larger share of high school graduates.

As for age, the male (female) type II migrants are 32 (26) years old on average, almost 10 year younger than their counterparts of local wage earners. It's also clear that female wage earners are much younger than their male counterparts on average. This is consistent with most of the literature. As for party membership, the fraction is much higher for local wage earners than for temporary migrants. And this is true for both genders.

We also calculate the working days per year and working hours per day for all the subgroups. Urban residents, both natives and migrants have more days of working than rural local wage earners and type II migrants, but less than type I migrants. And the rural local wage earners have the least working days. As for working hours per day, the temporary migrants, especially type I, work most.

The last thing we want to mention in this subsection is that, permanent migrants and temporary migrants are very distinct in almost every aspect. The previous research that emphasizes only the latter may bias our understanding. Next, we turn to the differentials of their economic outcomes (hour wages and per capita income).

Income distribution and income differentials

Figure 1gives the hour wage densities for rural local workers, urban native workers, and three types of migrants (upper-left and upper-middle panels for male and female respectively). It's clear that urban workers have much higher level of wages than rural local workers. The three types of migrants have very different wage distributions. For type I migrants, although they are sampled from the same areas as urban workers, their wage distribution is very close to that of rural but not urban workers. Type II migrants have even lower wage levels than rural workers. On the contrary, the wage distribution of permanent migrants almost overlaps with that of urban natives workers. All these indicate remarkable heterogeneity among different types of migrants. Even within permanent migrants, there is obvious heterogeneity. As is shown by the lower two panels of Figure 1, permanent migrants of different migration cohorts have very different wage distributions. The more recent migrants not only have lower wage levels, but have more wage dispersions. The above descriptions are true both for male and female.

Of course, the above description may be problematic, because wage earners are only parts of the whole labor force and there is larger proportion of the labor force that did not have wage income in rural areas. Thus wage density cannot fully reflect the rural-urban income gap. To overcome this, we use per capita annual income densities instead of those of wages (see upper-right panel of Figure 1). [13] The patterns are similar. One noteworthy dissimilarity is the distribution for type I migrants. It's more close to that of urban residents and permanent migrants instead of rural residents. This is partly because type I migrants have large number of working days and long working hours per day.

(Insert Figure 1 around here)

We also calculate the means and various measures of inequality for per capita income of different groups (see Table 4). The first row of Table 4 reports average per capita incomes for various groups. The first two columns are for rural residents and urban residents. It's straightforward to calculate the urban-rural income ratio, which is 3.01:1 (8174:2715). The urban residents are composed of two parts: urban natives and permanent migrants. The average incomes of these two subgroups are 8290 and 7678 Yuan. One group that is neither included as rural residents, nor as urban residents is type I migrant. Their average income is 6552 Yuan. The following rows report various inequality measures. The rural residents have the highest level of inequality, Gini coefficient being 0.3683. The type I migrants have the second highest Gini coefficient, 0.3484. The Gini coefficient for urban residents and for urban natives and permanent migrants are similar and are among the lowest (less than 0.33).

What would the urban-rural income gap and the income distributions would be like if there were not migrants? One simple but also naïve exercise is to separate the migrants samples from the urban residents sample, merge them with rural residents, and recalculate various inequality measures directly. The results are reported in columns 6, 7, and 8 in Table 4. As both type I migrants and permanent migrants have higher levels of per capita income than rural residents, it's not surprising to see that both rural income levels and inequality measures increase.

We call this exercise naïve because the rural residents, urban residents, and migrants face different skill prices. If migrants didn't migrate, they would not have

[13] In the 2002 urban survey, each working individual is asked "what's your total income in 2002?" The income can take various forms including basic wages, bonus, subsidies, ect. We sum all the incomes within a household and divide it by total numbers of individuals. The income data for rural households are more straightforward. They were asked what's their net income in 2002 on household basis, we divide the net income by family size to get the per capita income.

income levels as they are having as migrants. Therefore, we need to answer one basic question: what would the income distribution have been like if migrants were paid as in rural areas or as rural residents. We'll come back to Table 4 in section 5.

(Insert table 4 around here)

4. Constructing Counterfactual Income Distribution

In order to construct the counterfactual income distribution, we first discuss the methodology and then give the empirical results.

4.1 Methodology[14]

Let $f^{i}(w|x)$ be the density of income w in region i, conditional on a set of observed characteristics x. i takes two possible values, *Rural* and *Urban*. Therefore, differences in $f^{Rural}(w|x)$ and $f^{Urban}(w|x)$ capture differences in skill prices in the two areas. One way to see the difference is to estimate wage (income) equations for various subgroups. From Table A 2 in the appendix, we can see that urban residents (both natives and permanent migrants) have higher returns to education than rural residents and temporary migrants. The returns to experience, political status, and ethnicity status are also different.[15] Next, we define $h(x|i=Rural)$ as the density of observed characteristics of individuals in rural areas, and $h(x|i=Urban)$ as the density of observed characteristics among migrants in urban areas. Differences in $h(x|i=Rural)$ and $h(x|i=Urban)$ capture differences in the distributions of observed characteristics for migrant workers and for rural stayers. The observed density of income for individuals working in rural area is

$$g(w|i=Rural)=\int f^{Rural}(w|x)h(x|i=Rural)dx$$

Likewise, the observed density of income for migrants is

$$g(w|i=Urban)=\int f^{Urban}(w|x)h(x|i=Urban)dx$$

[14] The following part borrows heavily from Dinardo, Fortin, and Lemieux's (1996) and Chiquiar and Hanson (2002).

[15] A more complete description of the conditional distribution is to estimate quantile regressions. But this is not indispensible for the DFL approach. We think OLS estimation of wage equations is sufficient to illustrate that the skill prices are different between urban and rural areas.

Consider the density of wages that would prevail for migrant workers in urban areas if they were paid according to the prices of skills in rural areas:

$$g_{Urban}^{Rural}(w) = \int f^{Rural}(w \mid x)\, h(x \mid i = Urban)\, dx$$

This corresponds to the distribution of wages for rural stayers, except that it is integrated over the skill distribution for migrants in urban areas. While this distribution is unobserved, we can rewrite it as

$$\begin{aligned} g_{Urban}^{Rural}(w) &= \int f^{Rural}(w \mid x)\, h(x \mid i = Urban) \frac{h(x \mid i = Rural)}{h(x \mid i = Rural)} dx \\ &= \int \theta f^{Rural}(w \mid x)\, h(x \mid i = Rural)\, dx \end{aligned}$$

where

$$\theta = \frac{h(x \mid i = Urban)}{h(x \mid i = Rural)}$$

The key insight of Dinardo, Fortin, and Lemieux's (1996) is that a counterfactual density can be estimated by taking an observed density (e.g. for individuals in rural areas (stayers)) and re-weighting it (e.g., to reflect the distribution of characteristics of rural migrant workers). To compute the weight, use Bayes's Law to write,

$$h(x) = \frac{h(x \mid i = Urban) \Pr(i = Urban)}{\Pr(i = Urban \mid x)}$$

and

$$h(x) = \frac{h(x \mid i = Rural) \Pr(i = Rural)}{\Pr(i = Rural \mid x)}$$

Combining the above two equations we can obtain an expression for θ that is a function of the ratio of the probability that a rural-born individual works in rural areas (conditional on x) to the probability that a rural-born individual works in urban areas (conditional on x).

$$\theta^{M} = \frac{h(x \mid i = Urban)}{h(x \mid i = Rural)} = \frac{\Pr(i = Rural)}{\Pr(i = Urban)} \cdot \frac{\Pr(i = Urban \mid x)}{\Pr(i = Rural \mid x)}$$

Note that the first ratio $\Pr(i = Rural) / \Pr(i = Urban)$ is a constant given by the sample proportions of rural migrants and rural stayers. The second needs to be estimated, and Dinardo, Fortin, and Lemieux's (1996) suggest estimating these probabilities parametrically, using the estimates to calculate θ, and then applying the estimated θs

to standard kernel density estimators to obtain a counterfactual income density:

$$\hat{g}(w) = \sum_{j=1}^{n} \frac{\hat{\theta}_j}{h} K\left(\frac{w - W_j}{h}\right).$$

Therefore, by using the income data of rural local workers and the predicted values of θ, we can have the counterfactual income densities of migrant workers. And it's straight forward to calculate the counterfactual income inequalities (as we already pointed out in the literature section). Suppose the share of permanent migrants and rural residents are *s* and (*1-s*), and suppose all the estimated $\hat{\theta}$s for rural residents sum up to 1.[16] Then the counterfactual Gini coefficient can be estimated normally using the reweighted sample of rural residents. The weight for each observation j is $(1-s)/N + s \cdot \hat{\theta}_j$, where N refers to the number of observations of rural residents.

Before going to the empirical results part, we make two clarifications here. First, the income w can take various forms. Our first choice is log of hourly wages. But not everyone has wages, and those who have wages are not randomly drawn from the whole population. Therefore we can also think of w as per capita income at the household level. We do carry out some exercises in the following part by using per capita income to circumvent the sample selection issue of using wage data. At the moment, we just ignore the labor participation issue for simplicity, and we'll come back to this issue at the later part of this subsection. Second, the definition of region *i* is also worth more clarification. The two values (*Rural* and *Urban*) have different interpretations in different exercises. When we are using wage data, *Rural* refers to rural local workers, and *Urban* refers to different types of migrants such as permanent migrants or the two types of temporary migrants. This means that while we are studying the selection nature of different migrant samples, the sample of rural local workers is always being used as the reference group. When we are using per capita income, *Rural* refers to all labor force of rural residents in the rural household survey. The type II migrants are also included in this reference group because they also contribute to their rural household income. Therefore, *Urban* only refers to permanent migrants or type I temporary migrants.

---

[16] All the estimated $\hat{\theta}$s for rural residents do not necessarily sum up to 1. If not, we should first do some adjustment before using $(1-s)/N + s \cdot \hat{\theta}_j$ as weight. And when there are more than two groups, the extension is straight forward.

4.2 Basic Empirical Results

We apply the above method to our combined sample of rural local workers and permanent migrants first. To construct the counterfactual wage densities for permanent migrants, we estimate a logit model for $\Pr(i = Rural \mid x)$ using this combined sample. The predicted probability can be used to calculate the weight, $\left[1 - \Pr(i = Rural \mid x)\right] / \Pr(i = Rural \mid x)$. This can in turn be used to construct the counterfactual wage densities of permanent migrants. The variables we use include years of schooling, experience, experience squared, minority dummy, and party membership dummy. We report the predicted probability of being rural residents by gender, by party membership, by age, and by education in Figure 2. Clearly, education plays a very important role. For those with junior middle school education and below, the probability of permanent migration is almost negligible. This is especially true for the young and for males. Those with above college (inclusive) education have the highest probability of permanent migration. And this probability increases with age. Besides, party members have higher probability of permanent migration, female also have higher probability of permanent migration when holding others variables (age, party membership, education, etc) constant.

(Insert figure 2 here)

The most important empirical results of this paper are reported in Figure 3 for male (upper panels) and female (lower panels) respectively. The dark solid lines in the left and middle panels of Figure 3 are the estimated kernel wage densities of permanent migrants (left panel) and rural local workers (middle panel) respectively. The two dashed lines in the two panels for each gender are identical in fact. They're the counterfactual kernel densities for permanent migrants, assuming they are paid according rural skill prices. All these above estimates are based on a Gaussian kernel function, and the number of bins is set at 200.

It's clear that, the counterfactual wage densities for permanent migrants drift to the left to its actual densities. This is consistent with the fact that the skill prices are relatively lower in rural areas (or for rural residents). However, even permanent migrants were paid according to rural skill prices, the counterfactual wage densities and the actual wage densities of rural residents are far from close (see the upper middle panel). This means that permanent migrants are drawn from the rural residents disproportionately. For wages up to somewhere slightly over the peak of rural residents' wage density, the counterfactual wage density exhibits a smaller mass, and

it exhibits a larger mass after that point. It is not until the extremely high wages are reached that the two densities coincide, and for high levels of wages, the counterfactual density of permanent migrants seldom has smaller mass than the actual density of rural residents.

(Insert figure 3 here)

For convenience of comparison, the difference between the wage densities of rural residents and the counterfactual wage densities of permanent migrants is shown in the right panel. It appears that the people with the highest probability of becoming urban residents would have above-medium wages in rural areas, and those with the least probability would have medium-to-low wages. Were permanent migrant male in urban areas to return to rural areas and be paid according to rural skill prices, they would tend to fall disproportionately in the upper half of rural wage distribution. All these results are consistent with the hypothesis of "positive selection". This conclusion also applies to female. It appears that the discrepancy between the actual density of rural female residents and the counterfactual density of female permanent migrants are even larger than that for males.

All these results suggest that migration to urban areas permanently, by taking mass away from upper half of rural wage distribution, may reduce wage inequality in rural areas and increase income gap between rural and urban areas.

4.3 Robustness Check

By Age Cohorts and by Migration Cohorts

For robustness check, and to account for the fact that China experienced dramatic transition and development, we carry out several other exercises. One natural exercise is to divide the combined sample by age group. We have three age cohorts both for male and female, namely 18-30, 31-45, and 46-60. Figure 4 report the results. Instead of reporting the actual densities of rural residents and counterfactual densities of permanent migrants, we just report their differences. Before going into details, one conclusion can be made in advance. If the permanent migrants returned to rural areas and be paid as rural residents, their wages will fall disproportionately in the upper portion of rural income distribution. The interesting aspect is that the extent of disproportion varies with age.

We see the results for male first. From upper-left panel of Figure 4, we can see that the counterfactual densities for permanent migrants become more and more

closer to the actual wage densities for rural residents as the age increases. This means that for above-medium values in the residents' actual wage distribution, the mass of counterfactual density are larger than that of rural actual density to a larger extent for younger cohort than for older cohort. And for middle and lower wage values, the actual densities are above the counterfactual density to a larger extent for younger cohort also. As for female (upper-right panel), such trend is not as clear as for men, but this may be caused by the fact that the samples for female become very small.

The results for the current young cohorts are more supportive of the positive selection (at least that of men). This is consistent with the fact that the young individuals in the permanent migrant sample are mainly composed of those obtaining their urban hukou by education, while those older include not only those obtaining their urban hukou through education long before but those through buying house and/or "losing land". We can also interpret the results as that the positive selection effect disappears as people age. As these two explanations have distinct policy implications, we do the following exercise to determine which one is more feasible.

(Insert figure 4 here)

As we have mentioned in the previous sections, the process of permanent migration is influenced by different institutional factors in different time of periods. Another exercise we therefore carry out is to divide the sample into four "migration cohorts", namely those who obtained their urban hukou before 1978, those between year 1979 to 1992, those between 1993 and 2002, and those between 1998 and 2002. We investigate these four cohorts because 1978, 1992, and 1997 are three critical years in China's transitional process. It's clear that, positive selections remain obvious as the counterfactual densities are still above the actual densities for upper portions of wage distribution, and below for middle and lower portions (see lower panels in Figure 4). However, the positive selection effects seem to be more obvious for the 1978-1992 migration cohorts both for male and for female, because the distances between the zero horizontal line and the difference lines are the largest for these cohorts at most of the wage values (except for the -1978 cohorts). The difference lines for the 1993-2002, and 1998-2002 cohorts lie between those for the 1979-1992 cohorts and the zero line. The difference lines for the 1998-2002 cohorts are very close to the zero line in fact, indicating that the actual wage densities for rural residents are very close to the counterfactuals (especially for men).

All the results may indicate that, with the deepening of Chinese labor market reform, the process of permanent migration is becoming less and less selective. Of

course, this may also be due to the fact that during the later period, many rural residents' lands are occupied and therefore they obtained their urban hukou. Such kind of permanent migration processes are neither based on personal characteristics such as education level, nor based on socioeconomic status such as income or wealth level, but simply because they reside near city.

Accounting for Labor Participation Differentials

One major drawback of the above analysis is the neglect of the differentials in labor participation rates between urban residents and rural residents. In this subsection, we adopt the approach proposed by Chiquiar and Hanson (2002) to account for the fact that labor participation rates in rural and urban areas are very different from each other. The counterfactual density can be estimated by first taking the observed density for rural wage earners, and then reweighting it to reflect the not only differences in characteristics, but also differences in labor participation rates. Therefore, we need to modify our weight as follows:

$$\theta = \theta^P \theta^M = \frac{\Pr\left(i = Rural, D_i = 1\right)}{\Pr\left(i = Urban, D_i = 1\right)} \cdot \frac{\Pr\left(D_i = 1 \mid i = Urban, x\right)}{\Pr\left(D_i = 1 \mid i = Rural, x\right)} \cdot \frac{\Pr\left(i = Urban \mid x\right)}{\Pr\left(i = Rural \mid x\right)}$$

D_i is a dummy variable indicating whether the individual has wage jobs (Yes=1/otherwise=0). The first term in the r.h.s is simply some constant determined by the sample, and thus does not matter for our analysis. The last term is exactly what we estimated in the previous sections. The new term is in the middle,

$$\theta^P = \Pr\left(D_i = 1 \mid i = Urban, x\right) / \Pr\left(D_i = 1 \mid i = Rural, x\right)$$

This is used to account for the labor participation differentials.[17] We estimate logit models for *Rural* sample and *Urban* sample respectively, and use the predicted probabilities to construct $\hat{\theta}^P$, then we use $\hat{\theta} = \hat{\theta}^P \hat{\theta}^M$ as the new weight to get the counterfactual wage densities for permanent migrants.

The results by age cohorts and by migration cohorts are reported in Figure 5. The results are still supportive of positive selection. When we compare these results with those without considering labor participation, the distances between the difference lines and the zero line seem to be smaller. The patterns are still similar, however. As for results by age groups for men, the difference line for the 18-30 cohorts is the most

[17] For more details, please see Chiquiar and Hanson (2002).

distant one to the zero line for large range of the wage values, and the one for 46-60 cohorts is the closest. The result for female seems similar, but not as clear and clean as that for men.

As for results by migration cohorts, the results for the earlier migration cohorts are more supportive of positive selection. The more recent of the year of migration, the closer the difference line to the zero line. This is true both for male and female, but especially for men. As for the 1997-2002 migration cohorts, the difference line for male is very close to the zero line. This means permanent migrants would have very similar wage distributions as their counterpart rural stayers if they returned home and being paid as rural residents. For female however, and being paid as rural residents. For female however, the permanent migration process seems more selective even for the 1997-2002 cohorts (see lower-left panel of Figure 5).

(Insert figure 5 here)

Endogeneity of Education and Other Considerations

In order to circumvent the endogeneity problem of education, we drop all the individuals older than 16 years when obtaining their urban hukou in the above analysis. This practice, however are still not out of question. As have been mentioned above, formal education is a major channel through which rural residents get urban hukou. What's more, those who get their urban hukou through education almost invariably complete their final education in urban areas. The "positive selection" findings above may therefore simply reflect the fact that education is a channel of permanent migration. To alleviate this concern, we do another exercise by simply dropping the "education" permanent migration category. To avoid the problem of sample selection, we restrict this analysis to those aged 25-60. The results for male and female are reported in the lower-middle and lower-right panels of Figure 5, respectively. The black solid line depicts the difference between the actual rural wage distribution and the counterfactual wage densities for permanent migrants for the population aged 25-60. The pattern is similar to what we find in previous sections. The thick grey line depict the results after we drop the education category of permanent migration, the conclusion of "positive selection" is still valid, but to a lesser extent. This is true for both male and female. This means that even if there is not the formal education, through which rural residents get urban hukou, the process of permanent migration is still positive selective.

Another exercise we carried out is to drop the "losing land" category as well as

“education” category. The results are depicted by the dotted lines. They almost overlapped with the grey line, which means that our results are not sensitive to the exclusion of this category. This result is consistent with the fact that these “permanent migrants” obtain their urban hukou not because they self-select to, but forced to leave their land.

4.4 Counterfactual Wage Densities of Temporary Migrants

In this subsection, we investigate a slightly different but similar question. What would the wage distributions for the temporary migrants be like if they do not migrate and participate in wage jobs and earn wages as local wage earners.

The basic results for male and female are in Figure 6. For convenience of comparison, we put the three difference lines together, each representing the difference between the actual wage densities of rural local workers and the counterfactual densities of different types of migrants. One distinct feature is that difference lines for both types of temporary migrants are much closer to the zero line than that for permanent migrants. Another difference is that, for wages up to some point of the whole distribution, the difference line for each type of temporary migrants lies below the zero line, and after that point, the difference line lies above the zero line. This means that both types of temporary migrants would fall disproportionately in the medium-to-lower portion of the rural wage distributions if they stayed home, and those local wage earners would fall disproportionately in the medium and upper parts. This is supportive of “negative selection” as opposed to “positive selection” in previous sections. But as is shown in Figure 6, the negative selection effect of temporary migration is not as obvious as the positive selection effects of permanent migration, especially for male and for type I temporary migrants. We also report the difference lines for both types of temporary migrants by gender and by age cohorts (middle and right panels). The results are with more variability and are not consistent. What’s consistent is that distances between the difference lines and the zero horizontal line are close.

What we should keep in mind is that, we are using the rural local workers as our reference group. As the reference group is not randomly selected from the total rural population, the negative selection effect of temporary migration does not necessarily mean the temporary migrants are negatively selected from the total rural population. The selection effect of permanent migrants also suffers such problem. However, the positive selection effect is most probably reinforced as the reference group is also positively selected as is clear from the descriptive section.

(Insert figure 6 here)

5. The effects of migration on income distribution

Now we go back to the question raised in section 3，what would the urban-rural income gap and the inequality of rural areas be like if there were not migrants? To overcome the shortcoming of using wage data, and for the calculation of inequality measures, we use per capita income in this section.

We first construct the counterfactual densities for migrants, and see their selection effects. The reference group is labor force from the rural household survey. As type II temporary migrants also contribute to household income and it's difficult to separate them out, they are included in the reference group. The counterfactual densities and difference lines for type I migrants and permanent migrants are reported in Figure 7, respectively. The results are similar with those in the former sections although we use different income measures: the permanent migrants are positively selected and there is not obvious self-selection effect for type I temporary migrants.

(Insert figure 7 here)

To see the selection effects' implications for income distributions, we calculate the counterfactual income levels and inequality measures. Column 9 of Table 4 indicates that if permanent migrants returned to rural areas and be paid as rural residents, the per capita income of rural areas would increase 4% percent from 2715 to 2814 Yuan. If type I migrants returned, however, the per capita income level would remain unchanged (column 10). If both types of migrants retuned, the income level would become 2798 yuan (column 11). As permanent migrants are counted as urban residents when calculating factual income levels, we should drop them from the urban samples. Because permanent migrants have relative lower per capita income (7678 yuan) than urban natives, this makes the urban per capita income increase from 8174 to 8290 yuan. Therefore, if permanent migrants returned, the ratio of urban-rural income would fall from 3.01:1 (8174:2715) to 2.95:1 (8290:2814); if type I migrants returned, the ratio will not change; and if both types returned, the ratio would become 2.96:1. The counterfactual gap in terms of absolute value does not decrease however.

How about the inequality levels of rural areas if migrants returned? Take the Gini coefficient for example. If permanent migrants returned, the Gini coefficient would increase from 0.3683 to 0.3755; if the type I migrants returned, the Gini coefficient would remain unchanged; and if both types returned, the Gini coefficient would also

increase to 0.3740. Although the counterfactual inequality measures would not change much from the factual ones, the directions of changes are consistent with the selection effects of different types of migration.

6. Conclusions

As massive rural residents leave their home countryside for better employment, migration will have profound effects on income distributions such as rural-urban income gap and inequalities within rural or urban areas. The nature of the effects depends crucially on who are migrating and their migrating patterns. In this paper, we emphasize two facts. First, rural residents are not homogeneous, they self-select to migrate or not. Second, there are significant differences between migrants who successfully transformed their hukou status (permanent migrants) and those did not (temporary migrants). Using three coordinated CHIP data sets in 2002, we find that permanent migrants are positively selected from rural population especially in terms of education. As permanent migration takes more mass from the upper half of rural income density, both rural income level and inequalities decrease, the urban-rural income ratio increases at the same time. On the contrary, the selection effect for temporary migrants is almost negligible. As a result, it does not have obvious effect on rural income level and inequalities.

The policy implications are clear. Migration does not necessarily equalize rural-urban income gap. As more educated rural residents left, the rural areas suffered from brain-drain problem. Therefore, we need multi-facet development strategies instead of focusing mainly on facilitating migration. To increase living standard of those left behind, more public services should be provided, including education, training, and social securities.

However, the above conclusions must be taken with qualifications. To focus on self-selection, we de-emphasize several important channels, through which migration might influence income distributions. First, when we construct the counterfactual income densities, we assume that the conditional density of income does not depend on the distribution of attributes. This is a strong assumption. In a general equilibrium setting, as the distribution of attributes changes, so do their prices. We simply ignore this impact in this paper.[18] Remittance is another important channel through which

[18] The reasons are two. First, general equilibrium effect is difficult to address. Assuming no general equilibrium effect is a common practice in many other researches (Dinardo, et al, 1996, and Chiquiar and Hanson, 2002 for example). Second, general equilibrium effect may be not a severe problem. Until recent years, wages of migrant

migration has impacts on rural income distribution. Many other researches put emphasis on this (see Li, 1999, and Du et al, 2005 for example). We just ignore it here because we don't have remittance data for permanent migrants and because it's not our focus. Another important caveat is that our analysis addresses observable skills, only. However, if the correlation between observable skills and unobservable skills is positive and sufficiently strong, we expect that our results would apply to migrant selection in terms of unobservable skills, as well. Finally, we assume that education is predetermined before the migration decision of rural residents. However, the fact that higher education increased the probability of migration means that migration may have an incentive effect on rural residents' human capital investment decision. If not all those with relatively high level of education succeeded in migrating out, migration may have a positive effect on the human capital level of rural areas.[19] This possibility is now being studied intensively in other countries (see Beine et al. 2001 and Beine et al. 2006 for example). Although the 'Brain-Gain' story may counter-balance the 'Brain-Drain' story, we don't know any research in Chinese context yet. We believe 'Brain-Gain' story is worth further research, but it's not our focus here.

Reference

Autor, David, Lawrence Katz, and M. S. Kearney. (2005), "Trends in U.S. Wage Inequality: Re-Assessing the Revisionists." NBER Working Paper No. 11627.

Beine M., F. Docquier and H. Rapoport (2001) "Brain Drain and Economic growth: Theory and Evidence" Journal of Development Economics, Vol. 64, 275-289.

Beine M., Frederic D. and H. Rapoport (2006) "Brain Drain and Human Capital Formation in developing countries: Winners and losers" Manuscript IZA, Bonn May 2006.

Borjas, George J. (1987), "Self-Selection and the Earnings of Immigrants", *American Economic Review* 77(4): 531-553.

Borjas, George J. (1999), "The Economic Analysis of Immigration", In Orley C. Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Amsterdam: North-Holland, pp. 1697-1760.

Chiquiar Daniel, and Gordon H. Hanson, (2002), "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States", NBER Working Paper 9242, http://www.nber.org/papers/w9242

Cai Fang and Wang Dewen, (2007) "Impacts of Internal Migration on Economic Growth and

workers remained low. This is a strong indication that there's still surplus of labor in rural areas. With large amount of surplus labor, it's more probable that skill prices changed little in rural areas.

[19] Brain gain may also happen through the channel of return migration or knowledge spillover.

Urban Development in China" , Institute of Population and Labor Economics, CASS.

deBrauw, A., Huang, J., Rozelle, S., Zhang, L., and Zhang, Y., (2002), "The evolution of China's rural labor markets during the reforms: rapid, accelerating, transforming." *Journal of Comparative Economics*, June 30(2), pp. 329-53.

Deng, Quheng and Gustafsson, Bjorn, "China's Lesser Known Migrants" (May 2006). IZA Discussion Paper No. 2152., http://ssrn.com/abstract=908236

DiNardo, J., Nicole. Fortin, T. Lemieux, (1996), "Labor Market Institutions and the Distribution of Wages, 1973-1992: a Semi-parametric Approach", *Econometrica*,64, (5) :1001 ~ 1044.

Du, Yang, Albert Park, Sangui Wang, (2005), "Migration and Rural Poverty in China", Working Paper, Institute of Agricultural Economics, CASS.

Foster, Andrew D, and Rosenzweig, Mark R., (200?), "Economic Development and the Decline of Agricultural Employment", in Robert E. Evenson, and Prabhu Pingali (eds.) *Handbook of Agricultural Economics*, North Holland.

Harris, J.R., and Todaro, M.P., (1970), "Migration, unemployment and development: A two-sector analysis." *American Economic Review*, 60:126-42

Heckman, J., (1979), "Sample selection bias as a specification error," *Econometrica*, 47. pp.153-161

Jacoby, H., (1993), "Shadow Wages and Peasant Family Labour Supply: An Econometric Application to the Peruvian Sierra", *Review of Economic Studies*, 60(4): 903-921.

Lemieux, Thomas, (2006), "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *American Economic Review*, Vol(96), No. 3. June.

Li, Shi, (1999), "Labor Migration and Income Distribution", *China Social Sciences*, No.2, pp.16-33.

Li, Shi, and Yue Ximing, (2004), "A survey on China's rural urban disparity", *Finance*, No. 4.

Li, Shi, and Ding Sai, (2003), "Long-term Change in Private Returns to Education in Urban China", *China Social Science (Zhongguo Shehui Kexue)*, No.6.

Machado ,José and José Mata (2005), "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression." *Journal of Applied Econometrics*, 20, 445-465.

Meng, Xin, (1996), "An examination of wage determination in China's rural industrial sector." *Applied Economics*, 28(1), pp.715-24.

Nakosteen, Robert and Michael Zimmer, (1980), "Migration and Income: the Question of Self-selection", Southern Economic Journal, Vol.46, No.3. (Jan., 1980), pp.840-851.

National Bureau of Statistics, 2005, *China Statistical Yearbook 2005*, China Statistical Press.

Oaxaca, Ronald, 1973. "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, vol. 14(3), pages 693-709, October

Park, Albert, Dewen Wang and Fang Cai, (2006), "Migration and Urban Poverty and Inequality in China", Working Paper, Institute of Population and Labor Economics, CASS.

Parish, W.L., Zhe, X., and Li, F. "Nonfarm work and marketization of the Chinese countryside." China Quarterly, September 1995, 143 pp. 697-730.

Wang Dewen and Cai Fang, 2006, "Migration and Poverty Alleviation in China", Institute of Population and Labour Economics, CASS

Zhang, Junsen, Zhao, Yaohui, Albert Park, and Song Xiaoqing, (2005), "Economic returns to

schooling in urban China, 1988 to 2001," *Journal of Comparative Economics*, 33, 730-752.

Zhao Zhong, (2005). "Migration, Labour Market Flexibility, and Wage Determination in China: A Review," *Labour and Demography* 0507009.

Table 1 Using Urban Survey to Identify Permanent Migrants

Time obtain an urban *hukou*?		The Ways Rural Residents Get Urban Hukou (%)							
		Education	Cadre	Join Army	Lose Land	Buy House	Others	Missing	Total
born urban	16,278	-	-	-	-	-	-	-	-
-1950	133	18.05	8.27	15.79	2.26	9.02	37.59	9.02	100
1950-60	509	28.29	4.91	13.36	0.59	3.93	40.86	8.06	100
1960-70	484	18.18	3.93	31.82	1.45	2.27	33.88	8.47	100
1970-80	839	23.6	4.53	14.18	2.86	1.91	49.7	3.22	100
1980-90	1,312	33.16	1.45	5.34	8.08	3.96	43.45	4.57	100
1990-	884	16.52	0.68	2.71	12.44	7.81	53.05	6.79	100
Missing	193	-	-	-	-	-	-	-	-
Total	20,632								

Table 2a Summary Statistics, Male (Aged 18-60)

	Urban residents				Rural residents		rural migrants (urban survey)	
	Urban natives		Migrants					
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Years of schooling	11.14	2.99	11.67	3.46	7.76	2.41	8.32	2.62
By education level[a]								
Primary and below	0.03	--	0.05	--	0.25	--	0.20	--
Middle school	0.29	0.45	0.20	0.40	0.56	0.50	0.56	0.50
High school	0.39	0.49	0.31	0.46	0.18	0.39	0.21	0.41
College and above	0.29	0.45	0.44	0.50	0.01	0.12	0.03	0.16
Age	42.02	10.09	44.02	8.83	38.14	12.01	35.29	8.49
Minority	0.04	0.21	0.03	0.18	0.13	0.34	0.09	0.28
Party member	0.29	0.45	0.53	0.50	0.13	0.33	0.05	0.22
Wage earner	0.81	0.39	0.87	0.33	0.54	0.50	0.95	0.22
Number of household member	3.19	0.78	3.14	0.76	4.40	1.34	2.77	0.95
Obs.	5114		1291		12328		1977	

Note: a. “high school” category also includes professional school.

Table 2b Summary Statistics, Female (Aged 18-60)

	Urban residents				Rural residents		rural migrants (urban survey)	
	Urban natives		Migrants					
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Years of schooling	10.65	3.05	9.90	3.57	6.48	2.89	7.33	2.88
By education level[a]								
Primary and below	0.06	--	0.13	--	0.45	--	0.34	--
Middle school	0.30	0.46	0.33	0.47	0.43	0.50	0.52	0.50
High school	0.42	0.49	0.35	0.48	0.10	0.31	0.13	0.33
College and above	0.22	0.41	0.19	0.40	0.01	0.09	0.02	0.12
Age	41.54	9.70	42.38	9.44	37.56	11.63	33.80	8.30
Minority	0.04	0.20	0.04	0.20	0.13	0.33	0.09	0.28
Party member	0.18	0.39	0.18	0.39	0.02	0.15	0.01	0.10
Wage earner	0.63	0.48	0.62	0.49	0.22	0.41	0.76	0.43
Number of household member	3.15	0.80	3.22	0.90	4.46	1.34	2.85	0.89
Obs.	5481		1211		11372		1892	

Note: a. “high school” category also includes professional school.

Table 3a Summary Statistics for Wage Earners, Male (Aged 18-60)

	Urban residents				Rural residents					
	Native workers		Migrant workers		Local wage earners		Migrants (Rural survey)		Migrants (Urban survey)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Years of schooling	11.34	2.96	11.95	3.37	8.00	2.40	8.07	2.23	8.29	2.62
By education level[a]										
Primary and below	0.02	--	0.04	--	0.21	--	0.18	--	0.20	--
Middle school	0.27	0.44	0.17	0.38	0.54	0.50	0.63	0.48	0.56	0.50
High school	0.39	0.49	0.31	0.46	0.23	0.42	0.18	0.38	0.21	0.41
College and above	0.32	0.47	0.47	0.50	0.02	0.14	0.01	0.11	0.03	0.16
Age	41.56	9.24	43.07	8.48	40.91	10.43	31.63	9.88	35.27	8.40
Minority	0.04	0.20	0.03	0.18	0.08	0.27	0.11	0.31	0.09	0.29
Party member	0.30	0.46	0.54	0.50	0.20	0.40	0.06	0.23	0.05	0.21
Working days per year	274	49	275	47	168	107	207	97	326	58
Working hrs per day	8.31	1.50	8.22	1.22	7.75	1.91	8.57	1.50	10.29	2.51
Obs.	4152		1129		4122		2543		1874	

Note: a. “high school” category also includes professional school.

Table 3b Summary Statistics for Wage Earners, Female (Aged 18-60)

	Urban residents				Rural residents					
	Native workers		Migrant workers		Local wage earners		Migrants (Rural survey)		Migrants (Urban survey)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Years of schooling	11.42	2.78	10.97	3.29	7.60	3.01	7.85	2.42	7.48	2.87
By education level[a]										
Primary and below	0.02	--	0.05	--	0.31	--	0.25	--	0.32	--
Middle school	0.21	0.41	0.26	0.44	0.45	0.50	0.60	0.49	0.52	0.50
High school	0.46	0.50	0.40	0.49	0.21	0.41	0.13	0.34	0.14	0.34
College and above	0.30	0.46	0.29	0.45	0.02	0.16	0.01	0.11	0.02	0.13
Age	38.99	8.43	39.19	7.95	35.88	10.20	25.53	7.75	33.46	7.90
Minority	0.04	0.20	0.05	0.22	0.04	0.19	0.09	0.28	0.08	0.27
Party member	0.21	0.40	0.22	0.41	0.07	0.26	0.01	0.10	0.01	0.11
Working days per year	271	52	276	51	189	111	237	91	327	60
Working hrs per day	8.06	1.30	8.23	1.63	7.85	1.78	8.47	1.43	10.34	2.41
Obs.	3455		748		1378		1072		1432	

Note: a. “high school” category also includes professional school.

Table 4 Income Distribution (Per Capita Income)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Rural residents	Urban residents	Urban natives	Type III migrants	Type I migrants	(1)+(4)	(1)+(5)	(1)+(4)+(5)	(1)+(4)	(1)+(5)	(1)+(4)+(5)
						Naïve recalculation			Reweighting calculation		
Mean (yuan/year)	2715	8174	8290	7678	6552	3196	3248	3623	2814	2712	2798
Inequality measures											
relative mean deviation	0.2624	0.2328	0.2335	0.2289	0.2458	0.2985	0.2946	0.3112	0.2677	0.2619	0.2666
coefficient of variation	0.8379	0.6756	0.6717	0.6892	0.9155	0.9664	1.031	1.0384	0.8693	0.836	0.8642
standard deviation of logs	0.6801	0.6082	0.6088	0.6021	0.6178	0.7465	0.7414	0.7785	0.6896	0.6793	0.6877
Gini coefficient	0.3683	0.3278	0.328	0.3245	0.3484	0.4127	0.4086	0.4275	0.3755	0.3676	0.3740
Mehran measure	0.4876	0.4442	0.4445	0.4406	0.4593	0.5349	0.5315	0.5541	0.4948	0.4869	0.4933
Piesch measure	0.3086	0.2696	0.2698	0.2664	0.2929	0.3517	0.3472	0.3642	0.3158	0.3079	0.3144
Kakwani measure	0.1195	0.0957	0.0957	0.0944	0.1093	0.1473	0.1447	0.1569	0.1239	0.1191	0.1231
Theil entropy measure	0.2441	0.1832	0.1827	0.1832	0.2382	0.3088	0.31	0.3323	0.2563	0.2432	0.2541
Theil mean log deviation measure	0.2336	0.1827	0.1827	0.1803	0.2079	0.2907	0.2869	0.3142	0.2427	0.2328	0.2409
Entropy measure GE -1	0.3213	0.2304	0.2314	0.2232	0.2455	0.4022	0.3973	0.445	0.333	0.3202	0.3307

Note: (1) Rural residents of column one include temporary migrants in rural survey (type II migrants) but not migrants in urban survey (type I migrants); (2) Urban residents include permanent migrants

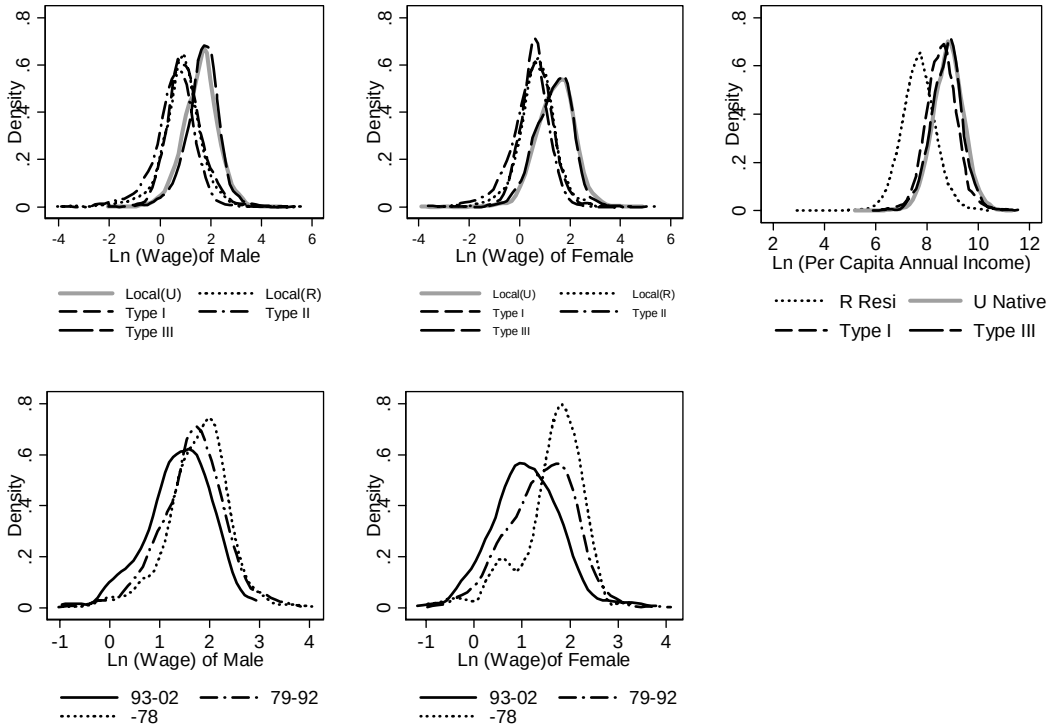

Figure 1 Wage (Income) Distributions

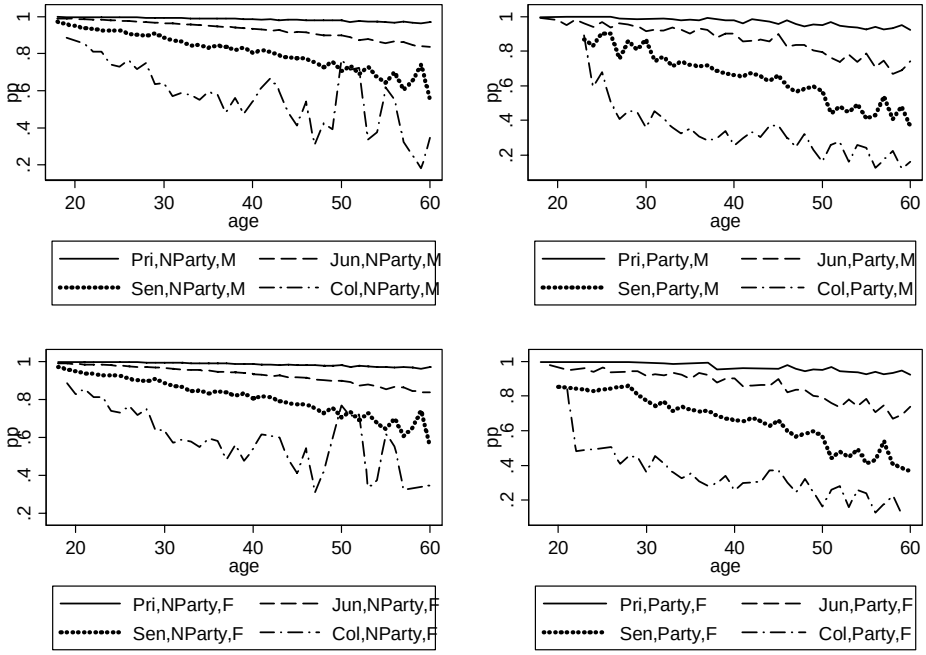

Figure 2 Predicted Probability of Being Rural Residents

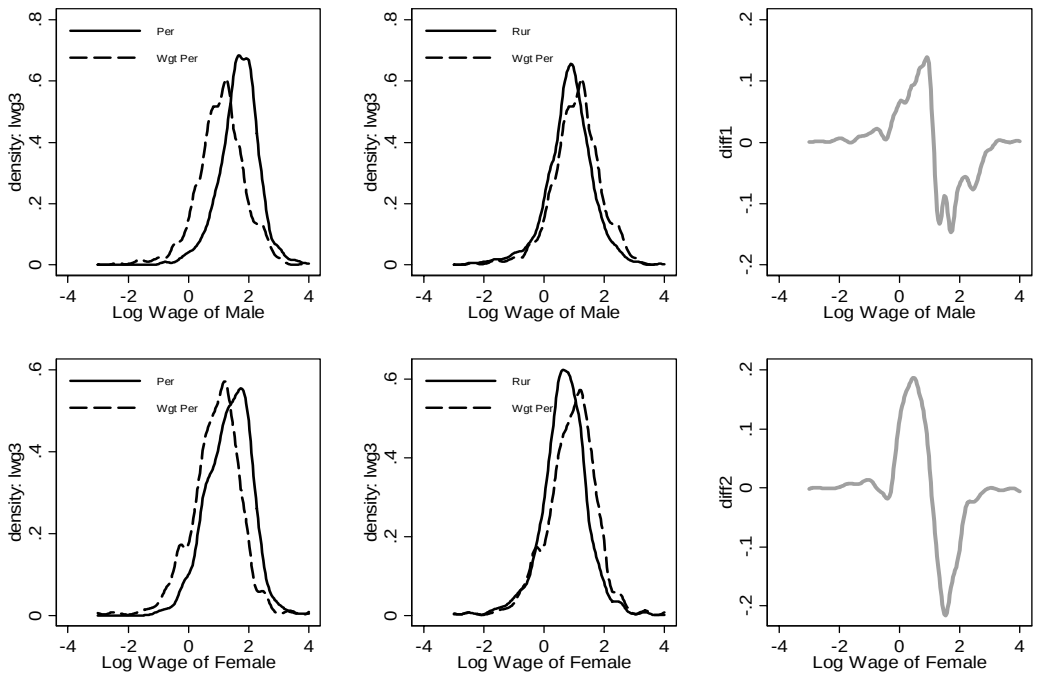

Figure 3 Counterfactual Wage Densities for Permanent Migrants

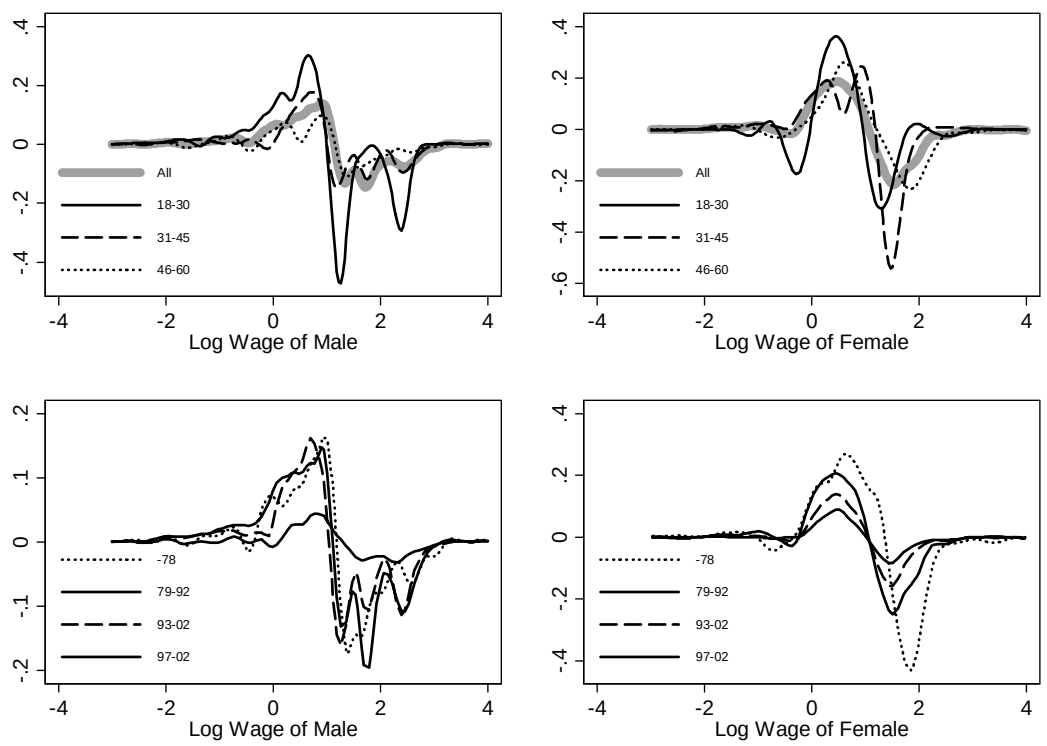

Figure 4 Differences between rural factual and counterfactual densities by age and by year of migration

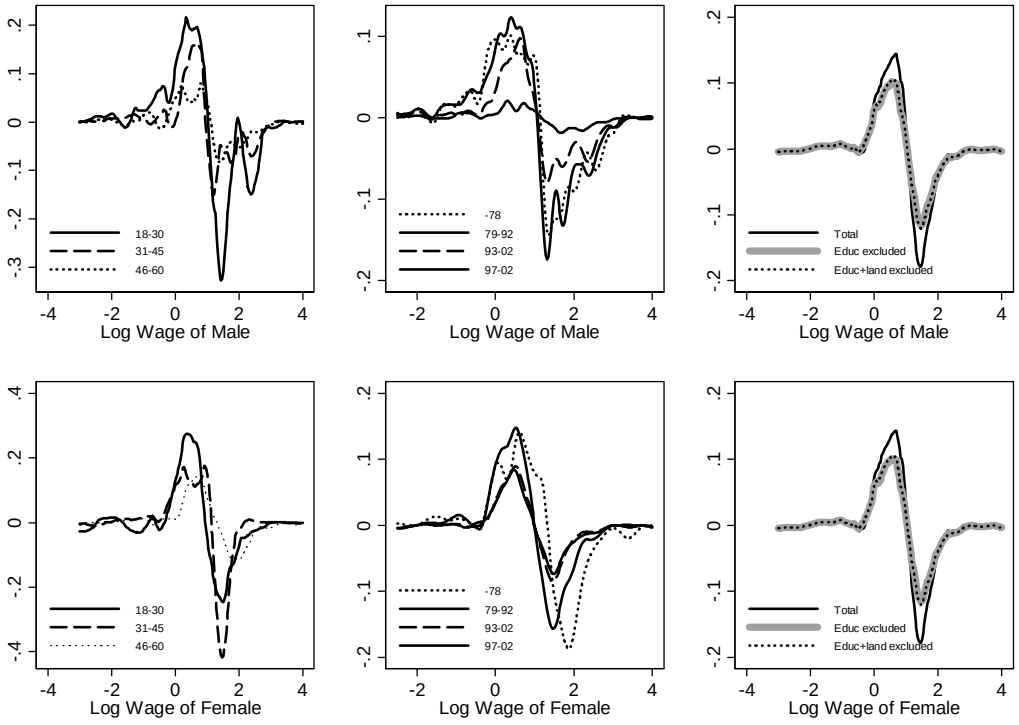

Figure 5 Robustness Check

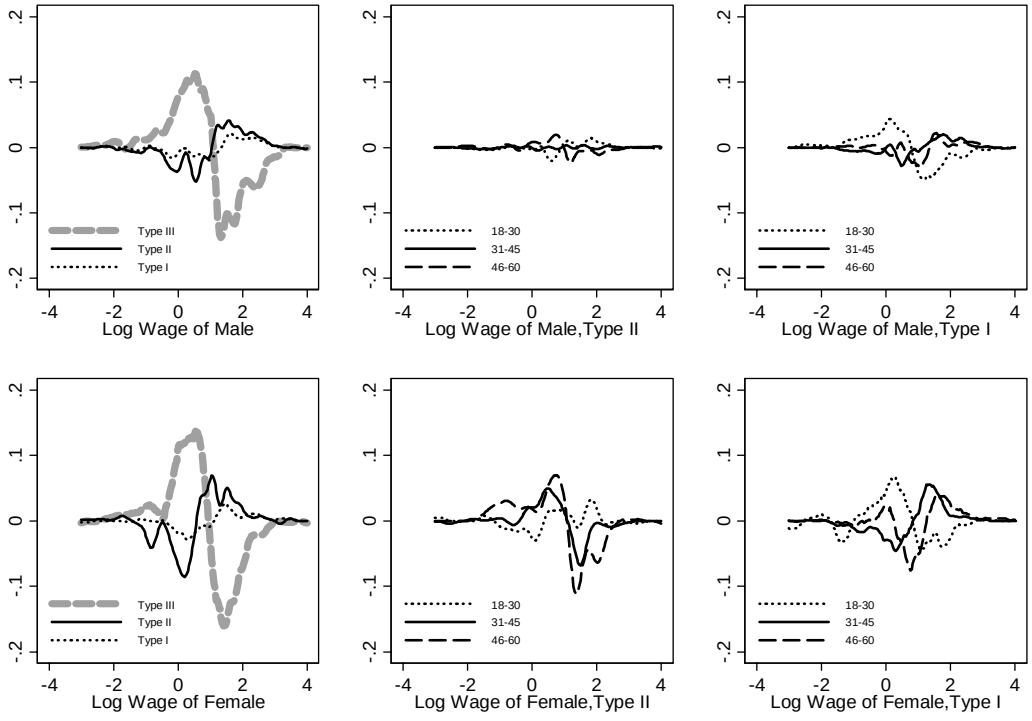

Figure 6 Differences between factual and counterfactual densities

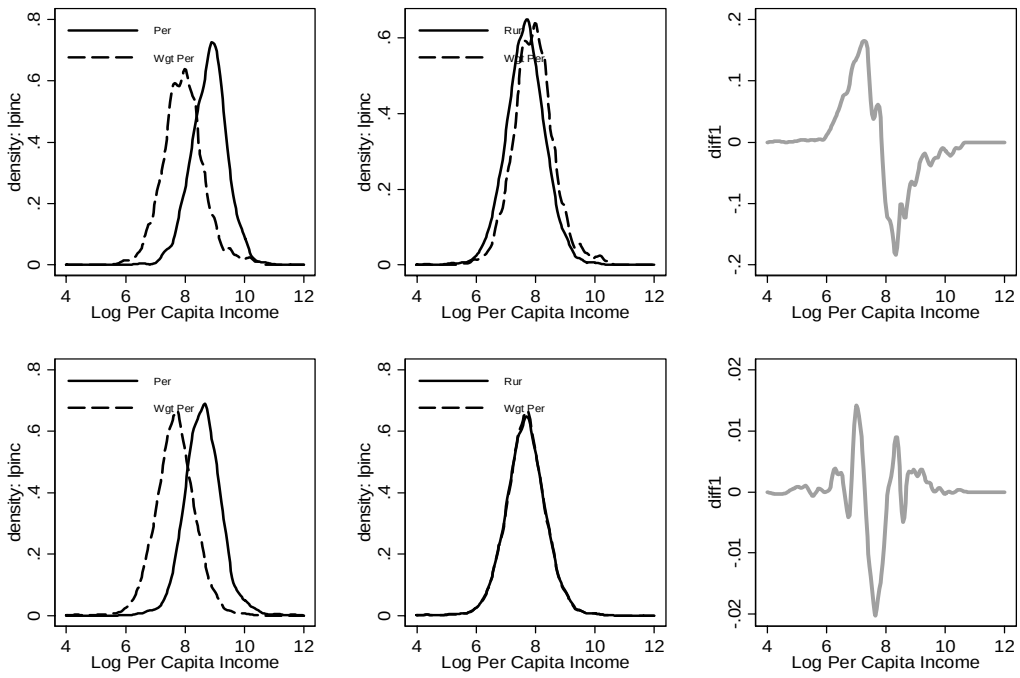

Figure 7 Counterfactual Densities of Per Capita Income for Permanent Migrants (upper) and Type I Migrants (lower)

Appendix

Table A 1a Summary Statistics for Permanent Migrants

	How did you obtain your urban hukou?					
Variable	Education	Cadre	Army	Land	House	Others
Age obtaining hukou	19.51	21.48	22.30	25.28	24.40	21.83
	(5.52)	(6.98)	(7.05)	(12.07)	(13.47)	(11.17)
Years of schooling	13.40	10.70	10.20	8.40	8.81	8.82
	(2.78)	(3.37)	(3.50)	(3.29)	(3.88)	(3.73)
Age	42.87	53.65	53.01	39.86	44.17	43.22
	(13.06)	(12.09)	(11.16)	(13.87)	(17.78)	(15.73)
Gender	0.39	0.40	0.05	0.61	0.59	0.61
	(0.49)	(0.49)	(0.22)	(0.49)	(0.49)	(0.49)

Note: Standard deviations are in parentheses.

Table A 1b Summary Statistics for Permanent Migrants

	When did you obtain your urban hukou?					
	-1950	1950-60	1960-70	1970-80	1980-90	1990-
Age obtaining hukou	17.22	17.52	18.56	21.30	22.09	23.26
	(8.53)	(7.09)	(7.56)	(7.68)	(10.32)	(13.05)
Years of schooling	7.41	8.74	10.20	10.62	11.36	9.29
	(4.46)	(4.26)	(3.78)	(3.62)	(3.68)	(4.01)
Age	69.41	63.62	54.05	46.71	38.47	30.40
	(8.54)	(7.46)	(7.85)	(7.97)	(10.61)	(13.23)
Gender	0.35	0.44	0.31	0.40	0.51	0.62
	(0.48)	(0.50)	(0.47)	(0.49)	(0.50)	(0.49)

Note: Standard deviations are in parentheses.

Table A 2a OLS Wage Regressions, Male

	Urban residents		Rural residents		
	Natives	Migrants	Local wage earners	migrants in rural survey	migrants in urban survey
Years of schooling	0.084***	0.059***	0.022***	0.023***	0.045***
	(0.004)	(0.007)	(0.006)	(0.008)	(0.007)
Experience	0.030***	0.008	0.029***	0.048***	0.024***
	(0.004)	(0.010)	(0.005)	(0.005)	(0.006)
Experience sqrd/100	-0.032***	-0.013	-0.049***	-0.071***	-0.050***
	(0.008)	(0.019)	(0.009)	(0.011)	(0.013)
Minority	-0.058	0.014	0.011	-0.017	0.059
	(0.047)	(0.100)	(0.060)	(0.059)	(0.057)
Party member	0.152***	0.205***	0.166***	-0.052	0.045
	(0.022)	(0.038)	(0.034)	(0.065)	(0.075)
Migrant cohort					
1961-70		-0.476**			
		(0.198)			
1971-80		-0.446**			
		(0.202)			
1981-90		-0.528**			
		(0.206)			
1990-		-0.637***			
		(0.212)			
Constant	0.541***	1.826***	0.620***	0.209	0.615***
	(0.074)	(0.298)	(0.117)	(0.161)	(0.131)
R-squared	0.266	0.253	0.094	0.220	0.095
N	4152	1129	4122	2543	1874

Note: 1). Region dummies are controlled. 2). "Migrant cohort" refers to when the migrants obtain his/her urban hukou. 3). Standard errors in parentheses. 4). *significant at 10% level, **significant 5%level, ***significant at 1% level.

Table A 2b OLS Wage Regressions, Female

	Urban residents		Rural residents		
	Natives	Migrants	Local wage earners	migrants in rural survey	migrants in urban survey
Years of schooling	0.109***	0.077***	0.025***	0.026**	0.042***
	(0.005)	(0.010)	(0.008)	(0.011)	(0.007)
Experience	0.026***	0.029**	0.013**	0.007	0.021***
	(0.004)	(0.012)	(0.006)	(0.009)	(0.006)
Experience sqrd/100	-0.020*	-0.067***	-0.015	0.009	-0.051***
	(0.011)	(0.025)	(0.013)	(0.022)	(0.013)
Minority	-0.012	0.008	-0.266**	-0.125	-0.018
	(0.057)	(0.117)	(0.134)	(0.097)	(0.060)
Party member	0.135***	0.131**	0.319***	0.624***	0.201
	(0.029)	(0.060)	(0.079)	(0.223)	(0.138)
Migrant cohort					
1971-80		0.080			
		(0.169)			
1981-90		-0.072			
		(0.169)			
1990-		-0.228			
		(0.177)			
Constant	0.159*	0.817**	0.478***	0.095	0.409***
	(0.091)	(0.347)	(0.148)	(0.278)	(0.124)
R-squared	0.295	0.253	0.131	0.268	0.126
N	3455	748	1378	1072	1432

Note: 1). Region dummies are controlled. 2). "Migrant cohort" refers to when the migrants obtain his/her urban hukou. 3). Standard errors in parentheses. 4). *significant at 10% level, **significant 5%level, ***significant at 1% level.