Centre National de la Recherche Scientifique

GATE
Groupe d'Analyse et de Théorie Économique
UMR 5824 du CNRS

UNIVERSITÉ LUMIÈRE LYON 2

DOCUMENTS DE TRAVAIL - WORKING PAPERS

W.P. 08-08

Migrants as second-class workers in urban China? A decomposition analysis

Sylvie Démurger, Marc Gurgand, Li Shi, Yue Ximing

Mars 2008

GATE Groupe d'Analyse et de Théorie Économique
UMR 5824 du CNRS
93 chemin des Mouilles – 69130 Écully – France
B.P. 167 – 69131 Écully Cedex
Tél. +33 (0)4 72 86 60 60 – Fax +33 (0)4 72 86 60 90
Messagerie électronique gate@gate.cnrs.fr
Serveur Web : www.gate.cnrs.fr

Migrants as second-class workers in urban China?
A decomposition analysis

Sylvie Démurger, Marc Gurgand, Li Shi, Yue Ximing*

This version: March 13, 2008

Abstract:

In urban China, urban resident annual earnings are 1.3 times larger than long term rural migrant earnings as observed in a nationally representative sample in 2002. Using microsimulation, we decompose this difference into four sources, with particular attention to path dependence and statistical distribution of the estimated effects: (1) different allocation to sectors that pay different wages (*sectoral effect*); (2) hourly wage disparities across the two populations within sectors (*wage effect*); (3) different working times within sectors (*hours effect*); (4) different population structures (*population effect*). Although sector allocation is extremely contrasted, with very few migrants in the public sector and very few urban residents working as self-employed, the *sectoral effect* is not robust to the path followed for the decomposition. We show that the migrant population has a comparative advantage in the private sector: increasing its participation into the public sector does not necessarily improve its average earnings. The opposite holds for the urban residents. The second main finding is that *population effect* is significantly more important than wage or hours effects. This implies that the main source of disparity is pre-market (education opportunities) rather than on-market.

JEL classification: J71, J31, O15, P23

Keywords: Chinese labor market, earnings differentials, migration, discrimination.

* Sylvie Démurger: University of Lyon, Lyon, F-69003, France; CNRS, UMR 5824, GATE, Ecully, F-69130, France; ENS LSH, Lyon, F-69007, France ; Centre Léon Bérard, Lyon, F-69003, France, demurger@gate.cnrs.fr; Marc Gurgand: Paris School of Economics and Crest, Paris, France, gurgand@pse.ens.fr; Li Shi: School of Economics and Business, Beijing Normal University, China, lishi@bnu.edu.cn; Yue Ximing: School of Finance, Renmin University of China, Beijing, China, yueximing@yahoo.com. This paper has benefited from financial support from the CEPREMAP Trade and Development program.

1. Introduction

The urban labor market in China has gone through tremendous changes over the last three decades of economic reforms. One of the most dramatic over the recent years is related to rural-urban migration that has soared with the loosening of administrative controls over population movements between rural and urban areas. Although it is difficult to evaluate precisely the actual number of rural migrants in Chinese cities, estimations reported by the National Bureau of Statistics amount to 132 million rural workers in cities in 2006 (National Bureau of Statistics, 2007)[1].

For decades since the late 1950s, the overall distribution of the Chinese population had been shaped by the strict policy of the household registration (*hukou*) system, which aimed at restricting migrations both between rural and urban areas and across regions. The main institutional barrier to mobility was then the exclusion of rural residents from the urban welfare system, which provided food ration, housing, medical care, education, childcare, and pension to urban residents. This system made it practically very difficult, if not impossible for rural *hukou* holders to survive in cities.

Economic reforms implemented from the late 1970s onwards have increased both the supply of and the demand for rural migrants in urban areas. As a consequence, population movements have risen sharply although not smoothly especially during the 1990s. Labor surplus in agriculture combined with reduced employment opportunities in township and village enterprises, increasing demand for labor by the booming urban private sector as well as a central government's policy of laisser-faire towards rural migration first fostered rural population movements to urban areas in the early 1990s. As a consequence, the number of rural migrants jumped from about 30 million in 1989 to 62 million in 1993 (Li, 2007). However, the reform of State-owned enterprises (SOEs) dramatically changed the situation faced by rural migrants because millions of urban laid-off workers entered the urban labor market after 1997. The harder competition on the urban labor market between urban unemployed and rural migrants was further reinforced by administrative regulations against rural migrants. By the end of the 1990s, several city governments had implemented local regulations to restrict rural migrants' employment and even forced enterprises to lay off migrant workers in favor of urban local workers. Other administrative regulations included a restricted access to certain job positions to urban residents only, or the imposition of fees to

[1] Given that 20% of rural migrant workers are estimated to live with their family in urban areas, the total number of rural-urban migrant population is estimated at around 160 million in 2006.

migrant workers and their employers (Knight *et al.* 1999, Appleton *et al.* 2004, Knight and Yueh 2004a, 2004b, Zhao 2005). This official administrative discrimination against rural migrants prevailed until recently when the central government issued a series of documents explicitly requiring local governments to enforce equal opportunities in employment and rights for rural migrants (Li, 2007). This change in policy towards rural labor mobility made the flow of rural migrants jumping again from about 80 million in 2001 to 132 million in 2006.

As it is often the case when two distinct labor force groups are competing in a labor market, the growing participation of rural migrants in the Chinese urban labor market raises the issue of potential discrimination behaviors against migrants. And indeed, various difficulties faced by rural migrants in terms of income and working conditions have been highlighted in the literature. Besides low income, the delayed payment of wages is a common feature for rural migrants. Their job mobility is much higher than local urban workers (Knight and Yueh 2004a) and they seldom have a contract signed with their employers. Their working conditions are tough, and they usually work much longer than the legal working time in low-end jobs that local urban workers do not want to take[2] (Yao 2001, Li 2007). As emphasized by Zhao (2005), the *hukou* system still makes it very difficult for rural migrants to enter the formal sector. Measures of perceived discrimination by rural migrants themselves illustrate these points by showing that in many aspects rural migrants consider that they do not enjoy the same treatment as urban workers. As an example, the Chinese Household Income Project (CHIP) 2002 survey data used in this paper indicate that 70% of rural migrants perceive discrimination in terms of wage paid for equal work, 71% in terms of type of work and 61% in terms of working hours. The disadvantaged position of rural migrants is not limited to earnings differentials and working conditions. Again, according to CHIP data, 81% of rural migrants consider being discriminated against in their chance to be promoted, 82% in housing provision, and 85% in social security.

This paper intends to contribute to the understanding of the discriminatory behaviors against rural migrants by specifically focusing on the explanation of earnings differentials between rural migrants and urban residents. To that purpose, we use a nationally representative sample of urban residents and rural migrants for the year 2002 and we propose an extended form of Oaxaca-Blinder decompositions to explain the observed earnings differentials between the two populations. One important issue raised in the literature

[2] The construction sector is a typical example here. According to the Project Team of Research Council, State Council (2006), data from the 2000 Census indicate that 80% of all the jobs in the construction sector are taken by rural migrants. These jobs are typically low-skilled, hard and dangerous, and as such are not valued by urban residents.

concerns the respective contribution of different allocations into sectors *versus* different wages earned within each sector, what Liu *et al.* (2004) refer to as between- and within-occupation wage effect. We explore this question by focusing on two possible sources of segregation: (1) a differentiated access to sectors (we distinguish here the public sector, the private sector and self-employment); and (2) differentiated earnings within each sector.

Although the evolution of the labor market in urban China has received a large attention, evaluations of the earnings gap between rural migrants and urban residents in China remain limited, mainly because of the paucity of relevant data. Even when adequate data is available, their scope tends to be limited to a few regions or cities (Knight *et al*. 1999, Meng 2001, Meng and Zhang 2001). Given the huge regional differences across China, this limitation makes cross-study comparisons difficult and any generalization irrelevant. The most in-depth analysis of earnings differentials between rural migrants and local urban workers in China to date is certainly that of Meng and Zhang (2001). Using two comparable household survey data sets for Shanghai in 1995, they find evidence of discrimination against rural migrants in terms of both occupational attainment and earnings. Following the methodology of Brown *et al.* (1980), they analyze the extent to which earnings differentials between rural migrants and urban residents are due to inter- or intra-occupational gaps and find that 82 percent of the hourly wage differential is due to unequal payment within occupation.

This paper tries to provide a general overview of earnings differentials between rural migrants and urban residents in China, using data from a nationally representative sample for 2002 made of comparable surveys for urban residents and rural migrants. The data set was collected under the China Household Income Project (CHIP), and contains about 8,000 observations for working individuals. It not only provides a wider scope than previous analyses, but also enables the comparison of two sub-populations that may be in strong competition for jobs in urban China. Indeed, rural migrants surveyed in the CHIP data were selected from resident communities (Khan and Riskin, 2005). Although not capturing the wide spectrum of rural migrants (those living in construction sites and factories were excluded from the sampling process), these data are relevant for the purpose of our study since the surveyed migrants, already settled in cities, can be expected to have characteristics closer to urban residents against whom they are competing in the labor market.

Furthermore, we adopt a decomposition analysis based on microsimulation that substantially departs from the traditional approach based on the Brown *et al.* (1980) extension of Oaxaca-Blinder decompositions. Our approach, formally based on sector allocation models allows for the evaluation of direct as well as indirect effects of changes in sector allocation on earnings differentials. In particular, it shows differences in comparative advantages between sectors for rural migrants and urban residents in the urban labor market.

The paper is structured as follows. The next section presents the decomposition methodology and section 3 describes the data used. Occupational distribution and wages and hours structures are discussed respectively in section 4 and 5. Section 6 presents the results of the decomposition analysis and discusses the various effects at stake. Concluding remarks are given in section 7.

2. Decomposition methodology

On average, our dataset shows that urban residents earn 1.3 times as much as rural migrants in 2002. We decompose this gap into 4 complementary effects: (1) the effect of different allocations between self-employment, public jobs and private jobs; (2) the effect of different hourly earnings structures; (3) the effect of different working times; and (4) the effect of the distribution of observed individual characteristics in the two populations. The decomposition is implemented by first estimating job allocation, earnings and hours equations, and then simulating counterfactual job status, earnings and hours.

The traditional approach to decomposition in this context follows the Brown *et al.* (1980) extension to Oaxaca-Blinder decompositions, which explicitly treats differences in occupational distributions between the two groups under investigation. This method has been applied in particular by Liu *et al.* (2004) for Hong Kong and Meng and Zhang (2001) for China. Our approach, however, is substantially different from the Brown *et al.* model in that it takes into account the fact that participation changes have indirect effects on within sector average earnings, as they affect population composition in the sectors. As will be illustrated, taking this dimension into account may greatly affect the results.

2.1. Model

To decompose the difference in average earnings between urban residents and rural migrants, we start with the following model. We consider two groups of workers, urban residents (u) and rural migrants (m), who can work into 3 different sectors indexed by k = 1, 2, 3 (self-employment, public sector and private sector). If Z is a vector of individual characteristics, the individual latent propensity to work in sector *k* for a person *i* belonging to group $g = \{u, m\}$ is assumed to be of the form:

$$P_i^k = Z_i \delta^{gk} + \eta_i^k$$

where the parameters δ^{gk} are group and sector specific. A person is observed working in sector *k* if $P_i^k = \max_j \{P_i^j\}$.

Within each sector *k*, hourly earnings are given by:

$$\log w_i^k = X_i \beta^{gk} + u_i^k$$

where X is a subset of Z that contains human capital variables. Working time is given by:

$$h_i^k = Z_i \gamma^{gk} + v_i^k$$

These parameters can have various interpretations. As a general setting, the propensities to be found working in a sector, as well as the working time in that sector, may depend on: (*i*) the expected income in sector *k*, (*ii*) individual preferences for sector *k*, and (*iii*) a restricted access to some sectors for some groups. Since the above specifications are reduced forms, the parameters δ^{gk} and γ^{gk} can capture both preferences and constraints. In particular, a difference between δ^{uk} and δ^{mk} can be explained by occupational segregation (demand driven) between urban residents and rural migrants as well as by different preferences (supply driven) across the two populations. Differences in earnings parameters β^{gk} are more readily interpretable in terms of segmentation, although they could also reflect compensating differentials.

The sector allocation model is estimated by maximum likelihood using a multinomial logit model. Residuals η_i^k are thus assumed to be i.i.d, with a Gumbell distribution (McFadden, 1973). In this model, the probability to work in sector *k* is:

$$\Pr(k|Z_i) = \frac{\exp(Z_i \delta^{gk})}{\sum_j \exp(Z_i \delta^{gj})}$$

if *i* belongs to group *g*. The sectoral choice model can also be used to correct for selectivity in earnings equations (Bourguignon *et al.*, 2007). Yet, achieving identification for sector choice is problematic since no available variable can be considered as a fully exogenous instrument. In an attempt to control for possible selectivity bias, we also estimated selectivity-corrected earnings functions with available imperfect instruments and found no quantitative difference for the general results (see Appendix 3). We therefore present results based on models estimated without selectivity correction (earnings and hours equations are estimated by ordinary least squares) to keep the presentation and discussion simple.

2.2. Decomposition

To present the decomposition methodology, the above model is embedded into a more general notation. If we note δ^g the full vector of sector choice parameters specific to any group *g*, and β^g and γ^g accordingly, the observed differential in average income between the two populations can be decomposed into four parts as follows:

1. The part due to different activity generating processes (δ^u versus δ^m), through the simulation of the sector in which urban residents would be working if they had the same activity allocation rule as migrants, and inversely.
2. The part due to different hourly earnings generating processes (β^u versus β^m), through the simulation of how much urban residents would be paid if they were paid according to the migrants' earnings generating model in a given activity, and inversely.
3. The part due to different hours generating processes (γ^u versus γ^m), through the simulation of how long urban residents would be working if their working time model was that of migrants, and inversely.
4. The part due to a different distribution of observed characteristics Z in the two populations.

Dropping hours for clarity, a typical decomposition in the Brown *et al.* (1980) approach would be:

$$\begin{aligned}\bar{y}^u - \bar{y}^m \quad &= \sum_{k=1}^{3} \bar{w}^k(u,\beta^u)[\bar{s}^k(u,\delta^u) - \bar{s}^k(u,\delta^m)] \\ &+ \sum_{k=1}^{3} \bar{s}^k(u,\delta^m)[\bar{w}^k(u,\beta^u) - \bar{w}^k(u,\beta^m)] \\ &+ \sum_{k=1}^{3} [\bar{s}^k(u,\delta^m)\bar{w}^k(u,\beta^m) - \bar{s}^k(m,\delta^m)\bar{w}^k(m,\beta^m)]\end{aligned}$$

where $\bar{s}^k(u,\delta^u)$ is the proportion of urban residents (*u*) in sector *k* when the sector allocation rule is δ^u, and $\bar{w}^k(u,\beta^u)$ is the average hourly earnings of urban residents actually observed in activity *k* when the hourly earnings generating process is β^u.

This algebraic decomposition works well when the activity dimension is absent, as in the original Blinder-Oaxaca approach. However, including activity in the decomposition generates composition effects across sectors: by changing the sector allocation rule (e.g. from δ^u to δ^m) while keeping the rest constant (notably the earnings generating process), the average earnings in each sector is modified accordingly because the individuals allocated to each sector are not the same any more. This is part of the activity effect. In contrast, the above decomposition uses observed average earnings ($\overline{w}^k(u,\beta^u)$) to evaluate the effect of activity changes. Such earnings level has no counterfactual meanings. This is an important issue if some variables do affect the sectoral allocation differently in the urban and the migrant models. This problem is specific to the introduction of activity choice and does not appear in the original Oaxaca-Blinder approach.

As a result, we propose the following decomposition:

$$
\begin{aligned}
\bar{y}^u - \bar{y}^m \quad &= \frac{1}{U}\sum_{i\in u}\sum_{k=1}^{3} w^k(Z_i,\beta^u)1[s(Z_i,\delta^u)=k] \\
&- \frac{1}{U}\sum_{i\in u}\sum_{k=1}^{3} w^k(Z_i,\beta^u)1[s(Z_i,\delta^m)=k] \\
&+ \frac{1}{U}\sum_{i\in u}\sum_{k=1}^{3} 1[s(Z_i,\delta^m)=k][w^k(Z_i,\beta^u)-w^k(Z_i,\beta^m)] \\
&+ \frac{1}{U}\sum_{i\in u}\sum_{k=1}^{3} w^k(Z_i,\beta^m)1[s(Z_i,\delta^m)=k] \\
&- \frac{1}{M}\sum_{i\in m}\sum_{k=1}^{3} w^k(Z_i,\beta^m)1[s(Z_i,\delta^m)=k]
\end{aligned}
\qquad (1)
$$

where 1[.] is an indicator variable, $i \in u$ means all individuals i belonging to the urban resident population and U is the total size of this population (resp. m and M for migrants). Line 1 minus line 2 gives the *sector allocation effect* (δ^u vs. δ^m), the third line gives the *earnings effect* (β^u versus β^m) and line 4 minus lines 5 gives the *population effect*[3].

Implementing this decomposition requires simulating individual counterfactual occupations. To that aim, we initially draw values of η for each individual, conditional on Z and his/her observed activity. We then use these drawn values to determine the allocation into counterfactual sectors. For instance, if individual i is a migrant and has received ($\hat{\eta}_i^1 ... \hat{\eta}_i^K$) compatible with her observed sector, her urban resident sector allocation counterfactual will be $[s(Z_i,\delta^u)=k]$ if $(Z_i\delta^{uk}+\hat{\eta}_i^k)=\max_j\left\{Z_i\delta^{uj}+\hat{\eta}_i^j\right\}$. Earnings and hours counterfactuals also plug in residuals $\hat{u}$ and $\hat{v}$ based on observed status[4].

A last issue, often overlooked in the literature, is path-dependence. The sector allocation effect in equation (1) is computed on the urban resident population and using the urban resident earnings determination rule β^u. But it could also be based on the rural migrant population, or with β^m, or both. There is no reason to expect that the effect will be identical

[3] If hours had not been dropped for legibility, there would be an additional term showing the *working hours effect*.
[4] Whenever the individual is not simulated in her original sector, she is given a residual value from the destination sector observed distribution, so that her rank in her initial distribution is preserved.

for all combinations. The same holds for each term of the above decomposition, so that the contribution of the various terms can be sensitive to the chosen path. It is thus necessary to compute every variant and check the robustness of the results.

An interesting situation where path-dependence can be important occurs when the optimal sector allocation rules are very different for the two populations. In that case, no allocation is generally superior. Consider for example the case where urban residents are better paid than migrants in every sector, but migrants receive higher wages in the private sector, whereas urban residents receive higher wages in the public sector. Since urban residents are far more concentrated in the public sector than migrants, the result of moving from δ^u to δ^m (*sector allocation effect*) will be to decrease the share of the public sector. Depending on the model chosen for generating earnings, this shift would either decrease (urban residents earnings model) or increase (rural migrants earnings model) overall income. This is a situation that will appear in the data.

3. Data

The data used in this paper has been collected during spring 2003 under the *China household income project* (CHIP), coordinated by the Institute of Economics, Chinese Academy of Social Sciences, with assistance from the National Bureau of Statistics (NBS). The urban section of the household income survey contains two distinct sub-samples, one on urban *hukou* households and the other on households living in urban areas without urban *hukou*. These rural-urban migrant households are selected from the same twelve provinces[5] as urban households, but not from all of the cities included in the urban survey. Since rural-urban migrants are mostly concentrated in large cities, the provincial capital cities plus one or two medium-sized cities in each province have been selected for the migrant survey. We restrict our analysis to cities common to the two sub-samples and to individuals aged 16 to 60 who declared working at least part of the year and earning wages or income from self-employment. The sample contains 4,978 observations for urban *hukou* working individuals and 3,035 observations for rural migrants.

The sample scheme for the rural migrant survey was to allocate 200 households to each province in the coastal and interior regions and 150 households to each province in the western region. Within each province, 100 households were drawn from the capital city and

[5] The twelve provinces are: Anhui, Beijing, Chongqing, Gansu, Guangdong, Henan, Hubei, Jiangsu, Liaoning, Shanxi, Sichuan and Yunnan.

50 households from other cities. Within cities, only rural-urban migrant households living in residential neighborhoods were sampled. This implies that migrant workers living on construction sites or in factory dormitories are not accounted for (Khan and Riskin, 2005)[6].

With little outside knowledge about the distribution of the migrant population by age, gender and location, it is difficult to make a judgment on how representative the migrant data is. This judgment also depends on how migrant households are defined. As indicated in some statistics from rural-urban individual migrants, the majority of migrants are single workers living in dormitories or construction sites. However, these single migrant workers are not our analysis unit since they experience temporary and short-term migration, with lower expectation to settle down with their families in cities. The migrant households covered in our sample are rather representative of long-term migrants living with their family and, as such, more directly comparable to local urban households. The average length of stay in cities for rural migrants is 7.34 years at survey year and 50% have been living in cities for more than 6 years; more than 75% had been staying in cities for the whole year.

Although the information collected in the two sub-surveys is meant to be consistent (with similar questions asked in the two surveys), earnings deserve particular attention. For urban *hukou* holders, questions concerning earnings are rather comprehensive. As discussed in the literature, the CHIP data is particularly careful with earnings measures: although it does not fully account for all fringe benefits provided by the public sector (such as implicit contribution to pensions, health insurance, or preferential housing rents), it includes some important non-monetary benefits (e.g. housing, medical care, child care and regional subsidies).

For urban *hukou* holders, the earnings variable is thus defined as the individual income of active workers earned from their own private business or work units[7]. For wage-earners, it is the sum of cash labor compensations (basic salary, bonuses, allowances, subsidies and other wages or income) and income in kind. For rural migrants, earnings are computed from the reported average monthly income in 2002 from their current job and the total (net) income from other sources. Available data do not allow us to take income in kind into account for rural migrants, which may slightly bias upward the observed income gap between migrants

[6] A full description of the sampling method and the data can be found in Li *et al.* (2007).

[7] Measuring income for self-employed is a highly debated issue. A recent paper by de Mel *et al.* (2008) compares the relative quality of direct reports of profits with details of revenues and expenses, and concludes that the former are likely to be more accurate in measuring firms' profits. In this vein, our earnings variable is based on reported net income of private businesses.

and urban residents. However, given that rural migrants have restricted access to subsidized services, this should not be a serious issue.

An important component of earnings differentials in China may arise from differences in living standards between different cities. To account for this issue, earnings are adjusted for provincial purchasing power differences, using Brandt and Holz (2006) urban provincial-level spatial price deflators. This adjustment makes cross-provinces data much more comparable than the non-deflated data usually used in the literature.

Table 1 shows statistics on sector allocation, average earnings and working time by sector (self-employment, public sector and private sector). Urban residents' annual earnings are 1.3 times larger than rural migrant earnings: 11,881 yuan vs. 9,335 yuan. This results from a large difference in hourly wage between the two populations that is compensated by longer migrant working hours. The hourly wage of urban residents is on average twice that of rural migrants, the ratio being much higher in the public sector (2.3) than in both self-employment (1.3) and the private sector (1.5). However, rural migrants work on average 69 hours per week, whereas urban residents work on average 44 hours a week. The fact that rural migrants, who receive lower hourly earnings, also tend to work longer may imply a strong income effect in labor supply behavior, but may also result from working constraints imposed by employers to workers with limited negotiating power.

Table 1 also shows that occupational distributions are extremely contrasted across the two groups and, as such, this is a potentially important source of income differences, as sectors have different wage-setting structures (Chen *et al.*, 2005). Indeed, there is a very strong concentration of rural migrants in self-employment and, to a lower extent, in the private wage-earning sector (respectively 57% and 36%), whereas urban residents are overwhelmingly employed in the public sector (71%) and only slightly in self-employment (4%).

The comparison between the public and the private sector hourly wage structure also reveals an interesting difference between urban residents and rural migrants. Indeed, while urban residents working in the public sector get a much higher hourly wage than those working in the private sector (1.4 times higher), rural migrants earn slightly *less* in the public sector than in the private sector.

Table 2 provides a description of individual characteristics, which highlights very important endowment differences between the two groups. Hence, urban residents are on average older and much more educated (almost 4 years difference) than rural migrants. As compared to the Shanghai sample used by Meng and Zhang (2001), migrants in our sample are slightly older (34.31 *versus* 27.07) and most are married (90% *versus* 55%), which is consistent with the fact that we are focusing on less temporary migrants. Urban residents are also far more often members of the Communist Party and have much more experience than rural migrants, experience being measured by the actual number of years of work in urban areas. Last, in terms of job status, the lower average qualification of rural migrants is illustrated by their very low share in white collar jobs: only 4.5% of rural migrants hold professional or technician positions (to be compared with 32.7% of urban residents) and 2.2% are office workers (against 19.6% for urban residents).

4. Occupational distribution

As described in section 2, we can evaluate the extent to which the occupational distribution is "biased" against rural migrants by simulating the occupational distribution of each group using the other group's sector allocation model. We first run a multinomial logit model over the choice of activity (self-employment, public sector, private sector), whose results are reported in Table 3. Explanatory variables include individual characteristics (education, age, gender, communist membership, geographical residence) as well as household characteristics (household size, number of children less than 6 years old).

Education influences the choice of both urban residents and rural migrants towards the public sector, the estimated effect being significantly stronger for urban residents[8]. This implies that the urban resident model selects educated workers into the public sector much more strongly than the rural migrant model does. Moreover, although education increases the probability for rural migrants to work in the private sector as compared to self-employment, it does not significantly increase their chances to work in the public sector as compared to the private sector (which is not the case for urban residents).

The impact of age on activity choice only appears significant for rural migrants. Estimations show an inverted U-shape relationship between age and the probability of entering self-employment: rural migrants in their early 40s have the highest probability to work as self-employed. A potential explanation is that young migrants entering the urban

[8] Although coefficients absolute values are not directly comparable in Table 3, the significance of the difference between urban residents and rural migrants equations has been checked by pooling the data and adding interaction terms for all variables with a "migrant" dummy.

labor market mostly start working in the wage-earning sector and only switch to a more risky position when they have acquired enough economic, human and social capital[9].

From the multinomial logit model, we can simulate the sector allocation of each individual in the sample, under the rules that prevail in the other population. In other words, the simulation answers the question: "in which sector would urban residents (rural migrants) work if they were allocated to activities according to the rural migrants (urban residents) model?". Table 4 compares the observed and the simulated marginal distributions.

Shifting from the migrant model to the urban model decreases the share of self-employment from 57% to 11% if applied to the migrant population and from 50% to 4% if applied to the urban population. Inversely, the same model change increases public sector share from 7% to 52% in the migrant population and from 14% to 72% in the urban population. The fact that the amplitude of these effects is sensitive to the population indicates that contrasted occupational distributions are only partly explained by a segregation (or model) effect and that population characteristics do play a role. However, it is interesting to note that the distribution of occupations based on the rural migrant model is much less sensitive to population changes than the distribution based on the urban model. This suggests that observed individual characteristics play a stronger role in the urban resident model than in the rural migrant model, as was already apparent from Table 3[10].

5. Earnings and hours structures

Six earnings and hours equations have been estimated, for self-employment, the public and the private sectors, and for the two populations (see Appendix A1 and A2). Table 5 provides a synthetic view of the corresponding structures[11]. In order to neutralize within-population composition effects, simulated hourly earnings (and working hours) are computed separately for the *whole* urban and migrant populations, and for each of the six self-employment/public/private urban/migrant model combinations. For instance in panel A, the first row shows in the first column, the average hourly earnings for the *whole* urban resident population under the 'urban/self-employment' earnings model (i.e. based on $\beta^{u,self}$), and in the second column, the average hourly earnings for the *whole* urban resident population under the 'migrant/self-employment' earnings model (i.e. based on $\beta^{m,self}$). It indicates that, if paid

[9] This result is consistent with Meng (2001) who finds that, in 1995, rural migrants in Jinan city are more likely to be self-employed in the informal sector as their city work experience increases.

[10] In this respect, the population effect in the decomposition should be stronger whenever based on the urban occupation model (see section 6).

according to the urban rules, the urban population would earn 3.90 yuan per hour on average in self-employment, whereas, according to the migrant rules, the same population would earn only 3.11 yuan per hour in the same sector.

Unsurprisingly, whatever the earnings model, the simulated wage for the urban population (left table) is always higher than the corresponding simulated wage for the migrant population (right table). Hence, for any wage structure, rural migrants earn less than urban residents, which is clearly a composition effect.

Table 5 also shows that the urban resident's earnings structure is always more favorable than the migrant's earnings structure when applied to the urban population, but not when applied to the migrant population. Indeed, according to the urban resident earnings model, the average hourly earnings for the urban population would be 3.90 yuan in self-employment, 5.92 yuan in the public sector and 4.93 yuan in the private sector, whereas according to the rural migrant earnings model, the corresponding average hourly wage for the same population would respectively be only 3.11 yuan, 5 yuan and 3.52 yuan. Results for the migrant population highlight a different pattern since, *in the private sector*, the migrant model is actually more favorable to migrants than is the urban resident model: the simulated hourly wage under the migrant earnings model (2.83 yuan) is higher than the simulated wage under the urban resident earnings model (2.61 yuan).

Another interesting result is that the public sector always pays better than self-employment and the private sector, except for migrants under the migrant model. Indeed, *in the migrant model*, both self-employment and the private sector are more favorable to migrants (with a simulated hourly wage of respectively 2.90 yuan and 2.83 yuan) than is the public sector (2.59 yuan per hour).

These results suggest that migrants may have a comparative advantage with respect to self-employment and the private sector. Since this feature does not apply to the urban resident population, it is certainly related to some specific combinations of productive characteristics. One explanation lies in the nature of public jobs offered to migrants and in returns to human capital. Indeed, rural migrants generally hold low-end non-tenured jobs in the public sector, which are very poorly paid on average, but with some returns to education. On the other hand, self-employment and the private sector provide better paid jobs to migrants, but with smaller

[11] The simple structure of weekly hours given in panel B does not deserve much explanation and will not be discussed here. The main result is that working hours are longer in self-employment and the private sector, as well as in the migrant model, with no exception. As urban residents are more concentrated in the public sector, this provides two reasons for the lower average working hours observed in this population.

returns to education[12]. As a result, in a population with a low education level, higher returns to education in the public sector do not make up for low baseline wages: public wages are lower on average. Hence, rural migrants are relatively better off in self-employment and the private sector where the returns to human capital are lower for them. In the more educated urban population, a higher return to schooling ensures that public wages are higher on average. As mentioned in section 2, this feature is likely to generate strong path-dependence in the sectoral decomposition because shifting migrants away from self-employment will not always make them better off.

6. Decomposition analysis

Putting all these elements together, we can decompose the earnings differential between urban residents and rural migrants into activity, earnings, hours and population effects. As mentioned in section 2, there are several paths to the decomposition. For instance, the effect of changing the sector choice model can be computed either on the urban or on the migrant population, and in each case, using either urban or migrant earnings and hours models. This results in 8 different possibilities. To start with an overview, Table 6 presents the average effects over all paths, both in absolute value and as a percentage of the observed gap.

The average observed difference in annual earnings is 2,546 yuan. If the two populations differed only by their allocation into sectors, this difference would be only -32 yuan on average, or 2% of the total difference. Unsurprisingly, moving from the migrants' hourly earnings model to the urban residents' hourly earnings model would increase the earnings gap to 1,162 yuan (46% of the total). Inversely, everything else equal, migrants would earn 2,068 yuan *more* than urban residents as a result of their much longer working time (82% of the total). Finally, the strongest effect is related to differences in observed characteristics between the two populations: by itself, it generates a 3,487 yuan earnings difference. In a nutshell, this means that given the significant working time effect, urban residents would not earn much more on average if they did not have much better endowments, such as education and city work experience. This is reinforced by the fact that part of the hourly earnings effect may also capture unobserved productive characteristics that are not evenly distributed in the two populations (and should belong to the population effect)[13].

[12] Table A1 illustrates these differences in returns to schooling across sectors for migrants.

[13] A common limitation of the earnings differentials literature is that the constants incorporate differences in means of unobserved characteristics across populations. In this respect, we arbitrarily incorporate into

There are, however, several sources of uncertainty over the meaning of these averages. One is path dependence. Another comes from the fact that all counterfactual earnings are based on estimated parameters: if these parameters are not precisely estimated, we may well shift on one side or the other by mere chance. Therefore, it is important to take into account the variances of the estimators. Finally, some residuals follow from random draws, which brings additional randomness.

In order to assess the robustness of our results, we have bootstrapped (200 times) the data so as to generate a distribution of occupation, earnings and hours equation parameters. For each of the 200 iterations, we have drawn new residuals and computed the effects along all possible paths. Figures 1 to 4 draw the distribution of each of the four decompositions presented in Table 6, expressed as a percentage of the observed earnings gap. Each figure shows the full distribution for the different paths. For example, Figure 1 shows the distribution of the activity effect for the four possible paths: the urban population under the urban earnings/hour model (*urbpop/urbwage*), the urban population under the migrant earnings/hour model (*urbpop/migwage*), the migrant population under the urban earnings/hour model (*migpop/urbwage*), the migrant population under the migrant earnings/hour model (*migpop/migwage*). The sign of the hours and population decompositions (Figures 3 and 4) is robust, although the population effect is rather sensitive to the model of reference, as anticipated in section 4[14].

Both the activity effect and the hourly earnings effect deserve specific attention. Figure 1 and Table 7 indicate that the small average activity effect actually hides strong path dependence. They both show the effect on total earnings of moving from the migrant to the urban occupation model, when applied alternatively to the urban and the migrant population and using the urban or the migrant earnings and hours structures. The effect is clearly positive when applied to the urban population using the urban earnings structure. In contrast, it is clearly negative when applied to the migrant population using the migrant earnings structure. Indeed, the effect of shifting from migrant activity (e.g. self-employment) to urban resident activity (e.g. public sector) increases total earnings, whereas the effect of being more into the public sector decreases income under the migrant wage model. The main reason for this negative effect has been anticipated in section 5: migrants have a comparative advantage in

differences in wage and participation structure (δ, γ, and β) across the two populations, what may in fact belong to an unobserved *composition* effect. As there is usually no natural experiment able to create exogenous variations in group identity, this is a general identification problem in the segmentation/discrimination literature.

[14] There are only two paths for the population decomposition because we have only computed exact decompositions for the sake of internal consistency: population change is computed once changes in all other parameters have been introduced, so that they rely either on all urban parameters or on all migrant parameters.

the private sector or self-employment given their earnings generating structure. Since the urban occupation rules imply fewer people in self-employment and the private sector and more in the public sector, these rules are suboptimal for migrants under the migrant wage structure, and would imply lower incomes on average. Our results corroborate Meng and Zhang (2001) findings of a small impact of occupational segregation on the earnings gap between migrants and urban workers in Shanghai, but they also highlight the possibility of a strong path dependence that reveals interesting patterns.

From Table 7, it seems that using the urban population and the migrant wage model generates a strong positive effect. However, Figure 1 clearly shows that this is an average over a very imprecise simulation, driven by a large right tail. As a result, the activity decomposition lacks robustness mainly because it is extremely path dependent and, to some extent, because of statistical imprecision.

Regarding the hourly earnings effect, Figure 2 shows generally positive effects, although small and quite imprecise when applied to the migrant population using the migrant activity structure. Indeed, the urban resident earnings structure is most of the time more favorable, but the private sector pays better under the migrant model for the migrant population (Table 5). The impact of different hourly earnings structures is small when the private sector has a limited weight (*i.e.* under the urban activity model) but it is strong under the migrant activity model.

Figures A1 to A4 in Appendix 3 show the robustness of these results to the inclusion of selectivity correction in the earnings equations, with the activity decomposition being again very path dependent. The main difference lies in the fact that the distributions have much larger variances within each path, which comes from the fact that the selectivity model is less precisely estimated.

Finally, it should be noted that employment in the public sector encompasses a variety of statuses. For urban residents, 95% are tenured jobs, most of the time as civil servants. However, rural migrants employed in the public sector are mostly under short-term contracts. As such, some of the differences in the public wage structures between urban and migrant employees may result from contrasted employment status and could arguably be interpreted as sectoral effects rather than wage segmentation. Therefore, we have reproduced the whole estimation and simulation procedure with all non-tenured public jobs allocated to the private sector category[15]. As shown in Table A4, the decomposition is not affected by this definition

[15] We are left with only 2% of the migrant population in tenured public jobs (instead of 7%).

change: average effects are similar in magnitude and the small activity effect is again the result of contrasted paths.

7. Conclusion

This paper assesses the sources of the strong income differences between urban residents and long term rural migrants in contemporary urban China, using nationally representative data. In particular, we disentangle the effect of different earnings structures within self-employment, the private and the public sectors from the effect of a different allocation into sectors with different payments. This is important because migrants' access to the public sector is very restricted, so that this could be expected to be an important source of income differences.

A decomposition analysis based on microsimulations indicates that despite a much contrasted sectoral allocation, the impact of sector allocation on earnings differences is neither strong nor robust. We find a stronger, but only partly robust within sector earnings discrimination effect between urban residents and rural migrants. Explanations of these results can be found in the fact that the sector allocation and, to a lesser extent, the earnings effects, are path dependent in the decomposition because rural migrants have a comparative advantage into self-employment and the private sector: shifting into the public sector is not always advantageous to them, whereas it is for urban residents or using the urban resident earnings structure. This result may cast doubt on the literature that concludes to limited sectoral effects on wage contrasts without checking path dependence.

Our findings on a segmented labor market between urban residents and rural migrants that reflects different comparative advantages are consistent with previous studies based on smaller datasets. Using data on 2,900 migrants surveyed in 1995 in 118 enterprises located in four cities, Knight *et al.* (1999) find that urban residents and rural migrants are highly imperfect substitutes in urban firms' production function. Rural migrants are both able to bear hardships and easily manageable, two "assets" that make them accept jobs that non-migrants would not. Using data on 1,500 migrants in Jinan city in 1995, Meng (2001) finds that among rural migrants, those who possess higher human capital are more likely to be self-employed in the informal sector and are better off than those who work in the formal sector, which is consistent with our findings.

Are rural migrants second-class workers in urban China? Our analysis suggests that segregation in terms of both access to jobs and on-the-job earnings is not the major explanation for a large share of earnings differences between urban residents and rural migrants in 2002 as compared to differences in endowments. The strongest source of earnings differences is indeed found to be related to differences in population structures. The two populations are substantially different: rural migrants are younger, much less experienced and much less educated than urban residents. Pre-market discrimination, resulting mainly from lower education opportunity in rural areas, is thus more important in explaining earnings differences than any form of on-market discrimination resulting from sector allocation or earnings generating processes. The key policy implication of this result is to emphasize the importance of public policies towards rural education in order to reduce the endowment gap between rural migrants and urban residents in the urban labor market.

References

Appleton, Simon, John Knight, Lina Song and Qingjie Xia (2004) "Contrasting Paradigms: Segmentation and Competitiveness in the Formation of the Chinese Labor Market", *Journal of Chinese Economic and Business Studies* 2(3): 185-205.

Bourguignon, François, Martin Fournier and Marc Gurgand (2007) "Selection bias corrections based on the multinominal logit model: Monte-Carlo comparisons", *Journal of Economic Surveys* 21(1): 174-205.

Brandt, Loren and Carsten A. Holz (2006) "Spatial Price Differences in China: Estimates and Implications", *Economic Development and Cultural Change* 55(1): 43-86.

Brown, Randall S., Marilyn Moon and Barbara S. Zoloth (1980) "Incorporating occupational attainment in studies of male-female differentials", *Journal of Human Resources* 15(1): 3-28.

Chen, Yi, Sylvie Démurger and Martin Fournier (2005), "Earnings differentials and ownership structure in Chinese enterprises", *Economic Development and Cultural Change* 53(4): 933-958.

de Mel Suresh, David J. McKenzie and Christopher Woodruff (2008) "Measuring Microenterprise Profits: Must we ask how the sausage is made?", *Journal of Development Economics*.

Khan, Azizur R. and Carl Riskin (2005) "China's household income and its distribution, 1995 and 2002", *The China Quarterly* 182: 356-384.

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

Knight, John, and Linda Yueh (2004a) “Job mobility of residents and migrants in urban China”, *Journal of Comparative Economics* 32: 637-660.

Knight, John, and Linda Yueh (2004b) “Urban insiders versus rural outsiders: Complementarity or competition in China’s urban labor market?”, Department of Economics, *Discussion Paper Series 217*, University of Oxford.

Li, Shi (2007) “Rural Migrant Workers in China: Scenario, Challenges and Public Policy”, School of Economics and Business, Beijing Normal University.

Li, Shi, Chuliang Luo, Zhong Wei and Ximing Yue (2007) “1995 and 2002 Household Surveys: Sampling and Data description”, in B. Gustafsson, Li Shi and Terry Sicular (eds.), *Inequality and Public Policy in China*. Cambridge University Press, forthcoming.

Liu, Pak-Wai, Junsen Zhang and Shu-Chuen Chong (2004) “Occupational segregation and wage differentials between natives and immigrants: evidence from Hong Kong”, *Journal of Development Economics* 73: 395-413.

McFadden, Daniel (1973) “Conditional logit analysis of qualitative choice behavior”, pp. 105–142 in P. Zarembka (ed.), *Frontiers in Econometrics,* Academic Press, New York.

Meng, Xin (2001) “The Informal Sector and Rual-Urban Migration – A Chinese Case Study”, *Asian Economic Journal* 15(1): 71-89.

Meng, Xin and Junsen Zhang (2001) “The Two-Tier Labor Market in Urban China – Occupational Segregation and Wage Differentials between Urban Residents and Rural Migrants in Shanghai”, *Journal of Comparative Economics* 29: 485-504.

National Bureau of Statistics (2007) *Situation of rural migration in China in 2006*. Internal Research Report No. 18.

Project Team of Research Office, State Council (2006) *Research Report of Rural Migrant Workers in China*, China Yanshi Press.

Yao, Yang (2001) “Social Exclusion and Economic Discrimination: The Status of Migrants in China’s Coastal Rural Areas”, Working Paper No. E2001005. Beijing: China Center for Economic Research, Peking University.

Zhao, Zhong (2005) “Migration, Labor Market Flexibility, and Wage Determination in China: A Review”, *The Developing Economies*, 43(2): 285-312.

Table 1 – Earnings, working time and activity of urban residents and rural migrants in 2002

	Urban residents		**Rural migrants**		*Mean test*
Rural migrants	*Mean or %*	*Std. Dev.*	*Mean or %*	*Std. Dev.*	
Annual earnings	11,881	7,713	9,335	13,387	**
Self-employed	10,223	12,015	10,077	11,536	NS
Public sector	12,742	7,435	6,967	5,139	**
Private sector	9,639	7,061	8,625	11,929	*
Hourly wage	5.60	4.34	2.83	4.03	**
Self-employed	3.64	4.23	2.86	3.89	**
Public sector	6.14	4.51	2.64	2.16	**
Private sector	4.33	3.38	2.82	4.50	**
Weekly working time	44	11	69	19	**
Self-employed	61	20	74	17	**
Public sector	42	8	57	18	**
Private sector	46	12	64	18	**
Activity					
Self-employed	4.16%		56.97%		
Public sector	71.47%		7.08%		
Private sector	24.37%		35.95%		
# of obs.	4,978		3,035		

Notes: The mean test column indicates the significance level of mean differences between urban residents and migrants. NS non significant, * significant at 5%; ** significant at 1%.

Source: CHIP data, authors' calculation.

Table 2 – Individual characteristics of urban residents and rural migrants in 2002

	Urban residents		**Rural migrants**	
	Mean or %	*Std. Dev.*	*Mean or %*	*Std. Dev.*
Years of schooling	11.72	2.88	7.97	2.71
City work experience	20.71	9.73	7.15	5.01
Age	40.71	9.07	34.31	8.09
Male	56.33%		56.51%	
Ethnic minority	3.42%		8.27%	
Communist membership	29.07%		3.13%	
Married	87.06%		89.95%	
Household size	3.12	0.68	2.77	0.91
Long-term tenure	74.16%		5.09%	
Professional or technicians	32.74%		4.48%	
Office workers	19.62%		2.17%	
Workers	42.01%		32.52%	
Self-employed	4.16%		56.97%	
# of obs.	4,978		3,035	

Notes: Mean differences between urban residents and migrants are all significant at 1% level, except for the "male" variable. "City work experience" is measured as actual number of years of work in urban areas.

Source: CHIP data, authors' calculation.

Table 3 – Estimation of the Multinomial Logit Model for occupational choice

	Public sector *versus* **Self-employment**		**Private sector** *versus* **Self-employment**		**Private sector** *Versus* **Public sector**	
	Coef.	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
Urban residents						
Years of schooling	0.255**	*0.027*	0.112**	*0.028*	-0.143**	*0.014*
Age	-0.072	*0.092*	-0.064	*0.095*	0.008	*0.037*
Age^2	0.002	*0.001*	0.001	*0.001*	-0.001	*0.000*
Coast	0.423**	*0.197*	0.692**	*0.203*	0.268**	*0.082*
West	0.039	*0.171*	0.086	*0.179*	0.047	*0.088*
Spouse of the household head	-0.480**	*0.165*	-0.347**	*0.172*	0.133*	*0.079*
Married	-0.357	*0.395*	-0.794**	*0.403*	-0.437**	*0.150*
# of children below 6	-0.107	*0.258*	-0.182	*0.270*	-0.075	*0.132*
Household size	-0.133	*0.108*	-0.008	*0.112*	0.125**	*0.052*
Communist membership	0.931**	*0.230*	0.234	*0.241*	-0.697**	*0.095*
Male	-0.559**	*0.164*	-0.578**	*0.170*	-0.019	*0.073*
Constant	1.303	*1.727*	2.448	*1.766*	1.145	*0.719*
Rural migrants						
Years of schooling	0.072**	*0.030*	0.035**	*0.016*	-0.037	*0.031*
Age	-0.256**	*0.065*	-0.274**	*0.039*	-0.019	*0.065*
Age^2	0.004**	*0.001*	0.003**	*0.001*	0.000	*0.001*
Coast	1.655**	*0.202*	0.326**	*0.099*	-1.329**	*0.204*
West	1.251**	*0.203*	-0.303**	*0.098*	-1.554**	*0.209*
Spouse of the household head	-0.201	*0.191*	0.165	*0.102*	0.366*	*0.197*
Married	-0.404	*0.309*	-0.408**	*0.176*	-0.004	*0.306*
# of children below 6	0.009	*0.217*	-0.082	*0.109*	-0.090	*0.223*
Household size	-0.223**	*0.088*	-0.055	*0.048*	0.168**	*0.090*
Communist membership	1.020**	*0.307*	0.123	*0.247*	-0.897**	*0.323*
Male	0.168	*0.181*	0.108	*0.099*	-0.060	*0.186*
Constant	1.505	*1.144*	4.885**	*0.659*	3.380**	*1.134*

# of observations	*Urban residents*: 4,978	*Rural migrants*: 3,035
Log likelihood	*Urban residents*: -3,309	*Rural migrants*: -2,495

Notes: * significant at 10%; ** significant at 5%.

Source: CHIP data, authors' calculation.

Table 4 – Observed and simulated occupational distribution of urban residents and rural migrants in 2002

	Urban residents		**Rural migrants**	
	Frequency	*%*	*Frequency*	*%*
# of obs.	4,978		3,035	
Occupation				
Observed				
Self-employed	207	4.2	1,729	57
Public sector	3,558	71.5	215	7.1
Private sector	1,213	24.4	1,091	36
Simulated with the other group model				
Self-employed	2,499	50.2	328	10.8
Public sector	712	14.3	1,582	52.1
Private sector	1,767	35.5	1,125	37.1

Source: CHIP data, authors' calculation.

Table 5 – Counterfactual hourly earnings and working hours structures under various models by population group

Panel A: Counterfactual hourly earnings structure under various models by population group

	Urban resident population (n=4,978)			***Migrant population (n=3,035)***	
	Urban resident earnings model	*Migrant earnings model*		*Urban resident earnings model*	*Migrant earnings model*
Self-employment model	3.90	3.11		3.08	2.90
Public wage model	5.92	5.00		3.45	2.59
Private wage model	4.93	3.52		2.61	2.83

Panel B: Counterfactual weekly working hours structure under various models by population group

	Urban resident population (n=4,978)			***Migrant population (n=3,035)***	
	Urban resident wage model	*Migrant wage model*		*Urban resident wage model*	*Migrant wage model*
Self-employment model	61.2	71.4		63.9	74
Public wage model	41.9	51.1		44.1	55.5
Private wage model	45.7	59.5		50.2	64.4

Source: CHIP data, authors' calculation.

Table 6 – Decomposition of observed earnings gaps between urban residents and rural migrants, 2002

Means over all observed paths

	(Yuan)	(%)
Observed earnings gap	2,546	
Activity effect	-32	-2%
Hourly wage effect	1,162	+46%
Working time effect	-2,068	-82%
Population effect	3,487	+138%

Note: means over 200 bootstrapped replications – see Figures 1-4 for the full distributions.

Source: CHIP data, authors' calculation.

Table 7 – Path dependence for the activity effect

Means by path

Population	**Wage/hours model**	**Activity effect**	
		(Yuan)	*(% of observed)*
Urban residents	Urban residents	451	18%
Urban residents	Rural migrants	1,376	53%
Rural migrants	Rural migrants	-1,270	-51%
Rural migrants	Urban residents	-689	-27%

Note: means over 200 bootstrapped replications – see Figures 1-4 for the full distributions.

Source: CHIP data, authors' calculation.

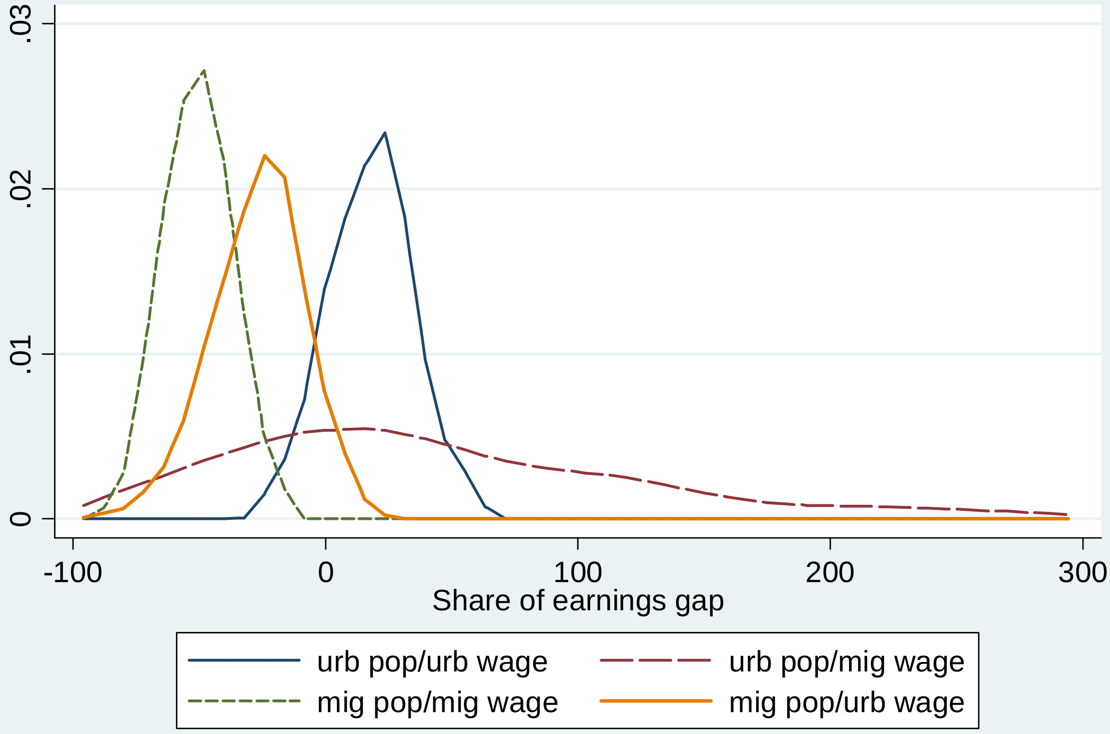


Figure 1: Distribution of activity decompositions

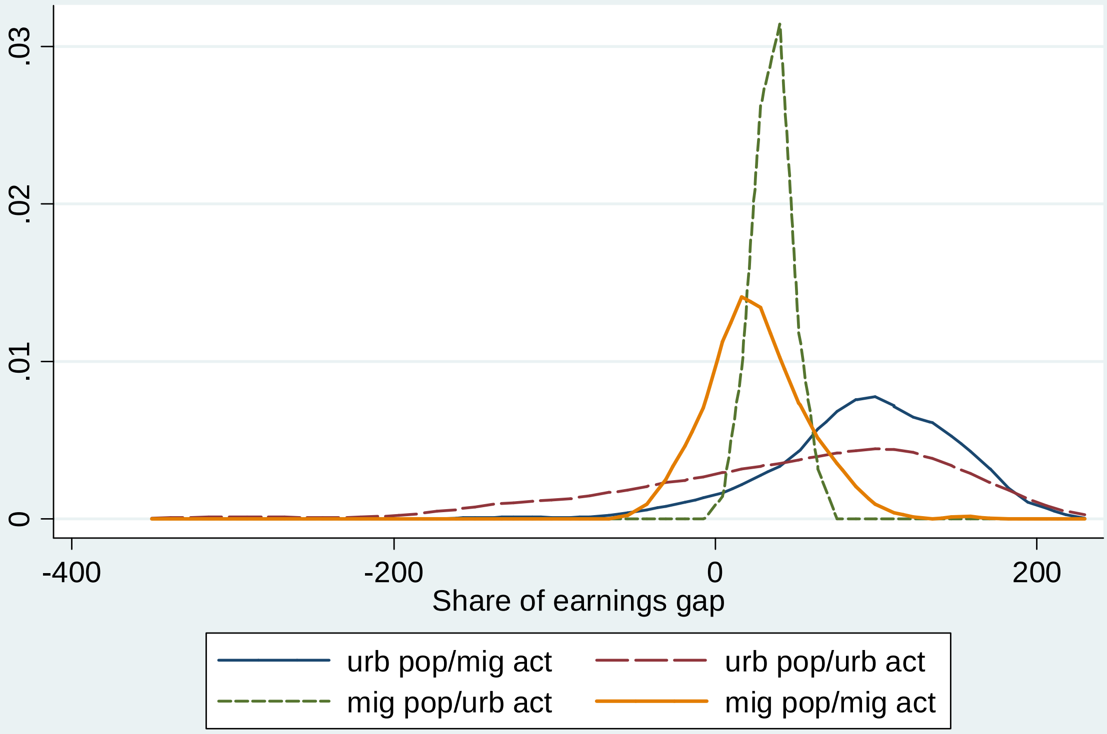
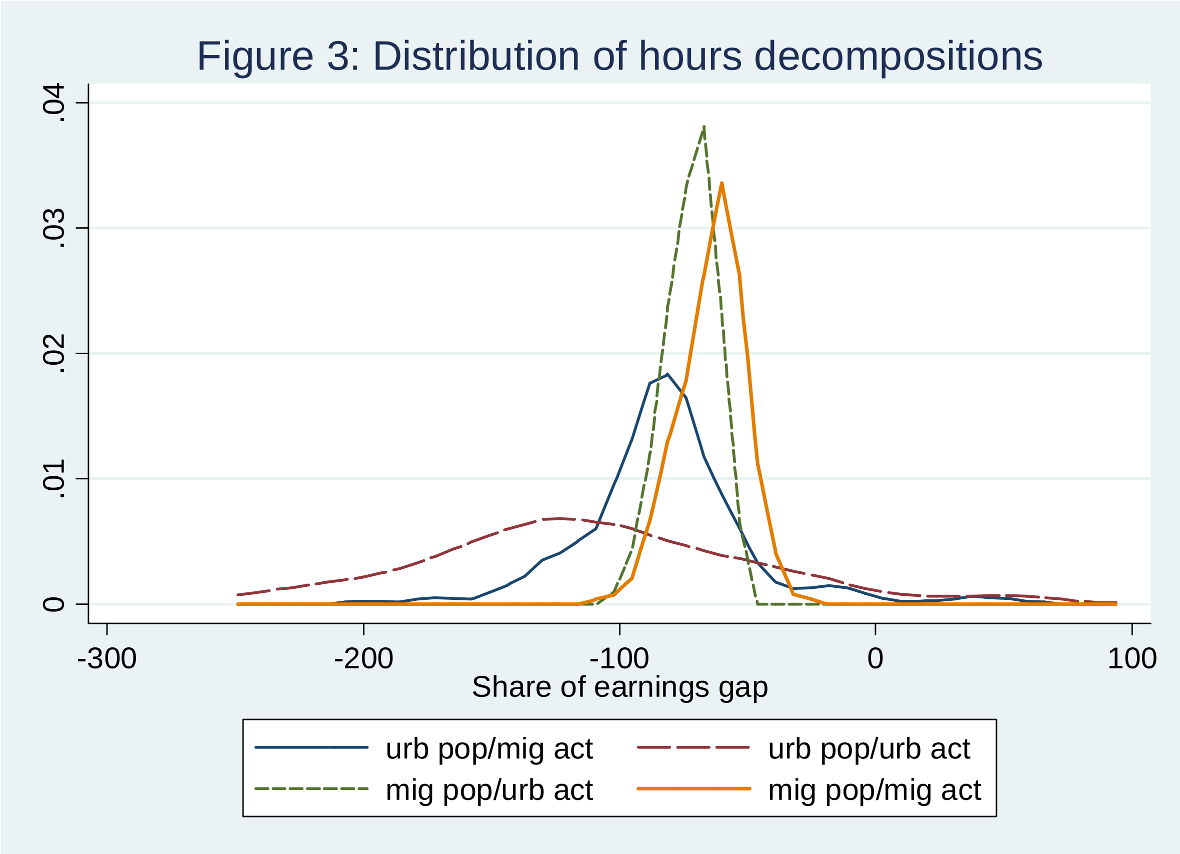
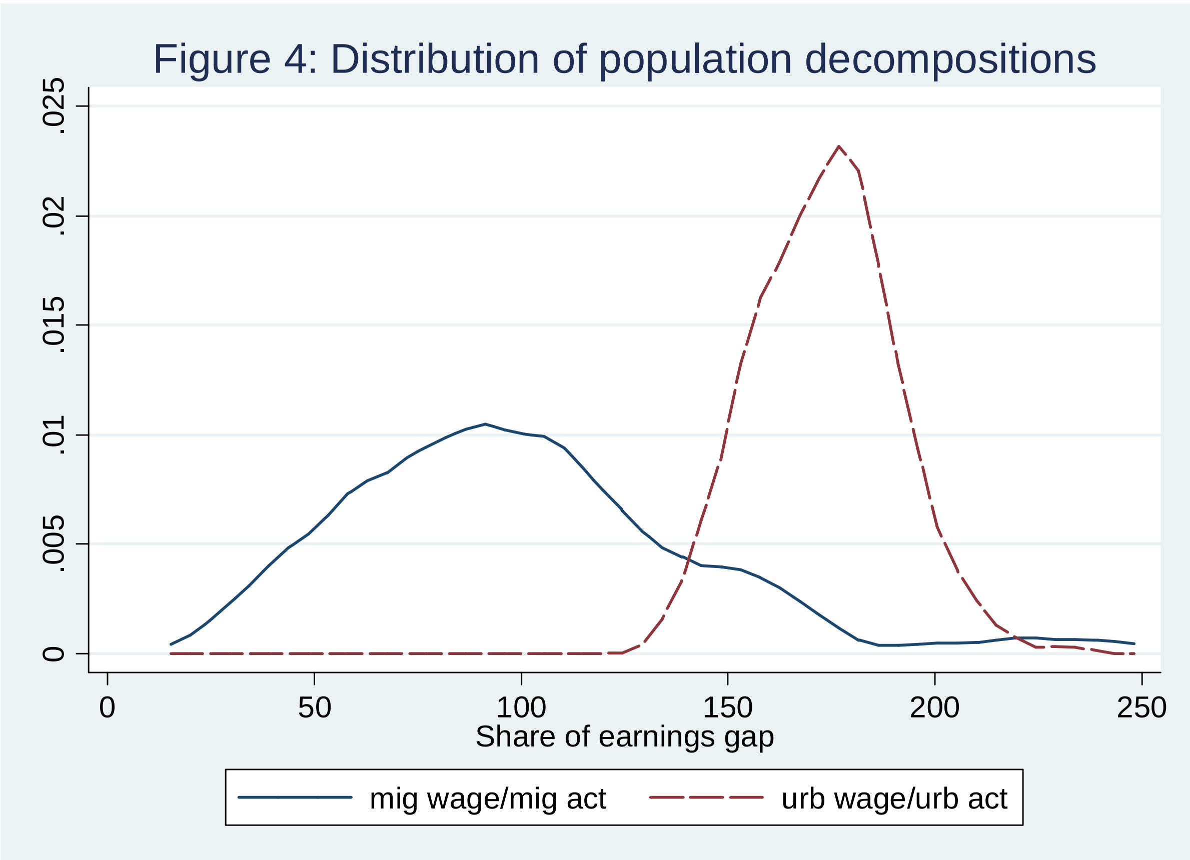


Figure 2: Distribution of hourly wage decompositions





Source: CHIP data, authors' calculation.

Appendix 1

Table A1 - Earnings equations

	Self-employment		Public sector		Private sector	
	Urban residents	***Rural migrants***	***Urban residents***	***Rural migrants***	***Urban residents***	***Rural migrants***
Years of schooling	0.042	0.037**	0.063**	0.073**	0.074**	0.036**
	1.64	*5.98*	*17.38*	*5.38*	*8.75*	*4.37*
City work experience	0.008	0.059**	0.048**	0.005	0.041**	0.063**
	0.27	*5.96*	*11.71*	*0.2*	*5*	*6.03*
City work experience^2	0.0001	-0.0020**	-0.0009**	0.0007	-0.0007**	-0.0021**
	-0.16	*-4.1*	*-8.98*	*0.55*	*-3.08*	*-4.41*
Coast	0.144	0.181**	0.210**	-0.173	0.397**	0.180**
	0.88	*4.22*	*9.13*	*-1.57*	*7.92*	*3.62*
West	-0.332*	-0.026	0.079**	-0.285*	0.115*	0.003
	-2.41	*-0.69*	*3.26*	*-2.55*	*2.1*	*0.06*
Communist	0.010	0.057	0.156**	-0.053	0.288**	0.504**
	0.05	*0.56*	*7.08*	*-0.37*	*4.5*	*3.97*
Male	0.390**	0.176**	0.109**	0.231*	0.140**	0.318**
	3.15	*5.21*	*5.42*	*2.61*	*3.21*	*7.43*
Constant	0.223	0.054	0.128*	0.066	-0.369**	-0.097
	0.59	*0.84*	*1.99*	*0.39*	*-2.88*	*-1.11*
Observations	*207*	*1,729*	*3,558*	*215*	*1,213*	*1,091*
R^2	0.12	0.10	0.18	0.25	0.18	0.14

Notes: 1. The dependent variable is the logarithm of total earnings.
2. The reference category for regional location is the central region.
3. Numbers in italics are t-statistics. * Significant at 5 percent level ** Significant at 1 percent level

Source: CHIP data, authors' calculation.

Appendix 2

Table A2 – Working hours equations

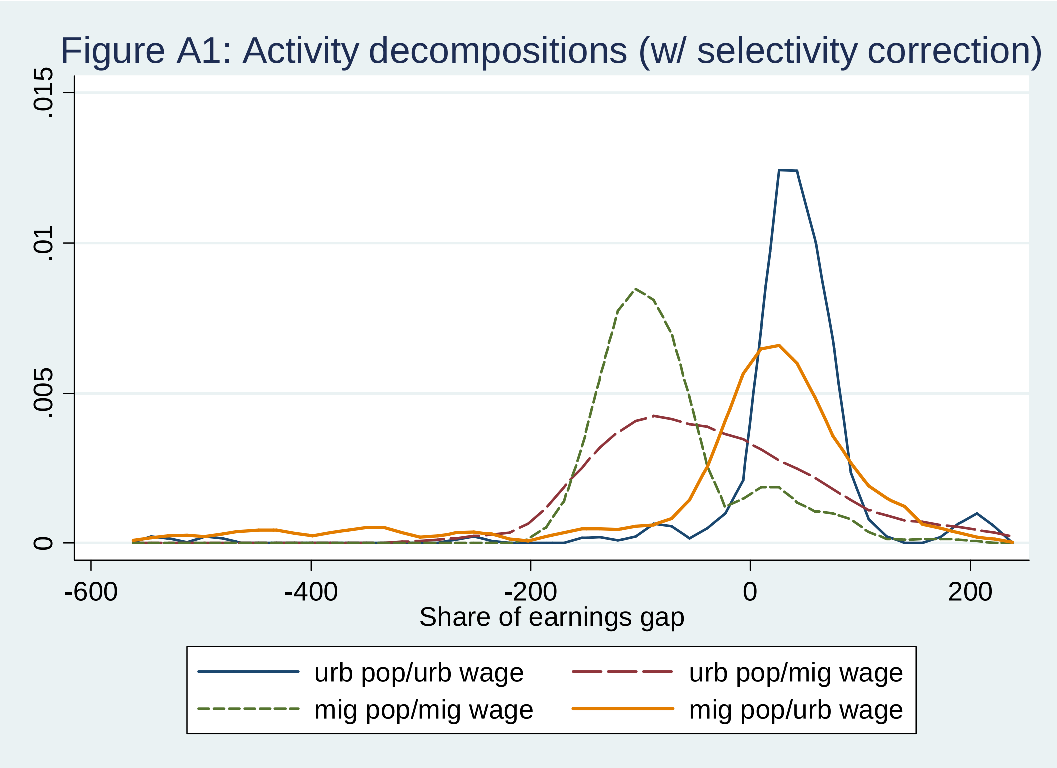
	Self-employment		Public sector		Private sector	
	Urban residents	*Rural migrants*	*Urban residents*	*Rural migrants*	*Urban residents*	*Rural migrants*
Years of schooling	20.72	-1.17	-5.69*	-23.87	-23.96**	-48.24**
	0.68	*-0.14*	*-2.29*	*-1.09*	*-3.29*	*-4.31*
City work experience	-10.38	-2.66	-18.96**	-22.21	-14.59	1.39
	-0.31	*-0.20*	*-5.09*	*-0.55*	*-1.59*	*0.10*
City work experience^2	0.06	-0.03	0.37**	0.39	0.19	-0.12
	0.06	*-0.04*	*4.45*	*0.20*	*0.84*	*-0.20*
Coast	245.74	65.21	-57.80**	23.94	-209.18**	-207.54**
	1.28	*1.15*	*-3.64*	*0.14*	*-4.91*	*-3.13*
West	687.07**	363.46**	-25.05	488.37*	-124.04*	93.50
	4.29	*7.39*	*-1.51*	*2.73*	*-2.69*	*1.31*
Communist	-184.32	-242.90	0.17	-75.73	-141.00*	-250.93
	-0.80	*-1.80*	*0.01*	*-0.34*	*-2.59*	*-1.49*
# of children below 6	74.83	-180.73**	-3.71	-255.09	-70.85	16.53
	0.34	*-3.45*	*-0.14*	*-1.52*	*-1.00*	*0.23*
# of children at school	-101.32	-35.46	18.42	34.91	13.76	123.88*
	-0.81	*-1.14*	*1.25*	*0.31*	*0.33*	*2.68*
Married	-314.62	329.35**	66.25**	-97.82	68.87	-95.88
	-1.18	*3.49*	*2.18*	*-0.46*	*0.99*	*-1.18*
Male	-43.50	-4.83	64.55**	3.78	182.09*	-61.26
	-0.30	*-0.11*	*4.67*	*0.03*	*4.94*	*-1.08*
Constant	3274.73**	3514.66**	2362.30**	3185.18**	2817.03**	3841.76**
	6.58	*28.63*	*53.06*	*10.32*	*25.65*	*29.03*
Observations	*207*	*1,729*	*3,558*	*215*	*1,213*	*1,091*
R^2	0.11	0.05	0.02	0.10	0.06	0.05

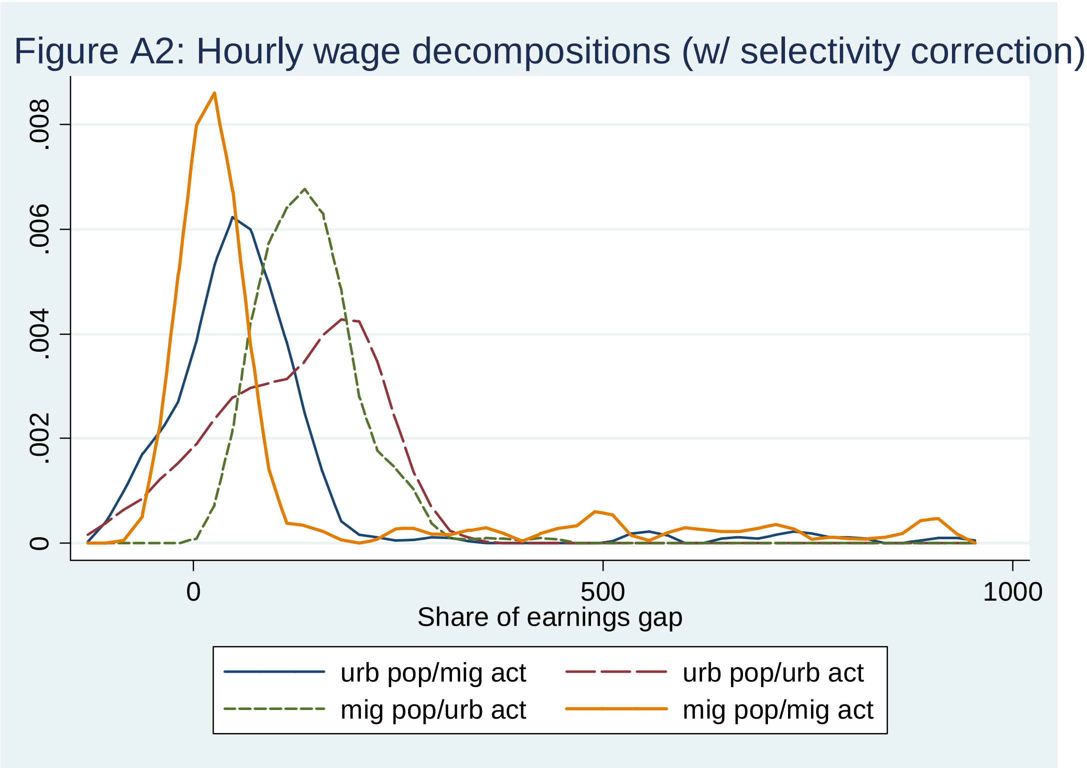
Notes: 1. The dependent variable is the number of hours worked in a year.
2. The reference category for regional location is the central region.
3. Numbers in italics are t-statistics. * Significant at 5 percent level ** Significant at 1 percent level

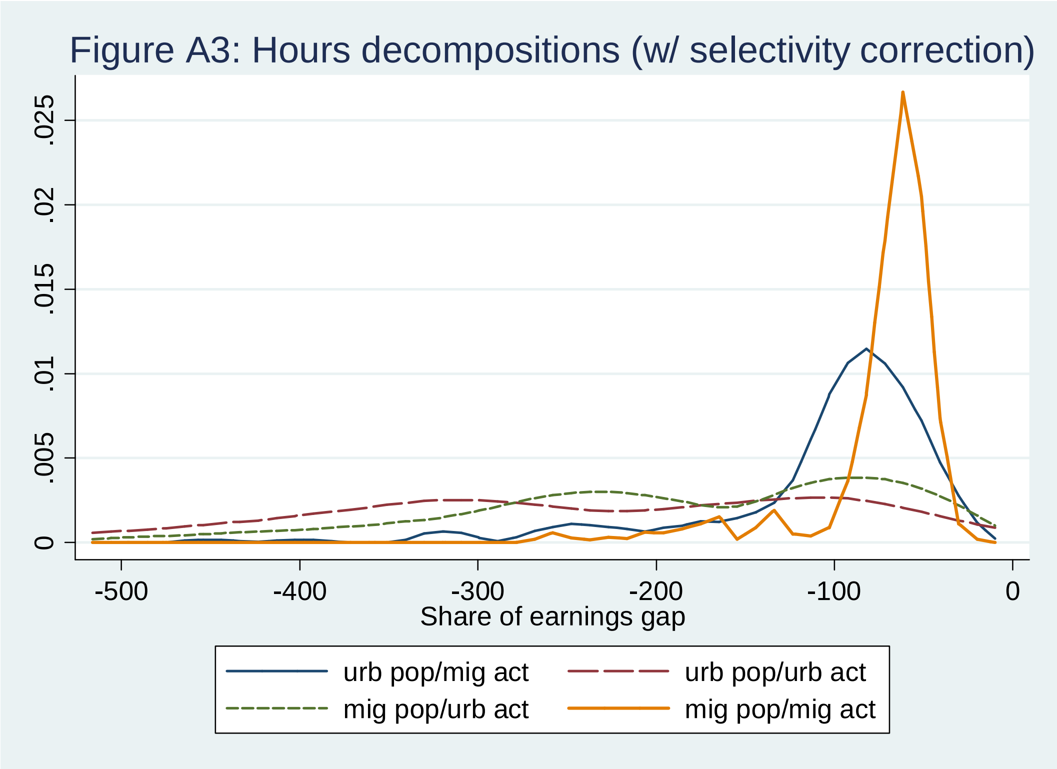
Source: CHIP data, authors' calculation.

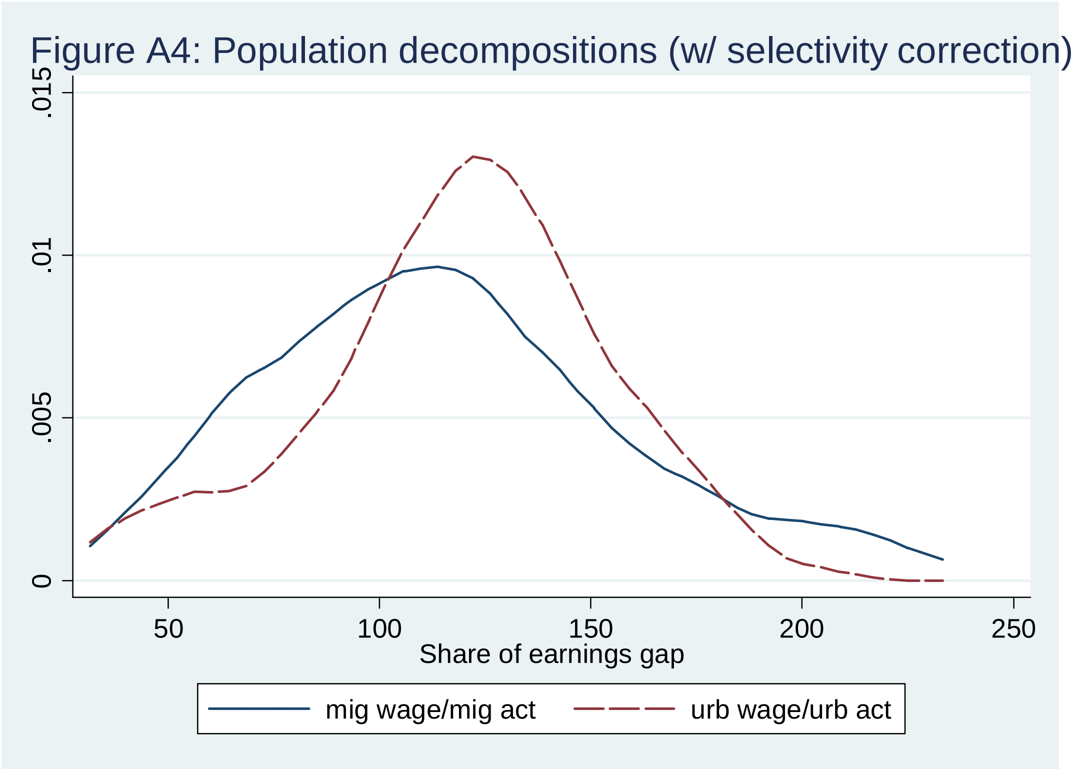
Appendix 3: Decomposition of observed earnings gaps

with selectivity correction









Source: CHIP data, authors' calculation.

Appendix 4

Table A4 – Decomposition of observed earnings gaps with public-tenure between urban residents and rural migrants, 2002

Means over all observed paths

	(Yuan)	(%)
Observed earnings gap	2,546	
Activity effect	-86	-3.5%
Hourly wage effect	1,837	67%
Working time effect	-1,835	-67%
Population effect	2,828	104%

Note: means over 200 bootstrapped replications.

Source: CHIP data, authors' calculation.