CENTRE D'ETUDES
ET DE RECHERCHES
SUR LE DEVELOPPEMENT
INTERNATIONAL

Document de travail de la série

Etudes et Documents

E 2012.05

Are workers close to cities paid higher non-agricultural wages in rural China?

Chloé DUVIVIER, Shi LI, Mary-Françoise RENARD

February, 2012

(*Revised November, 2012*)

CERDI
65 BD. F. MITTERRAND
63000 CLERMONT FERRAND - FRANCE
TEL. 04 73 71 74 20
FAX 04 73 17 74 28
www.cerdi.org

Les auteurs

Duvivier Chloé

Doctorant / PhD Student

Auvergne University, CNRS, UMR 6587, CERDI, F-63009 Clermont Fd.

Li Shi

Professeur / Professor

School of Economics and Business Administration, Beijing Normal University, China

Renard Mary-Françoise

Professeur / Professor

Auvergne University, CNRS, UMR 6587, CERDI, F-63009 Clermont Fd.

Corresponding author: Chloé Duvivier

Email : Chloe.DUVIVIER@u-clermont1.fr

La série des Etudes et Documents du CERDI est consultable sur le site :

http://www.cerdi.org/ed

Directeur de la publication : Patrick Plane

Directeur de la rédaction : Catherine Araujo Bonjean

Responsable d'édition : Annie Cohade

ISSN : 2114 7957

Résumé / abstract

In the present study, we investigate whether workers close to cities are paid higher non-agricultural wages than workers in outlying rural areas. We find that workers close to urban areas not only benefit from more opportunities to engage in non-agricultural activities, but also from better paid jobs. In addition, distance exhibits a strongly nonlinear impact. Distance always has a negative impact on wages but the effect is more detrimental, the closer the village is to the urban center. We also find evidence of urban hierarchy effects: workers living close to bigger cities are paid higher wages. Finally, we provide evidence on the transmission channels at work.

Mots clés /Key words : Remoteness, wage differentials, regional labor market, China

Codes JEL / JEL codes : J31, O53, R10, R23

Remerciements

We thank J-L. Combes, S. Démurger, M. Fafchamps, S. Guillaumont-Jeanneney, H. Guo, Z. Peng, M. Xueting, N. Zihan and the participants of the 8^{th} International Conference on the Chinese Economy, 23^{rd} Chinese Economic Association Annual Conference, Thematic Meeting of the French Economic Association "Economic Geography and Public Policies", 15^{th} Applied Economics Meeting, 4^{th} GREThA International Conference on Economic Development and of the 4^{th} Annual Joint Workshop on Socio-Economics for their constructive comments.

I. Introduction

It is widely recognized that non-agricultural employment enables rural households to get out of poverty and enhance the welfare of households. Indeed, non-agricultural work can enable households both to raise their income and to reduce instability (Ellis, 1998). This is particularly true for China, where farm size is extremely small and thus, where farmers have few opportunities to generate agricultural income.

Over the last thirty years, non-agricultural employment has developed unevenly across rural China, leading to a significant increase in intra-rural inequality (Kung and Lee, 2001). Specifically, as rural areas close to cities benefit from low transport costs and from significant transmission of ideas, they benefit from more non-agricultural employment opportunities than areas further away from cities (Henderson *et al.*, 2001). Thus, the closer the city, the higher the probability to engage in the non-agricultural sector (Knight and Song, 2003).

The present work aims at studying more deeply how urban proximity affects non-agricultural employment, by investigating whether rural workers closer to cities are paid higher non-agricultural wages. Therefore, unlike previous studies, our focus is not on the *level* but on the *kind* of non-agricultural employment that rural workers manage to get according to their location. Examining this issue in China is particularly relevant for three reasons. First, given the institutional restrictions on labor mobility, local conditions play a very significant role in determining rural workers' earnings and well-being (Xia and Simmons, 2004). Second, if on average non-agricultural activities are much more income-generating than agricultural activities, there is a significant variation in the remuneration of non-agricultural employment. There are even low-paid non-agricultural jobs where earnings are lower than agricultural earnings (Lanjouw, 1999), so that one cannot assume *a priori* that non-agricultural employment enables workers to increase their income. Third, it has been recognized that there are persistent spatial differences in wages, especially between workers in urban areas and in remote areas (Hanson, 2000).

According to the nationally representative 2002 Chinese Household Income Project (CHIP) survey, there are large differences in wages both across rural areas and between suburban[1] and other villages in China. First, the average daily wage was 2.5 times higher in the ninth decile than in the first decile. This data shows that intra-rural and intra-urban wage inequality are of comparable magnitude (see Combes *et al.* (2012) for data on intra-urban wage inequality). Second, suburban villages benefit from higher wages, as the average daily wage in these villages was about 1.25 higher than in other villages.

In spite of these spatial variations in wages in rural China, to our knowledge, there is no empirical evidence on the effect of urban proximity on rural non-agricultural wages in China. However, workers close to cities are likely to engage in more remunerative non-agricultural jobs for two reasons: (i) agglomeration externalities, leading to differences in wages across villages, and (ii) commuting to nearby urban centers, that enables workers to benefit from the higher urban wages. Firstly, nowadays Chinese suburban areas are highly urbanized and with densely concentrated industries (Naughton, 2007). Suburban villages are therefore likely to benefit from some kinds of agglomeration economies, leading to higher labor productivity and so, to higher wages. In addition, villages close to cities benefit from a large market potential. Thus, firms in these villages, which enjoy lower transport costs to reach their consumers, can afford to pay higher wages. Previous studies have highlighted that market potential plays a major role in determining wages in cities (Hering and Poncet, 2010) and regional growth (Bai *et al.*, 2012) in China. Market potential should also play a crucial role in determining rural wages given that rural non-agricultural production is closely tied to urban production, through subcontracting and technical assistance to urban firms. Secondly, workers close to cities are likely to benefit from higher wages because of commuting. Indeed, workers close to urban areas are much more likely to commute to nearby urban centers and thus, to benefit from the higher level of wages that is paid in urban areas.

Using data from the 2002 CHIP survey, we investigate whether workers close to cities are paid higher wages. We make two main contributions to the existing literature. First, we highlight that rural workers close to cities benefit from higher wages than workers in outlying rural areas. This issue has been largely ignored in the literature on spatial dispar-

[1] The municipal area (*shixiaqu*) of a city is composed by the urban core (*jianchengqu*) and the periphery or the suburb (*jiaoqu*). Suburban villages refer to villages located in the suburb of a middle-sized or large city (*jiaoqu*).

ities in China, which mainly focuses on disparities either between urban and rural areas or within urban areas. Second, to our knowledge, we offer the most comprehensive study on the impact of urban proximity on rural earnings in China. We find very robust evidence that workers close to cities are paid significantly higher wages. In addition, the closer to the urban center, the more detrimental is the impact of distance on wages. Workers closer to the biggest cities are also found to benefit from the highest wage premium. Finally, workers close to cities manage to engage in better remunerated jobs because they benefit of both higher wages in their villages and higher opportunities to commute.

The rest of the study proceeds as follows. Section II presents the data and Section III the methodology used. We describe the results in Section IV and finally, we conclude in Section V.

II. Data

To carry out the empirical analysis, we use the 2002 rural survey of the CHIP[2]. The database is composed both of an individual, a household and a village level survey. Thus, we benefit from detailed information on individual labor allocation and from household and village characteristics. In addition, this is a nationally representative survey which investigates 37 969 individuals of 9200 households from 961 villages belonging to 122 counties (*xiàn*) of 22 provinces[3]. As a result, compared to most microeconomic studies on rural areas in developing countries, we benefit from a great range of variability in terms of remoteness-proximity to urban areas.

Labor allocation of workers in the sample

We restrict the CHIP sample to workers. Every individual above 15 years old, who reports having earned some income or having spent some time working, is considered as a worker. We have classified rural workers according to their primary activity[4] in one of

[2] We do not use the 2007 CHIP survey as there is no detailed information on rural non-agricultural work to calculate hourly wages.

[3] There are five levels of administrative divisions in China, namely, from the highest to the lowest: province, prefecture, county, township and village.

[4] The primary activity is the activity to which the worker devotes most, if not all, of his working time. Many workers also declared having a secondary activity, which is an activity to which they devote a

the four following categories: (1) Local agricultural workers, (2) Local non-agricultural wage earners, (3) Local non-agricultural self-employed and (4) Migrant workers. Local agricultural workers are individuals whose primary activity consists of working on the family farm or as a farm-employee. Local non-agricultural wage earners include workers who spend most of their time working out of agriculture as wage earners. Local non-agricultural self-employed are workers who are self-employed in the non-agricultural sector. The three previous categories only include local workers, *i.e.* individuals working in their home county. On the contrary, we considered as migrant every individual whose primary activity takes place out of its home county (Zhao, 1999). Indeed, given the size of counties, it is impossible for a worker to commute from a county to another county. This ensures that individuals working out of their home county, *i.e.* migrants, are both working and living in towns and cities. On the contrary, this criteria ensures that commuters, who work out of their village but who come back to their home village every day, are classified as local workers (de la Rupelle *et al.*, 2010)[5]. As stated in the introduction, workers close to cities are expected to be paid higher wages partly because they can commute to cities and thus, benefit from the higher wages that are paid in urban areas. Thus, to capture the entire effect of urban proximity, commuters must be classified in the local workers category. Table 1 presents the classification of workers in our sample. Our sample is composed of 22 551 workers[6] and 4530 local non-agricultural wage earners.

[Table 1]

smaller part of their time. As some workers have both agricultural and local non-agricultural or migratory work, we have classified individuals according to their primary activity so that each worker belongs to only one category. The worker's primary activity has been demonstrated to be the most relevant criteria to classify rural workers with multiple activities (Deichmann *et al.*, 2009).

[5]Consistently, in our sample 75% of the local workers spent less than 14 days out of their households during the year, whereas 75% of migrants spent more than 180 days out of their households.

[6]On the 37 969 individuals surveyed, 7869 are children and 30 100 are adults. 26 065 adults are workers and 4035 are inactive. However, we have missing information on place of work, labor time and/or wage for 949 workers. Finally, there are 2565 individuals for whom explanatory variables are missing. As a result, our sample is composed of 22 551 workers.

Non-agricultural hourly wage

To study whether workers close to cities are engaged in better paid non-agricultural jobs, we focus on *local* non-agricultural workers. As described above, local non-agricultural work is composed by wage earners and self-employed workers. However, most information on labor time and earnings is not available at the individual level for self-employed workers. Thus, the present study focuses on local non-agricultural wage earners.

The explained variable is the individual non-agricultural hourly wage (hereafter NAHW). Another option could be to use annual non-agricultural earnings. However, annual earnings depend on both the intensity of participation in the non-agricultural sector and on the hourly wage. Given that urban proximity increases the intensity of participation in the non-agricultural sector (Knight and Song, 2003), using annual earnings would lead to over-estimating the effect of urban proximity. As a result, the NAHW is the most appropriate variable. This variable is calculated as:

$$NAHW_i = \frac{W_i}{D_i * H_i} \tag{1}$$

W_i is the annual wage[7] earned by individual i; D_i is the number of days worked during the year and H_i the number of hours worked per day. Both W_i, D_i and H_i refer to the worker's primary activity.

Variables of interest

The relationship between urban proximity and wages is likely to be characterized by two phenomena: nonlinearity and heterogeneity. First, distance is likely to have a nonlinear impact on wages. Indeed, there is extensive evidence that most urban agglomeration effects disappear quite rapidly across space (Rosenthal and Strange, 2001; Aminiti and Cameron, 2007). Thus, the closer to the urban areas, the more detrimental the impact of the distance should be. In remote areas, where almost all agglomeration effects have disappeared, distance should have a much lower effect, or no effect at all, on wages. Secondly, Partridge *et al.* (2009) have demonstrated that urban hierarchy effects were at work in the determination process of wages; specifically, if wages are higher close to cities, the effect

[7] Following Hering and Poncet (2010) and Démurger *et al.* (2012), this includes the basic wage, bonuses, in-kind earnings and subsidies and pension income.

is the strongest close to the biggest cities because they generate the largest agglomeration effects.

We use two indicators to measure the degree of urban proximity of workers' villages. Both indicators are designed to take into account nonlinearity and urban hierarchy effects. First, in the survey we have data on the number of kilometers between each worker's village and the nearest county seat (*Distance*). To account for the nonlinearity of the effect of distance, we have created four dummy variables (*Quartile*) to indicate which quartile of distance the village is located in ($q_1 = 10km$; $q_2 = 20km$; $q_3 = 30km$; $q_4 = 160km$). Thus, the dummy "$Quartile_1$" is equal to 1 if the village is located within 10 km from the county seat, the dummy "$Quartile_2$" is equal to 1 if the village is located between 10 and 20 km from the county seat and so on. To test whether distance has a non-linear impact on wages, we have introduced in the estimates the *Distance* variable, together with interactive terms between the *Distance* variable and the *Quartile* dummies. These interactive terms enable us to test whether an increase of 1 km in the distance between the county seat and the worker's village has a more detrimental impact on wages close to the county seat.

Second, to test whether wages are highest close to the biggest cities, we use the official codes of the counties available in the dataset to construct the two following variables. *Provincial City* is a dummy equal to 1 if the worker's village is located in the suburb of a provincial city, and 0 otherwise. *Low level City* is a dummy variable equal to 1 if the worker's village is a suburb of a prefecture city, or if it is located in the administrative area of a county-level city, and 0 otherwise. As provincial cities are much bigger and more economically developed than other cities, we expect workers located close to these cities to benefit from the highest wages.

[Table 2]

Table 2 gives descriptive statistics on the hourly wages in yuan according to the distance to urban areas. It appears that wages decrease with the distance to the county seat. In addition, they are significantly higher in suburban villages than in non-suburban villages.

III. Methodology

Selection bias correction based on a multinomial logit model

To test whether workers close to urban centers are paid higher wages, we estimate an income function on the sub-sample of local non-agricultural wage earners. To get unbiased estimates of the coefficients in the income function, we need to correct for the potential selection of workers. In our case, selection is over more than two choices, given that workers choose to engage in one of the following four activities: local agriculture, local non-agricultural wage-employment (hereafter NAWE), local non-agricultural self-employment (hereafter NASE) and migration. Several methods have been proposed to correct for selection bias, when selection is over more than two exclusive choices (Lee, 1983; Dubin and MacFadden, 1984; Dahl, 2002). Essentially, these generalizations of the Heckman procedure, consist in the following two steps. First, a multinomial logit model is estimated, accounting for all the different possible choices. Second, the results of the first step equation are used to compute the appropriate correction terms, which are included as control variables in the second-step earning equation. Bourguignon *et al.* (2007) show that the Dahl (2002) semi-parametric model (with full specification) should be preferred to the other models. Following their recommendation, we use the Dahl's method in the empirical analysis.

Baseline specification

In the first step multinomial logit model, the explained variable takes the four following values, according to the worker's primary activity: 0 if local agriculture, 1 if NAWE, 2 if NASE and 3 if migration. In the second step, we estimate a hourly earnings function by the OLS, by adding the correction terms calculated from the first-step model to the set of explanatory variables[8].

Regarding the income function, to test whether workers close to cities are paid higher wages, we introduce as determinants of the hourly wage the variables of interest described

[8] In the Dahl's model, the correction terms are a polynomial of choice probabilities. Following Bourguignon *et al.* (2007), we use all the probabilities, which are included as a fourth-order polynomial and with all interactions between them.

in Section II. We also introduce a wide range of controls, both at the worker and village levels, which are expected to affect the level of hourly wages. Thus, we control for worker's age and its square, education, experience and its square, gender, ethnic minority and Communist Party membership. We introduce two more variables, at the village level, to control for frictional distance (Bird and Sheperd, 2003): a dummy variable to control for the topography and a dummy variable indicating whether or not a road reaches the village. In addition, as wages are expected to be lower in poorer areas, we introduce a variable indicating whether the village is in a province level poverty township. Regional (East, Center, West) and provincial dummies are introduced to control for differences in development, endowments and policies. These dummies also partially control for living costs. However, living costs are also likely to vary within a given province, and especially between remote rural areas and other ones. As wages are expected to be an increasing function of living costs, and as living costs are expected to be higher close to urban areas, the coefficient associated with the variables of interest could be over-estimated (Hering and Poncet, 2010). To precisely control for living costs, we calculate an index of living costs at the village level, using information on the market price, in yuan per kg, of six non-staple foods (meat, eggs, edible oil, sugar, vegetables, fruit and melons).

As identifying restrictions, we use the quantity of land per capita in the worker's household and a dummy indicating whether the worker is unmarried. These variables are assumed to affect the participation choice of the worker but not his wage. Tests of joint significance are carried out to assess the validity of the identifying restrictions. Definition of the variables are given in Appendix 1 at the end of the paper[9].

IV. Results

Baseline results

Table 3 presents the baseline estimates of the Dahl's model. The first three columns give the results of the multinomial logit model. The reference category is made up of local agricultural workers. The results of the income equations are reported in columns (4), (5)

[9]Descriptive statistics are provided in the supplementary Appendix S.1, which is available upon request.

and (6). In Column (4), we only control for living costs by introducing provincial level dummies, whereas in Column (5) we add the index of living costs at the village level to the set of control variables. Finally, in Column (6) we use the index of living costs to calculate the real hourly wage, which is used as explained variable instead of the nominal wage.

The selection correction terms enter the income equation significantly, suggesting that the selection model is appropriate[10]. In addition, tests of join significance show that the identifying restrictions are valid.

The estimation results of the participation model are very consistent with previous findings (Xia and Simmons, 2004; Liu and Sicular, 2009). The variables of interest exhibit a different impact on local non-agricultural employment and migration, as shown by Knight and Song (2003). Workers close to cities have a higher probability of working locally out of agriculture due their more developed non-agricultural sector and to commuting opportunities.

Turning to the income equation, it appears that, as for urban areas (Hering and Poncet, 2010; Démurger *et al.*, 2012), hourly wages in rural China are an increasing function of a worker's age, education and experience. Regarding our indicators of interest, the distance[11] to the county seat has a negative impact on wages, whereas living in a suburban village significantly increases wages. Interestingly, distance exhibits a strongly nonlinear impact. A 1 km increase in the distance between a worker's village and the county seat has a significantly stronger impact on wages within 10 km from the county seat (first quartile). The effect is the lowest for villagers located at more than 30km from the county seat (last quartile), indicating that most agglomeration effects occur in the vicinity of the county seat. In addition, we find strong evidence of urban hierarchy effects. While

[10] We do not report the whole set of coefficients associated with the polynomials of the selection probabilities in the table because they have no direct interpretation and because of their high number. Instead, we report of the F-test indicating whether or not the selection correction terms enter the income equation significantly, which is much more informative.

[11] In order to make the interpretation of the results easier, we directly present in the estimation tables the coefficients of the distance variable by quartiles. To do this, we have recalculated the coefficients of distance for each quartile, together with their standard errors, using information on both the additive terms and the interactive terms between distance and the quartiles.

workers in the suburb of the county and prefecture-level cities earn about $7\% - 12\%$ more than workers not located in the suburb of a city, workers in the suburbs of provincial-level cities earn about $54\% - 62\%$ more. To our knowledge, we are the first to demonstrate that rural workers are paid different wages according to their location in rural China. As for physical distance, wages also decrease with frictional distance: wages are higher in villages linked by a road and lower in mountainous areas. Finally, the results are robust whatever the controls introduced for living costs. In the rest of the paper, we use the real wage as dependant variable so that results have to be compared with Column (6) of Table 1.

[Table 3]

Transmission Channels

One additional contribution of this study is to disentangle the role of the potential transmission channels. As stated in the introduction, workers close to cities can be paid higher wages because of both agglomeration externalities and commuting.

To assess the effect of agglomeration externalities, we test the specific effect of specialization[12], diversification and market potential. We use two indicators of specialization at the village level: the share of employees in township and village enterprises and the number of non-agricultural family businesses. Following Combes *et al.* (2008), to capture the effect of the diversity of the economy, we use the log of the inverse of the Herfindahl index. Specifically, in the village questionnaire, the data on the labor force is disaggregated into the five following sectors: agriculture; manufacturing; construction; wholesale, retail and food services; and other industries. Third, we construct two market potential variables. As is widely done, we have constructed a Harris market potential indicator as follows:

[12]The notion of specialization in the context of rural areas in developing countries is different than the usual notion used in the literature on agglomeration economies in urban areas. In agrarian economies, what matters for an economy is to specialize out of agriculture so that new non-agricultural knowledge can emerge, in addition to traditional agricultural knowledge. In this context, looking at the effect of specialization out of agriculture is more relevant than testing the effect of specialization in one particular non-agricultural sector.

$$Harris\ MP_i = \sum_{j=i}^{J} w_{ij} \cdot GDP_j \tag{2}$$

where GDP_j is the Gross Domestic Product of county j[13] and w_{ij} a spatial weighting matrix defined as follows:

$$w_{ij} = 1 \quad if \quad i = j \tag{3}$$

$$w_{ij} = \frac{1}{dist_{ij}} \quad if \quad dist_{ij} \leq 200km \tag{4}$$

$$w_{ij} = 0 \ if \quad dist_{ij} > 200km \tag{5}$$

where $dist_{ij}$ is the number of kilometers between county i and county j. The distance 200km is chosen as the cut-off parameter, *i.e.* beyond 200 km, interactions are considered as negligible[14]. Following Aminiti and Cameron (2007), we include in the measure of the market potential the GDP of the county in which the village is located (Equation 3). As the authors highlight, working with micro-level data alleviates the potential problem of the endogeneity of the own county population that affect studies at the aggregated regional level. Moreover, excluding the market of the own county would lead to an irrelevant measure of the market potential for villages located in the periphery of a city. In addition, we follow Partridge *et al.* (2009) by using the aggregate income in surrounding concentric rings, measured from the population-weighted center of the county, as an additional indicator of market potential. We use the aggregate income in surrounding concentric rings of 0-50km, 50-100km, 100-150km and 150-200km[15]. It is worth noting that the market potential is calculated at the county level. This is due to the fact that even if we know the precise name of every village of the sample, accurate geographical coordinates of the villages are not available so that distance is calculated using the geographical coordinates at the county level.

[13] Data at the county-level comes from the 2003 China Statistical Yearbook for Regional Economy.

[14] The cut-off value chosen is similar to the values used in previous studies (Ke and Feser, 2010; Chen and Partridge, 2011). Moreover, to check the sensibility of our results, we have used other cut-off values and the population of the county instead of the GDP. Results, which are available upon request, are robust to this change.

[15] The intervals chosen are consistent with previous studies (Ke and Feser, 2010).

To investigate the effect of commuting, we decompose local workers into two categories: individuals working in their village and individuals working out of their village (but within their home county) *i.e.*, commuters. To asses the effect of commuting, we estimate the income equation on the sub-sample of individuals working within their village. To do this, we consider commuters as a distinct choice in the multinomial logit model[16]. If commuting plays a significant role, once commuters excluded the coefficients of interest should become less significant.

We begin by testing the effect of agglomeration externalities. As shown in columns (1) and (2) of Table 4, workers living in villages where the economy is diversified are not paid higher wages. Regarding specialization, the share of employees in TVE has a positive effect on wages; however, the effect is not robust when excluding commuters. Turning to market potential, workers are paid higher wages, the higher the market potential of the county. The impact is robust, whatever the measure used. Our result is consistent with Hering and Poncet (2010) who estimate that standard agglomeration effects does not affect wages in Chinese cities, whereas market potential has a very significant impact. Compared with the baseline estimation (Column (6) in Table 1), the coefficients of *Provincial City* and *Low Level City* are less (or not) significant and of lower magnitude when we control for agglomeration externalities. Thus, market potential appears to be one significant transmission channel leading workers close to cities to be paid higher wages[17].

Secondly, we investigate whether workers close to cities are paid higher wages thanks to commuting. As shown in Column (3), when excluding commuters, the coefficients associated with the variables distance are no longer significant. This indicates that workers in villages close to the county seat are more likely to commute to the county seat, where they engage in better paid jobs. Thus, differences in commuting opportunities play a

[16] In this case, the explained variable of the first-step selection equation takes the following values: 0 if local agriculture, 1 if NAWE working within their home village, 2 if commuters, 3 if NASE and 4 if migration.

[17] The coefficients of the distance are almost not affected when we control for agglomeration externalities. This is consistent because market potential is measured at the county level. As every village located in the same county is given the same market potential, the market potential indicator does not capture the effect of the distance to the county seat variable, which varies within villages that belong to the same county.

very significant role in explaining wages disparities across rural workers. On the contrary, the coefficient of *Provincial City* remain strongly significant and its magnitude increases. Thus, even when commuters are excluded, workers close to cities are paid higher wages. This clearly arises from the higher level of wages in villages close to cities. As a consequence, workers close to cities are paid higher wages due to both spatial differences in wages across villages and to higher opportunities to commute. Finally, when controlling both for agglomeration externalities and for commuters, almost all the variables of interest are no longer significant, indicating that we have successfully capture the transmission channels at work.

[Table 4]

Robustness checks

We successively address the issues of endowments, ownership and endogeneity of agglomeration externalities. Results are reported in Table 5[18]. Part I of Table 5 presents the results obtained with additional controls for endowments and Part II the results when controlling for ownership. In parts I and II of Table 5, we have re-estimated the baseline model and the "augmented" model with transmission channels. To limit the number of results, we only provide the estimation results obtained with the Harris market potential indicator[19]. Thus, results obtained in parts I and II of Table 5 have to be compared with results presented in Column (6) of Table 3 and to columns (1), (3) and (4) of Table 4. Finally, we address the issue of the endogeneity of transmission channels in Part III of Table 5. In that case, we have only re-estimated the "augmented" model with transmission channels. Thus, results have to be compared with columns (1) and (4) of Table 4.

Endowments. Differences in regional characteristics are one major source of spatial differences in wages because endowments can affect workers' productivity (Hanson, 2000).

[18]Table 5 only presents the coefficients of interest. The full estimation results are available in the supplementary Appendix S.2.

[19]We only present the results with the Harris indicator for two additional reasons. First, the Harris indicator is much more common. Second, there are two available instruments for market potential so that we can only implement 2SLS with the Harris market potential indicator.

Moreover, endowments are one major source of spatial agglomeration, so that they may be correlated with our indicators of urban proximity, leading to estimation bias. According to Hering and Poncet (2010), endowments are likely to vary across Chinese provinces so that provincial dummies should control for such differences. However, to ensure robustness, we carry out two more tests. First, we substitute provincial dummies with county-level dummies[20]. Second, we follow Fally *et al.* (2010) by excluding from our analysis sectors which depend on natural resources. To do so, we consider non-agricultural wage employment in these sectors as a distinct choice in the multinomial logit model (as we did for commuters). Estimation results are reported in Part I of Table 5. With the exception of the coefficient of *Low level city*, which is no longer significant in Column (1), results are very similar to those obtained in Tables 3 and 4.

[Table 5]

Ownership structure. Wages in the public sector are higher than the average in urban China (Démurger *et al.*, 2012). If wages are also higher in the public sector in rural China, and if the public sector is concentrated close to cities, this would upwardly bias the coefficients associated with urban proximity. Appendix 2 provides descriptive statistics to assess whether or not the lack of control for ownership is likely to bias our results. Appendix 2.1. gives the average hourly wage in our sample for each specific sector: public, semi-public, private and other ownerships. Consistently, workers in the public sector benefit from the highest wages. Appendix 2.2. gives the share of non-agricultural workers in the village involved in each different ownership sector. It clearly appears that the ownership structure does not significantly differ between suburban and non-suburban villages. The semi-public sector is slightly more present in suburban areas. However, given that wages in the semi-public sector are not significantly different than average wages (Appendix 2.1.), this should not lead to estimation bias.

Even if the lack of control for ownership is not likely to lead to estimation bias, controlling for public ownership is an interesting robustness check. Indeed, wages in the

[20] It is worth noting that in this case, we cannot estimate the coefficient of the market potential, given that the market potential is measured at the county level.

public sector are likely to be much less influenced by market access compared with other sectors in China (Hering and Poncet, 2010). To investigate this issue, we estimate the income function by excluding public workers. To exclude the public workers, we consider non-agricultural wage-employment in the public sector as a distinct choice in the multinomial logit model. Estimation results are reported in Part II of Table 5. Excluding public workers does not lead to an increase in the coefficients of interest. This may be due to the fact that public workers only account for a small share of our sample (about 8%).

Endogeneity. In our study, wages are measured at the individual level whereas our indicators of agglomeration externalities are measured at the village (specialization and diversification) or at the county-level (market potential). Thus, a shock to a worker's wage is very unlikely to affect the indicators of agglomeration externalities, measured at a more aggregated level. In other words, using micro data greatly reduces the risk of endogeneity. However, we investigate the robustness of the transmission channels by addressing the potential endogeneity issue. First, following Partridge *et al.* (2009), we have substituted the contemporaneous values of specialization and market potential by their lagged values (measured in 1998)[21]. The use of the lagged value reduces the possible shocks that contemporaneously affect wages and the indicators of agglomeration externalities and thus, it mitigates endogeneity. Second, we instrument the Harris market potential indicator. Following Redding and Venables (2004), we instrument the market potential by the distance to the nearest central locations. In our case, there are two obvious types of central locations: provincial cities (Beijing, Tianjin, Shanghai and Chongqing) and the capital city of each province. Thus, we use two instruments for the Harris market potential: the distance to the nearest provincial city and the distance to the own-provincial capital city. We follow Wooldridge (2002) to implement the instrumental variable approach in the case of the selection model[22]. According to Part III of Table 5, taking into account the potential endogeneity issue does not change the results[23].

[21]We cannot compute a lagged indicator of diversification because the data on labor force in the village is disaggregated by sector only for the year 2002.

[22]First, we estimate the selection equation, in order to generate the selection terms. Second, we estimate the wage equation (augmented by the selection terms) by 2SLS.

[23]Instrument diagnostic tests for equations (3) and (4) are provided in the supplementary Appendix S2. For both equations, instruments are relevant and valid.

Additional issues. There may be some concern about the lack of control for the industry and about the spatial sorting of workers. We provide a discussion of each issue in the supplementary Appendix S.3. We show that both issues are highly unlikely to bias our estimates.

V. Conclusion

We provide a thorough analysis of the effects of urban proximity on rural non-agricultural wages, which are a crucial determinant of the level of earnings and well-being of rural households. We find that remote workers not only suffer from lower opportunities to diversify out of agriculture, but, in addition, when they manage to diversify, they engage in lower paid non-agricultural jobs. By demonstrating that non-agricultural wages vary according to the distance from urban centers, we shed additional light on intra-rural inequality and on the geographic repartition of poverty in China. In order to reduce poverty and inequality in rural China, rural development policies not only must pay attention to the individual determinants of job access and earnings but also to their spatial determinants. Our results also suggest that a minority of villages located close to urban areas benefit from significant and localized agglomeration effects (mainly, market potential). In this context, it may be difficult for rural policies to attract new industries or relocate existing ones to peripheral rural areas. This issue is extremely serious given that non-agricultural employment strongly determines rural welfare. Finally, the issue is exacerbated by the strong institutional restrictions on labor mobility which lead living conditions in an individual's birthplace to still significantly affect his well-being.

References

Aminiti, M. and Cameron, L. (2007) Economic geography and wages, *The Review of Economics and Statistics*, **89**, 15-29.

Bai, C. E., Ma, H. and Pan, W. (2012) Spatial spillover and regional economic growth in China, *China Economic Review*, **23**, 982-990.

Bird, K. and Shepherd, A. (2003) Livelihoods and chronic poverty in semi-arid Zimbabwe, *World Development*, **31**, 591-610.

Bourguignon, F., Fournier, M. and Gurgand, M. (2007) Selection bias corrections based on the multinomial logit model: Monte Carlo comparisons, *Journal of Economic Surveys*, **21**, 174-205.

Chen, A. and Partridge, M. (2011) When are cities engines of growth in China? Spread and backwash effects across the urban hierarchy, *Regional Studies*, 1-19

Combes, P. P., Démurger, S. S. and Li, S. (2012) Urbanization and agglomeration economies in China, Unpublished Paper, GATE - University of Lyon.

Combes, P. P., Duranton, G. and Gobillon, L. (2008) Spatial wage disparities: Sorting matters!, *Journal of Urban Economics*, **63**, 723-742.

Dahl, G.B. (2002) Mobility and the return to education: testing a Roy model with multiple markets, *Econometrica*, **70**, 2367-2420.

de la Rupelle, M., Quheng, D., Li, S., and Vendryes, T. (2010) Land rights insecurity and temporary migration in rural China, IZA Discussion Paper No. 4668.

Deichmann, U., Shilpi, F. and Vakis, R. (2009) Urban proximity, agricultural potential and rural non-farm employment: Evidence from Bangladesh, *World Development*, **37**, 645-660.

Démurger, S., Li, S. and Yang, J. (2012) Earnings differentials between the public and private sectors in China: Exploring changes for urban local residents in the 2000s, *China Economic Review*, **23**, 138-153.

Dubin, J. A. and McFadden, D. L. (1984) An econometric analysis of residential electric appliance holdings and consumption, *Econometrica*, **52**, 345-362.

Ellis F. (1998) Household strategies and rural livelihood diversification, *Journal of Development Studies*, **35**, 1-38.

Fally, T., Paillacar, R. and Terra, C. (2010) Economic geography and wages in Brazil: Evidence from micro-data, *Journal of Development Economics*, **91**, 155-168.

Hanson, G. H. (2000) Firms, workers, and the geographic concentration of economic activity, in *Oxford Handbook of Economic Geography*, (Eds.) G.L. Clark, M. Feldman, M. Gertler, Oxford: Oxford University Press, pp. 477-494.

Hering, L. and Poncet, S. (2010) Market access and individual wages: Evidence from China, *The Review of Economics and Statistics*, **92**, 145-159.

Henderson, J. V., Shalizi, Z. and Venables, A. J. (2001) Geography and development, *Journal of Economic Geography*, **1**, 81-105.

Ke, F. and Feser, E. (2010) Count on the growth pole strategy for regional economic growth? Spread-backwash effects in Greater Central China, *Regional Studies*, **44**, 1131-1147.

Knight, J. and Song, L. (2003) Chinese peasant choices: Migration, rural industry or farming, *Oxford Development Studies*, **31**, 123-148.

Kung, J. K. S. and Lee, Y. F. (2001) So what if there is income inequality? The distributive consequence of nonfarm employment in rural China, *Economic Development and Cultural Change*, **50**, 19-46.

Lanjouw, P. (1999) Rural nonagricultural employment and poverty in Ecuador, *Economic Development and Cultural Change*, **48**, 91-122.

Lee, L.F. (1983) Generalized econometric models with selectivity, *Econometrica*, **51**, 507-512.

Liu, X. and Sicular, T. (2009) Nonagricultural employment determinants and income inequality decomposition, *The Chinese Economy*, **42**, 29-43.

Naughton,B. (2007) *The Chinese Economy: Transitions and Growth*, MIT Press, Cambridge, MA.

Partridge, M.D., Rickman, D.S., Ali, K. and Olfert, M.R. (2009) Agglomeration spillovers and wage and housing cost gradients across the urban hierarchy, *Journal of International Economics*, **78**, 126-140.

Redding, S. and Venables, A. (2004) Economic geography and international inequality, *Journal of International Economics*, **62**, 53-82.

Rosenthal, S. S. and Strange W. (2001) The determinants of agglomeration, *Journal* *of Urban Economics*, **50**, 191-229.

Wooldridge, J. M. (2002) *Econometric analysis of cross section and panel data*, MIT Press, Cambridge, MA.

Xia, Q. and Simmons, C. (2004) Diversify and prosper: Peasant households participating in emerging markets in northeast rural China, *China Economic Review*, **15**, 375-397.

Zhao, Y. (1999) Labor migration and earnings differences: The case of rural China, *Economic Development and Cultural Change*, **47**, 767-782.

Table 1. Classification of workers

	Effective	%
Local agricultural workers	14506	64.32
Local non-agricultural wage-earners	4530	20.09
Local non-agricultural self-employed	863	3.83
Migrant workers	2652	11.76
Total workers	22551	100

Table 2. Distance to urban centers and non-agricultural hourly wages

	Mean	SD	Median	Difference†
All sample	3.07	3.92	2.31	
Distance to county seat				
[0-10] km	3.13	3.97	2.38	
]10-20] km	3.06	4.28	2.24	
]20-30] km	3.05	4.01	2.31	
]30-160] km	3.03	3.47	2.33	
Suburban village♭				
Yes	3.47	4.72	2.50	0.46 **
No	3.01	3.80	2.28	(−2.52)

Notes: †A test of difference between means has been conducted; *t-statistic* reported in parenthesis.
♭Suburban village are villages located in the suburb of a large or middle city (*jiaoqu*).
*, **, *** indicate significance at the 10%, 5% and 1% levels respectively.

Table 3. Baseline estimation

	Multinomial logit model			Income equation		
	NAWE	NASE	Migration			
	(1)	(2)	(3)	(4)	(5)	(6)
Individual characteristics						
Age	0.007***	0.002***	−0.0001	0.044***	0.045***	0.046***
	(0.000)	(0.001)	(0.576)	(0.000)	(0.000)	(0.000)
Age^2	−0.011***	−0.003***	−0.005***	−0.054***	−0.055***	−0.056***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education	0.011***	0.001***	−0.002***	0.043***	0.045***	0.043***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Experience	0.082***	0.015***	0.034***	0.122***	0.127***	0.116***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
$Experience^2$	−0.220***	−0.051***	−0.122***	−0.310***	−0.326***	−0.297***
	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.001)
Party member	0.053***	−0.010***	−0.022***	0.196***	0.200***	0.183***
	(0.000)	(0.006)	(0.001)	(0.000)	(0.000)	(0.000)
Male	0.154***	0.014***	0.032***	0.426***	0.440***	0.423***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Minority	−0.034***	−0.003	−0.018***	−0.097	−0.101*	−0.104*
	(0.009)	(0.319)	(0.000)	(0.108)	(0.088)	(0.063)
Village characteristics						
Distance q1	−0.004**	−0.001**	0.002**	−0.022***	−0.022***	−0.022***
	(0.018)	(0.023)	(0.025)	(0.000)	(0.000)	(0.000)
Distance q2	−0.001*	−0.0004**	0.001***	−0.006**	−0.006**	−0.006**
	(0.066)	(0.034)	(0.008)	(0.011)	(0.016)	(0.011)
Distance q3	−0.002***	−0.0003**	0.001***	−0.006***	−0.006***	−0.005***
	(0.000)	(0.026)	(0.004)	(0.003)	(0.005)	(0.005)
Distance q4	−0.002***	−0.0004***	0.0005***	−0.003**	−0.003**	−0.003**
	(0.000)	(0.000)	(0.001)	(0.024)	(0.018)	(0.018)
Low level city	0.058***	0.008***	−0.019***	0.119**	0.103**	0.077*
	(0.000)	(0.002)	(0.000)	(0.010)	(0.015)	(0.072)
Provincial city	0.108***	−0.002***	0.521***	0.543**	0.623***	0.567**
	(0.000)	(0.004)	(0.000)	(0.012)	(0.009)	(0.015)
Road	0.050***	0.005	−0.0001	0.192***	0.190**	0.170**
	(0.003)	(0.198)	(0.593)	(0.008)	(0.011)	(0.024)
Topography	−0.011	0.002	0.009***	−0.045*	−0.047**	−0.045**
	(0.149)	(0.361)	(0.001)	(0.059)	(0.034)	(0.037)
Township	−0.025**	−0.016***	0.009	−0.119**	−0.107**	−0.104*
	(0.018)	(0.000)	(0.164)	(0.028)	(0.044)	(0.079)
Living costs					0.479***	
					(0.005)	
Land per capita	−0.007***	−0.002**	−0.003***			
	(0.000)	(0.020)	(0.000)			
Unmarried	−0.002	−0.012***	0.061***			
	(0.480)	(0.005)	(0.000)			
Observations		22551		4530	4530	4530
R^2				0.2061	0.2071	0.1778
Selection correction terms				2.02***	1.99***	1.98***
Test for identifying restrictions:						
Wald test (first-step equation)	15.62***	13.99***	114.75***			
F-test (second-step equation)				1.82	1.62	1.71
Distance q1 = Distance q2				13.01***	12.88***	13.51***
Distance q2 = Distance q3				0.05	0.03	0.18
Distance q3 = Distance q4				4.81**	4.55**	3.47*
Provincial city = Low level city				4.31**	5.68**	4.81**

Notes: *, **, *** indicate significance at the 10%, 5% and 1% levels respectively. *p-values* in parenthesis. The models also include a constant term, regional and provincial dummy variables, but coefficients for these variables are not reported. The models have been estimated using a bootstrap procedure with 500 replications.

Table 4. Transmission channels

	Agglo. ext.		No commuters	All channels	
	(1)	(2)	(3)	(4)	(5)
Individual characteristics					
Age	0.046 * **	0.043 * **	0.063 * **	0.052 * **	0.052 * **
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age^2	−0.056 * **	−0.053 * **	−0.068 * **	−0.060 * **	−0.059 * **
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education	0.041 * **	0.042 * **	0.048 * **	0.039 * **	0.042 * **
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Experience	0.109 * **	0.105 * **	0.152 * **	0.096 * **	0.113 * **
	(0.000)	(0.000)	(0.003)	(0.008)	(0.001)
$Experience^2$	−0.278 * **	−0.267 * **	−0.406 * **	−0.261 * **	−0.298 * **
	(0.000)	(0.000)	(0.004)	(0.010)	(0.002)
Party member	0.174 * **	0.170 * **	0.372*	0.200	0.279
	(0.000)	(0.000)	(0.088)	(0.252)	(0.100)
Male	0.423 * **	0.408 * **	0.432 * **	0.384 * **	0.403 * **
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Minority	−0.058	−0.103*	−0.182 * *	−0.102	−0.191 * *
	(0.276)	(0.053)	(0.031)	(0.219)	(0.022)
Village characteristics					
Distance q1	−0.019 * **	−0.019 * **	−0.027 * *	−0.026 * *	−0.025 * *
	(0.000)	(0.000)	(0.013)	(0.019)	(0.028)
Distance q2	−0.006 * **	−0.006 * **	−0.004	−0.007	−0.007
	(0.003)	(0.006)	(0.456)	(0.207)	(0.252)
Distance q3	−0.005 * **	−0.005 * **	−0.006	−0.008	−0.007
	(0.001)	(0.004)	(0.230)	(0.112)	(0.182)
Distance q4	−0.002 * **	−0.002 * *	−0.003	−0.003*	−0.003
	(0.006)	(0.016)	(0.160)	(0.098)	(0.167)
Low level city	0.038	0.078 * *	0.101	0.035	0.083
	(0.242)	(0.020)	(0.111)	(0.560)	(0.139)
Provincial city	0.227	0.108	1.734 * **	0.106	0.752*
	(0.194)	(0.597)	(0.000)	(0.496)	(0.077)
Road	0.162 * **	0.143 * *	0.417 * *	0.294 * *	0.333 * *
	(0.008)	(0.018)	(0.016)	(0.046)	(0.023)
Topography	−0.016	−0.032*	0.030	0.015	0.017
	(0.387)	(0.080)	(0.616)	(0.790)	(0.750)
Township	−0.101 * *	−0.090*	−0.346 * **	−0.301 * **	−0.272 * **
	(0.033)	(0.056)	(0.000)	(0.000)	(0.001)
Transmission channels					
Share of workers in TVE	0.040 * **	0.041 * **		0.065	0.109*
	(0.009)	(0.006)		(0.289)	(0.091)
NA family businesses	4.52E-05	4.35E-06		1.57E-05	4.02E-05
	(0.777)	(0.978)		(0.953)	(0.879)
Diversification	0.032	0.035		−0.015	0.016
	(0.335)	(0.294)		(0.767)	(0.771)
Harris market potential	0.096 * **			0.075 * *	
	(0.000)			(0.015)	
Agg inc 0 - 50 km		1.03E-04***			1.43E-04***
		(0.001)			(0.005)
Agg inc 50 - 100 km		8.79E-05**			1.17E-04**
		(0.011)			(0.020)
Agg inc 100 - 150 km		1.14E-04***			1.26E-04***
		(0.000)			(0.006)
Agg inc 150- 200 km		1.09E-04***			1.19E-04**
		(0.000)			(0.020)
Observations	4530	4530	1997	1997	1997

Notes: *, **, *** indicate significance at the 10%, 5% and 1% levels respectively. *p-values* in parenthesis. The models also include a constant term, regional and provincial dummy variables, but coefficients for these variables are not reported. The models have been estimated using a bootstrap procedure with 500 replications.

Table 5. Robustness checks: endowment, ownership and endogeneity

	(I) Endowments							
	(i) County dummies				(ii) No NR workers			
	No channels (1)	Agglo. ext. (2)	No commuters (3)	All channels (4)	No channels (1)	Agglo. ext. (2)	No commuters (3)	All channels (4)
Distance q1	−0.020***	−0.019***	−0.022**	−0.019*	−0.020***	−0.017***	−0.023**	−0.024**
	(0.000)	(0.000)	(0.028)	(0.078)	(0.000)	(0.001)	(0.030)	(0.035)
Distance q2	−0.006***	−0.006***	−0.003	−0.002	−0.006**	−0.005***	−0.004	−0.007
	(0.007)	(0.009)	(0.525)	(0.670)	(0.013)	(0.007)	(0.483)	(0.221)
Distance q3	−0.007***	−0.007***	−0.006	−0.006	−0.005**	−0.005***	−0.006	−0.007
	(0.000)	(0.000)	(0.130)	(0.134)	(0.013)	(0.003)	(0.255)	(0.137)
Distance q4	−0.003***	−0.003***	−0.002	−0.002	−0.002**	−0.002**	−0.003	−0.003
	(0.000)	(0.000)	(0.262)	(0.297)	(0.033)	(0.018)	(0.167)	(0.103)
Low level city	0.128	0.120	0.085	0.099	0.060	0.037	0.087	0.030
	(0.151)	(0.182)	(0.540)	(0.482)	(0.136)	(0.269)	(0.194)	(0.641)
Provincial city	1.714***	1.645***	2.187***	1.980***	0.514**	0.251	1.850***	1.437***
	(0.000)	(0.000)	(0.004)	(0.005)	(0.011)	(0.163)	(0.000)	(0.001)
Share of workers in TVE		0.050***		0.082		0.033**		0.056
		(0.000)		(0.102)		(0.035)		(0.365)
NA family businesses		1.64E-04		5.21E-05		1.01E-04		5.36E-05
		(0.325)		(0.834)		(0.542)		(0.855)
Diversification		0.017		0.008		0.037		0.005
		(0.637)		(0.893)		(0.269)		(0.923)
Harris market potential		-		-		0.087***		0.079**
						(0.000)		(0.015)
Observations	4530	4530	1997	1997	4414	4414	1932	1932

	(II) No public workers				**(III) Endogeneity**			
					Lagged values		2SLS	
	No channels (1)	Agglo. ext. (2)	No commuters (3)	All channels (4)	All NAWE (1)	No commuters (2)	All NAWE (3)	No commuters (4)
Distance q1	−0.022***	−0.019***	−0.018	−0.022*	−0.019***	−0.027**	−0.018***	−0.027***
	(0.000)	(0.000)	(0.144)	(0.097)	(0.000)	(0.016)	(0.000)	(0.007)
Distance q2	−0.006**	−0.005***	$-1.84E-04$	−0.005	−0.006***	−0.007	−0.008***	−0.010*
	(0.013)	(0.009)	(0.977)	(0.444)	(0.002)	(0.189)	(0.001)	(0.051)
Distance q3	−0.005***	−0.005***	−0.002	−0.005	−0.006***	−0.008*	−0.008***	−0.013***
	(0.009)	(0.002)	(0.680)	(0.339)	(0.000)	(0.092)	(0.000)	(0.007)
Distance q4	−0.002**	−0.002**	-8.56E-04	−0.001	−0.002***	−0.004*	−0.003***	−0.004**
	(0.034)	(0.023)	(0.750)	(0.549)	(0.003)	(0.065)	(0.002)	(0.019)
Low level city	0.066*	0.034	0.068	−0.005	0.053	0.058	0.009	0.019
	(0.100)	(0.318)	(0.322)	(0.929)	(0.112)	(0.351)	(0.780)	(0.766)
Provincial city	0.304**	0.247**	1.451***	0.961**	0.259	0.147	−0.301	1.017**
	(0.016)	(0.024)	(0.001)	(0.018)	(0.144)	(0.352)	(0.304)	(0.024)
Share of workers in TVE		0.048***		0.067	0.054**	0.111	0.055**	0.064
		(0.006)		(0.339)	(0.011)	(0.184)	(0.016)	(0.343)
NA family businesses		-2.88E-05		-3.77E-04	1.52E-04	1.55E-04	2.91E-04	1.94E-04
		(0.883)		(0.274)	(0.416)	(0.609)	(0.166)	(0.524)
Diversification		0.028		−0.011	0.029	−0.014	0.060*	0.017
		(0.419)		(0.841)	(0.371)	(0.784)	(0.096)	(0.739)
Harris market potential		0.091***		0.061*	0.093***	0.076**	0.264***	0.238*
		(0.000)		(0.063)	(0.000)	(0.015)	(0.006)	(0.099)
Observations	4181	4181	1862	1862	4530	1997	4530	1997

Notes: NAWE refers to non-agricultural wage-earners. *, **, *** indicate significance at the 10%, 5% and 1% levels respectively. *p-values* in parenthesis. These models use the same control variables as the models shown in Tables 3 and 4. The models have been estimated using a bootstrap procedure with 500 replications.

Appendix 1. Definition of explanatory variables

Variables	**Definition**	**Unit**
Age	Age of the worker	Year
Education	Number of years of schooling (not including years spent on repeating a grade)	Year
Experience	Number of years since when the worker starts a non-agricultural activity as his primary activity	Year
Party member	Dummy equal to 1 if the worker is member of the Communist Party, 0 otherwise	
Male	Dummy equal to 1 if the worker is a man, 0 otherwise	
Minority	Dummy equal to 1 if the worker is an ethnic minority, 0 otherwise	
Unmarried	Dummy equal to 1 if the worker is not married, 0 otherwise	
Land per capita	Total amount of land possessed per capita in the household	Mu
Distance	Distance from the nearest county seat	Kilometers
Quartile 1	Dummy equal to 1 if the worker's village is located within 10 km from the county seat (the first quartile of the distance is equal to 10 km)	
Quartile 2	Dummy equal to 1 if the worker's village is located within 10 and 20 km from the county seat (the median of the distance is equal to 20 km)	
Quartile 3	Dummy equal to 1 if the worker's village is located within 20 and 30 km from the county seat (the third quartile of the distance is equal to 30 km)	
Low level city	Dummy equal to 1 if the village is in the suburb of a prefecture city or if it is located in the administrative area of a county-level city, 0 otherwise	
Provincial city	Dummy equal to 1 if the village is in the suburb of a provincial-level city, 0 otherwise	
Road	Dummy equal to 1 if a road reaches the village, 0 otherwise	
Topography	Variable equal to 1 if the village is located in a plain, 2 if in a hilly area and 3 if in a mountainous area	
Township	Dummy equal to 1 if the township the village is in is a province level poverty township	
Living costs	Average market price of six non-staple foods (meat, eggs, edible oil, sugar, vegetables, fruit and melons) (in logarithm form)	Yuan
Share of workers in TVE	Share of employees in township and village enterprises in the village	%
Family businesses	Number of non-agricultural family businesses in the village	
Diversification	Inverse of the Herfindahl index, calculated using labor force data at the village level, disaggregated into five sectors: agriculture; manufacturing; construction; wholesale, retail and food services; other industries	
Harris market potential	Sum of the GDP of the counties, weighted by the distance in km between the county in which the worker's village is located and other counties (see Equation 2)	Yuan
Agg inc within 0-50 km ring	Aggregate income between 0 and 50 km radii from county centroid	Yuan
Agg inc within 50-100 km ring	Aggregate income between 50 and 100 km radii from county centroid	Yuan
Agg inc within 100-150 km ring	Aggregate income between 100 and 150 km radii from county centroid	Yuan
Agg inc within 150-200 km ring	Aggregate income between 150 and 200 km radii from county centroid	Yuan
Distance to own capital city	Number of km between the county seat (in which the worker's village is located) and the capital city of the province	km
Distance to nearest provincial city	Number of km between the county seat (in which the worker's village is located) and the nearest provincial city	km

Appendix 2. Descriptive statistics on ownership

2.1. Hourly wages disaggregated by ownership

	Nb. of workers	Average wage	SD	Difference[†]
All NAWE	4530	3.07	3.92	
According to ownership:				
Public	349	3.81	3.18	3.66 * **
Semi-public	526	3.11	2.65	0.24
Private	1694	2.76	3.79	−4.02 * **
Other	1961	3.19	4.39	1.79*

Notes: NAWE means non-agricultural wage earners.
[†] Tests of difference between means have been conducted to compare the wage in each sector to the average wage in the rest of the economy; *t-statistics* are reported in this column.
*, **, *** indicate significance at the 10%, 5% and 1% level respectively.

2.2. Share of non-agricultural wage-earners by ownership

	All villages[♭] (N=788)		Suburban (N=62)		Non-suburban (N=726)		Difference[†]
	Mean	SD	Mean	SD	Mean	SD	
All NAWE	20.55	18.03	34.32	23.14	19.40	17.06	6.71 * **
According to ownership:							
Public	8.77	18.95	9.78	17.94	8.68	19.05	0.44
Semi-public	9.08	20.11	13.85	22.29	8.67	19.87	1.95*
Private	34.21	33.08	32.60	27.20	34.35	33.54	−0.40
Other	47.94	35.45	43.76	30.58	48.29	35.83	−0.97
Total	100		100		100		

Notes: NAWE means non-agricultural wage earners.
[♭] Villages with no non-agricultural workers are excluded from the table.
[†] Tests of difference between means have been conducted to compared ownership and industry structures between suburban and non-suburban villages; *t-statistics* are reported in this column.
*, **, *** indicate significance at the 10%, 5% and 1% level respectively.