

Chapter 2

How Important Is Economic Geography for Rural Non-agricultural Employment? Lessons from Brazil

Co-authored with Steven Helfand

2.1. INTRODUCTION

Rural non-agricultural employment in developing countries has received increasing attention since the early 1990s. The share of rural household income that stems from non-agricultural sources ranges from 35 percent in Asia to 40 percent in Latin America and 45 percent in Sub-Saharan Africa, underscoring the fact that the rural economy consists of much more than just agriculture (Reardon et al., 2001). The rural non-agricultural (RNA) sector has been considered as able to absorb an underemployed rural labor force, and thereby slow rural-to-urban migration, to increase the income of the rural poor, and to contribute to national economic growth (Lanjouw and Lanjouw, 2001; Kay, 2005). Many rural development strategies have also emphasized RNA development as a pathway out of poverty for rural landless households and land-constrained family farmers, (de Janvry and Sadoulet, 1993; Echeverría, 2000; Quijandría et al., 2001; World Bank 2003 and 2007).

There is still an incomplete understanding of what determines rural non-agricultural employment (RNAE) opportunities, and to what extent such employment is able to reduce poverty and improve living standards for rural households (Haggblade et al., 2007). The empirical literature on the determinants of RNAE and its impact on poverty reduction has so far mainly focused on supply-side considerations, such as household and worker endowments, without controlling for other important factors. The general hypothesis posed in this chapter is that employment and earnings

opportunities in the RNA sector are determined jointly by individual and household characteristics (supply-side effects), labor market characteristics (demand-side effects), and the transaction costs of participating in markets. Household asset endowments on their own are unlikely to generate upward income mobility if there is insufficient demand for labor, or if market participation is very costly due to physical distance to markets or underdeveloped infrastructure. In this chapter we therefore seek to jointly assess the importance of supply, demand, and transaction costs on an individual's probability of engaging in RNAE and on earned income in the RNA sector. We devote particular attention to the role of participation costs and demand-side effects. Even though there is a consensus that location matters for the viability of the RNA sector, the empirical support so far relies on indirect locational indicators, which give us limited insight into the role that remoteness from markets and urban areas actually plays (Dirven, 2004).¹

Due to its size and regional diversity, Brazil provides an excellent case study to assess the importance of economic geography for rural non-agricultural employment. To reach a deeper understanding of demand-side effects and the role of transaction costs, our study utilizes a more fine-grained set of variables than previous studies to describe the local economic geography. In particular, by utilizing data from the Brazilian Demographic Census, we are able to test for the role of municipal-level economic factors such as local market size, infrastructure, and distance to population centers.²

As expected, the empirical results show that personal and household characteristics matter for employment outcomes and for income earnings potential. Demand-side factors and proxies for transaction costs, however, also have a strong influence on the probability of being engaged in RNAE. Market size and the degree of urbanization are associated with greater RNAE

¹ In her survey of the literature, Dirven (2004, p. 60) states: "Returning to the more economic view of 'distance' (i.e., that of transaction costs generated by physical distance), evidence as to RNFE [rural non-farm employment] is still scant, but there is no doubt that distance and the transaction costs that ensue play a role both directly and indirectly..."

² The Brazilian literature on RNAE has been based almost exclusively on the national household surveys (PNAD). Ney and Hoffmann (2007), who also utilize the 2000 Demographic Census, is the one exception that we are aware of. PNAD is only representative at the state level, thus providing little insight into how employment and income outcomes are conditioned by location.

opportunities. Similarly, distance to population centers has a large effect on outcomes. These factors do not render individual characteristics insignificant, but in some cases substantially alter their magnitude. Geographical variables have a weaker and less consistent relationship to earnings. Like nearly all of the literature on this topic, it is important to emphasize that this is not a causal analysis and that the results should be interpreted as conditional correlations. Given these limitations, our conclusions on the importance of the local economic geography stand up to a number of robustness checks that seek to address endogeneity and measurement concerns.

The next section of the chapter reviews how locational factors have been analyzed in the RNAE literature. Section 2.3 provides an overview of rural employment and the RNA sector in the case of Brazil. Section 2.4 contains the first part of the empirical analysis, which is concerned with the relation between local characteristics and RNAE. Section 2.5 extends the empirical analysis by assessing the dependence of rural non-agricultural income on geographical factors. Section 2.6 concludes the chapter.

2.2. PREVIOUS STUDIES ON ECONOMIC GEOGRAPHY AND RURAL EMPLOYMENT

It is widely recognized that geographical location and economic conditions specific to the local economy matter, in one way or another, for the employment outcome and earnings prospects of rural households. Dirven (2004) provides a valuable discussion of the literature. Previous studies have utilized a range of indicators to capture the effect of local economic conditions. In addition to regional dummy variables, locational variables that have been used include: distance to regional capital city and local population density (Abdulai and Delgado, 1999); rural sub-categories such as urban extension or rural town (Ferreira and Lanjouw, 2001); distance to nearest health center (Corral and Reardon, 2001); number of population centers within one hour's commuting distance (de Janvry and Sadoulet, 2001); distance to nearest market and local market size (Escobal, 2001); local road conditions and distance to nearest school (Lanjouw, 2001); neighborhood average household income, local urbanization, and electrification (Isgut, 2004); and altitude, distance to nearest pharmacy, and number of hostel beds as a proxy for tourism (Laszlo, 2005). Van de Walle and Cratty (2004) provide an illustration of the extent to

which geographical effects might matter. In their analysis of the probability of non-agricultural self-employment in Vietnam, commune dummies account for two thirds of the explained variance of the model.

A few observations are pertinent to the previous literature regarding the use of these different kinds of variables to account for the economic geography. First, when feasible, geographical dummy variables may be used as control variables to capture unobserved local factors. A weakness of regional dummies or fixed effects is that they do not lend themselves to interpretation. They may, however, be used as a benchmark to explore whether a set of interpretable geographical variables is sufficient to remove bias due to omitted local variables of the other coefficients in the model. Second, variables that relate to location in space can provide an attractive alternative to geographical dummies. Longitude, latitude, and altitude can help to control for the influence of unmeasured geographical variables but, like dummies, in many cases they do not have a natural economic interpretation. Variables that can serve as proxies for distance to markets or market potential are likely to be preferable (Baltenweck and Staal, 2007). Third, some variables are more informative than others, and a family of variables might be preferable to a single one. For example, distance to the nearest school, health clinic, pharmacy, or state capital all carry some information about remoteness, but the information is fuzzy. Certainly, it should matter if the nearest urban location has five thousand or five hundred thousand people, just as it should matter if a household has two cities with 10 000 people at less than 50 kilometers away rather than just one. Last, while it is clear that transaction costs should play an important role in influencing the probability of RNAE, proxies for these costs should be interpreted with caution. Population density or share of households with telephones, for example, are associated with better infrastructure in general, and lower costs of moving people and information. The magnitude of a coefficient of any single proxy, however, might vary considerably depending on if it is used to represent the entire group of transaction cost variables, or is included as only one of many of these variables.

More often than not, the decisions on which geographical variables to use are driven by data availability. Due to the abundance of data contained in the Brazilian Demographic Census, we seek to shed light on the extent to which alternative choices that are common in the literature are adequate for capturing the effects of the local economic geography.

2.3. THE RURAL NON-AGRICULTURAL SECTOR: THE CASE OF BRAZIL

2.3.1. *The data*

The description of the RNA sector that follows is based on the Brazilian Demographic Census long form of the year 2000. The long form was applied to a sample of more than 20 million observations (approximately 12 percent of the population), constructed to be representative at the municipal level. There were 5 507 municipalities at the time of the survey, with an average population of approximately 30 000 people. Our empirical analysis uses the rural adult labor force as the base sample, which includes about 1.7 million observations.³ Adults were defined as everyone aged 15 years or more. Anyone reporting an occupation was considered as a participant in the labor force, including unpaid workers. By RNAE we mean that a person resides in a rural domicile, yet has a principal occupation in a non-agricultural activity. Thus, this person could work at home producing handicrafts, in a rural home as a maid, in a rural area with tourism, or in an urban area in a non-agricultural occupation.⁴

The income data in the Demographic Census suffer from the same limitations as those from the Brazilian National Household Survey (PNAD). As described in Ferreira and Lanjouw (2001), the single question about earnings does not a) distinguish clearly between gross and net income for the self-employed, b) take proper account of seasonal earnings which are common in agriculture, or c) adjust for own consumption of agricultural production by farmers. These limitations with how income is measured in the Census and PNAD are most problematic for small farmers and the self-employed. For this reason, our econometric analysis of earnings is restricted to people employed in RNAE, and contains a robustness check limited to the sub-sample of wage earners.

³ There is considerable debate in Brazil about the appropriate definition of “rural” areas. We use the official definition of rural areas based on municipal government decisions. As Table 6 in Ney and Hoffmann (2007) shows, alternative definitions of “rural” have no impact on the qualitative results about the relative importance of variables in earnings equations, and have only a minor impact on the magnitude of these effects.

⁴ Many authors, such as Reardon et al. (2001), use the term rural non-farm employment (RNFE) in the same way that we use RNAE. We prefer RNAE because it emphasizes the distinction between location of residence and sector of work. RNAE is distinct from *off-farm* employment, which includes agricultural wage labor.

With only 19 percent of its population residing in rural areas, Brazil is a highly urbanized country. While the rural population share is close to the average for Latin America, it is much lower than in other developing regions such as South Asia (72 percent) and Sub-Saharan Africa (64 percent). With 22 people per square-kilometer, Brazil also has a low population density, with rural households often being widely dispersed and far away from major population centers. Some of this is captured directly by the Demographic Census. It classifies the rural census tracts into five sub-categories: 1) rural agglomerations that are urban extensions, 2) isolated rural agglomerations or towns that have some service provision, 3) isolated rural agglomerations linked to a single landowner, 4) other isolated agglomerations, and 5) rural areas exclusive of agglomerations. The vast majority of the rural population, 86 percent, falls into the fifth category, and the Census provides no information that assists us to identify the degree of remoteness of these households. Around 11 percent of the rural population live in rural towns or agglomerations, and only three percent are found in urban extensions.

The information on household income available in the Census data suggests that rural remoteness tends to go hand in hand with poverty. Rural poverty was above 70 percent in the less urbanized North and Northeast, and below 45 percent in the other three macro regions (South, Southeast, and Center-West). Poverty rates within each region also increase the further away from urban areas one gets, rising from 42 percent in rural areas classified as urban extensions to 62 percent in rural areas exclusive of agglomerations.⁵

2.3.2. The rural non-agricultural sector

Table 2.1 shows that, of the rural labor force, 70 percent have their principal employment in agriculture (cultivation, animal rearing, and forestry). The remaining 30 percent are employed in RNA activities. Empirical evidence shows that the share working in RNA activities has increased over time (Graziano da Silva and del Grossi, 2001). There are regional variations in the composition of the rural labor force. The Northeast is not only the poorest

⁵ The poverty headcount ratio reported here uses a poverty line set at R\$75 per month, which corresponds to half the minimum wage of August 2000.

How Important Is Economic Geography for Rural Non-agricultural Employment?

TABLE 2.1. *Share of rural labor force by sector of principal occupation*

	Agriculture			Non-agriculture
	Cultivation	Animal rearing	Forestry	
<i>Region</i>				
Brazil	0.56	0.12	0.02	0.30
North	0.52	0.12	0.04	0.32
Northeast	0.66	0.07	0.03	0.25
Southeast	0.43	0.16	0.01	0.39
South	0.56	0.15	0.02	0.27
Center-West	0.27	0.41	0.02	0.30
<i>Rural sub-category</i>				
Urban extension	0.08	0.02	0.00	0.90
Rural towns	0.38	0.06	0.02	0.54
Rural exclusive	0.60	0.13	0.02	0.25
<i>Employment status</i>				
Wage labor	0.31	0.15	0.02	0.52
Self-employed	0.60	0.11	0.03	0.26
Unpaid	0.83	0.10	0.02	0.05
<i>Gender</i>				
Men	0.59	0.14	0.02	0.25
Women	0.48	0.07	0.03	0.42

Source: Demographic Census 2000, authors' calculations.

region, but is also the region with the lowest share in the non-agricultural sector (25 percent). RNAE is greatest in the relatively urbanized Southeast region (39 percent). Table 2.1 also shows that rural areas that are extensions of urban areas are dominated by non-agricultural work. Only 10 percent of the labor force in these areas is involved in agriculture. Non-agricultural activities also employ more people than agriculture in rural towns.

As a residual concept, the rural non-agricultural sector contains a wide range of activities, including everything from low-return street-vending to well-paid jobs in the formal sector. Table 2.2 shows that the five largest RNA sectors are manufacturing, commerce, domestic services, education and construction, which together employ almost 70 percent of the non-agricultural labor force. Manufacturing employ a considerably larger share in the North and South than in the other regions. Domestic services play a larger role in Southeast and Center-West. Among the self-employed engaged in non-agricultural activities, manufacturing and commerce are the two major sectors. Among wage laborers, the largest sector of non-agricultural employment is domestic services. The most noticeable difference between male and female non-agricultural work is

TABLE 2.2. *Share of rural non-agricultural employment by sub-sector*

	<i>Region</i>						<i>Employment</i>		<i>Gender</i>	
	Brazil	North	North-east	South-east	South	Center-West	Wage Labor	Self-employed	Men	Women
Manufacturing	0.20	0.25	0.18	0.18	0.29	0.16	0.18	0.22	0.23	0.17
Commerce	0.14	0.13	0.14	0.15	0.15	0.15	0.09	0.27	0.17	0.10
Domestic Services	0.14	0.08	0.12	0.21	0.13	0.23	0.21	0.00	0.05	0.28
Education	0.11	0.10	0.14	0.06	0.07	0.11	0.16	0.01	0.03	0.22
Construction	0.10	0.05	0.11	0.12	0.09	0.07	0.10	0.12	0.16	0.00
Public administration	0.06	0.05	0.07	0.04	0.05	0.06	0.09	0.00	0.05	0.07
Other sectors	0.25	0.34	0.24	0.24	0.22	0.22	0.17	0.38	0.31	0.16
<i>Total</i>	<i>1.00</i>	<i>1.00</i>	<i>1.00</i>	<i>1.00</i>	<i>1.00</i>	<i>1.00</i>	<i>1.00</i>	<i>1.00</i>	<i>1.00</i>	<i>1.00</i>

Source: Demographic Census 2000, authors' calculations.

that women dominate the jobs classified as domestic services and education, while men are engaged to a higher extent in activities such as construction and transportation.

Traditionally, the RNA sector has been considered largely dependent on backward and forward linkages to agriculture (Mellor, 1976; Tomich et al., 1995). A significant share of Brazilian agriculture, however, is characterized by large-scale, commercial, highly mechanized export-oriented production. Thus, it is unclear how strong such linkages are in Brazil relative to countries with smaller farms, lower levels of technology, and weaker linkages to the world market. In this spirit, Graziano da Silva and del Grossi (2001) argue that the composition of the RNA sector in Brazil often bears little relation to regional agricultural development, and that its dynamism depends more on the degree of urbanization and the size of cities in a given region. Ferreira and Lanjouw (2001) also argue that proximity to urban areas is an important determinant of employment in the RNA sector, a view supported by Figures 2.1A and 2.1B (pages 26–27) of the Brazilian Southeast and Northeast. The maps depict the share of the rural labor force whose principal occupation is in RNAE in each municipality. Non-agricultural activities are more prevalent in the proximity of capital cities and highly urbanized areas, a pattern most pronounced in the densely populated areas surrounding São Paulo, Rio de Janeiro, and Belo Horizonte in Figure 2.1A. In these areas, RNAE is above 50 percent, whereas in some of the remote hinterlands the share falls below 15 percent.

2.3.3. Rural non-agricultural income

On average, people earn higher incomes in the rural non-agricultural sectors than in agriculture. This is true for wage laborers and the self-employed, as well as for men and women. Table 2.3 shows average monthly earnings in the six non-agricultural sectors that employed the majority of the RNA labor force. The average earnings in agriculture in the year 2000 were R\$280 when considering earned monetary income from principal employment and excluding those with zero reported income. Domestic services is the only major RNA sector in which average earnings are lower than in agriculture. The self-employed earn more than wage laborers, and in all sectors men earn more than women.

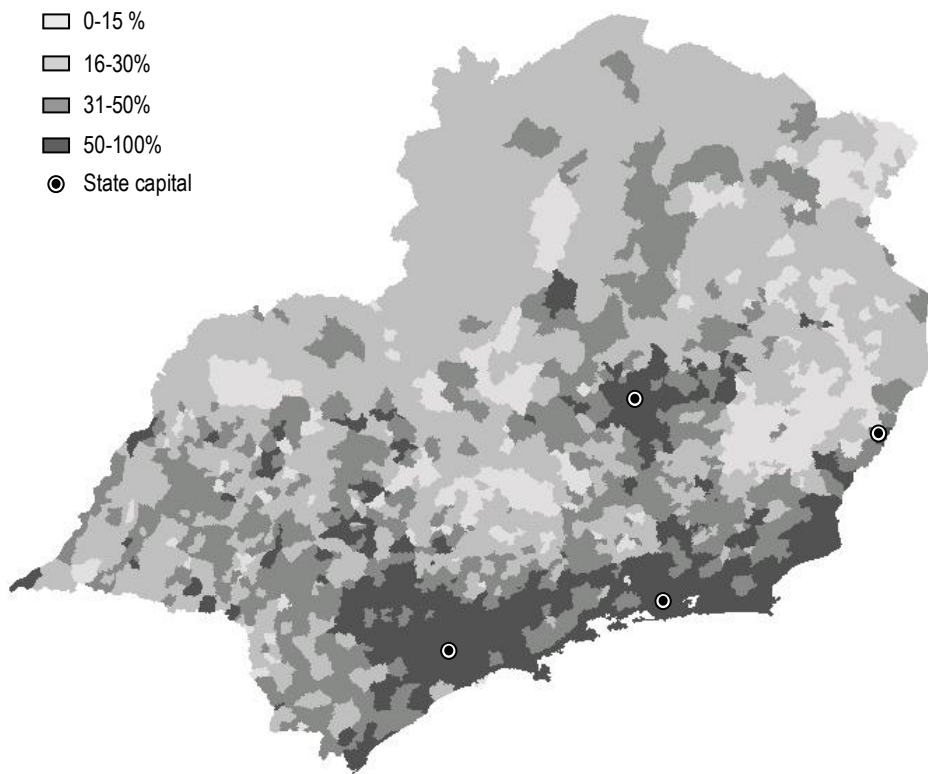
Even though average earnings in most of the RNA sectors are higher than in agriculture, there are also many low-paid non-agricultural jobs. We divide individuals with RNAE into two groups depending on earnings relative to agriculture. If an individual is engaged in RNAE and has earnings below the average municipal earnings of wage laborers in agriculture, we consider the individual as being engaged in low-productivity RNAE. Those who earn above this average are classified as being engaged in high-productivity RNAE. With this categorization, only 53 percent of the non-agricultural labor force is engaged in high-productivity RNAE, although average earnings in RNAE are 25 percent higher than in agriculture. In the educational sector more than two-thirds of the labor force have high-productivity jobs. In domestic services, in contrast, only 20 percent of employment is high productivity.

Table 2.3. *Rural non-agricultural income by sector (R\$ per month, 2000)*

	Brazil	Wage Labor	Self-employed	Men	Women	Share high productivity
Manufacturing	337	314	385	390	209	0.51
Commerce	449	310	578	492	329	0.57
Domestic Services	160	160	n/a	223	140	0.21
Education	295	292	411	394	274	0.68
Construction	334	299	402	335	321	0.65
Public administration	387	387	n/a	507	256	0.64
All RNA sectors	345	294	479	416	236	0.53
<i>Agriculture</i>	280	198	346	296	170	n/a

Note: The exchange rate US\$/R\$, August 2000, was 0.56. *Source:* Demographic Census 2000, authors' calculations.

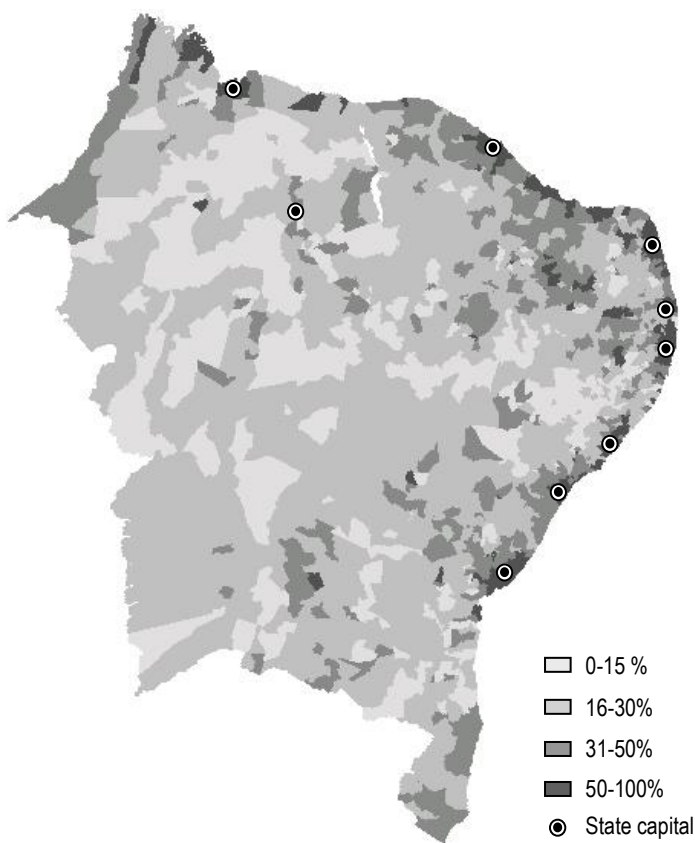
FIGURE 2.1A. Rural non-agricultural employment in the Brazilian Southeast



Source: Demographic Census 2000, authors' calculations.

Non-agricultural activities are often viewed as a means of income diversification among rural households (Ellis, 2000). For households in rural Brazil, however, engaging in RNAE for this purpose does not appear to be a deliberate strategy of the majority of households. We define households as specialized in agriculture if they derive 90 percent or more of their earned income from agriculture, specialized in non-agriculture if they derive 90 percent or more from RNAE, and pluriactive otherwise. Only 14 percent of rural

FIGURE 2.1B. *Rural non-agricultural employment in the Brazilian Northeast*



Source: Demographic Census 2000, authors' calculations.

households are considered pluriactive by this definition. Noticeable in terms of specialization is that richer households are engaged in RNAE to a larger extent than poorer households.

Differences in average earnings suggest that the rural non-agricultural sector could potentially provide a pathway out of rural poverty. To assess this potential, we analyze the importance of supply, demand, and transaction costs in the following two sections, first by assessing what influences the probability

that people in the rural labor force engage in non-agricultural activities, and second by examining what affects their earnings.

2.4. EMPIRICAL ANALYSIS OF RURAL NON-AGRICULTURAL EMPLOYMENT

This section reports the results of a probability analysis of engagement in rural non-agricultural employment. First, we estimate a binomial probit model in which the dependent variable indicates whether the individual was engaged in RNAE as opposed to agriculture. Second, motivated by the heterogeneity of earnings in RNAE, we use a multinomial probit model to estimate jointly the probabilities of engaging in high- and low-productivity RNAE in comparison with agriculture.

2.4.1 Estimation method

The binomial model is specified based on the assumption that a set of exogenous variables determines an endogenous, but unobserved (latent), variable V . If V exceeds a certain threshold value, V^* , the individual is engaged in RNAE; otherwise, he or she is engaged in agriculture. The latent variable can be thought of as the rural worker's expected earnings if participating in the rural non-agricultural sector. The threshold could be the shadow wage for agricultural work on the own farm or the wage rate on the agricultural labor market. The probability that individual i is engaged in RNAE, P_i , is modeled as the probability that V_i exceeds V_i^* . If v_i denotes the difference $V_i - V_i^*$, then the probability is given by:

$$P_i = \text{prob}(RNAE_i = 1 | X_{ijk}, H_{jk}, M_k) = \text{prob}(v_i \geq 0) \quad (1)$$

where X , H , and M denote vectors of individual, household, and municipal variables, respectively. Subscript i refers to individuals, j to households, and k to municipalities. The potential net benefit of RNAE, v_i , is assumed to be a linear function of X , H , and M :

$$v_i = X_{ijk} \beta_1 + H_{jk} \beta_2 + M_k \beta_3 + \varepsilon_{ijk} \quad (2)$$

where the β s are vectors of coefficients to be estimated, and ε is a residual assumed to be normally distributed with zero mean and variance σ^2 . Let $F(\cdot)$ be the standard normal cumulative distribution function of ε . The individual's probability of engaging in RNAE is estimated as:

$$P_i = \text{prob}(X_{ijk} \beta_1 + H_{jk} \beta_2 + M_k \beta_3 \geq -\varepsilon_{ijk}) = F(X_{ijk} \beta_1 + H_{jk} \beta_2 + M_k \beta_3) \quad (3)$$

In the second approach, which involves the estimation of a multinomial probit model, we distinguish three forms of employment (EMP): agricultural work, low-productivity RNAE, and high-productivity RNAE. The threshold that is used to separate the two RNAE types is the average agricultural earnings of wage laborers in each municipality. The model is specified as:

$$P_i^e = \text{prob}(EMP_i = e | X_{ijk}, H_{jk}, M_k) = F(X_{ijk} \beta_1^e + H_{jk} \beta_2^e + M_k \beta_3^e) \quad (4)$$

where P^e denotes the probability that individual i has employment type e (e being any of the three defined employment forms).

2.4.2. Variables used in the empirical analysis

Table 2.4 provides descriptive statistics and definitions of the variables. The binary variable that indicates that the individual is engaged in RNAE is based on reported principal occupation. The individual characteristics included in X are age, gender, race/color, education, and migrant status. Age and years of schooling serve as proxies for human capital. Even though human capital matters for agricultural labor productivity, the non-agricultural sector is likely to contain those jobs with the highest returns to education, and would hence attract the relatively well-educated workers in the rural labor force. Human capital can also have the allocative effect of allowing households to make optimal labor allocation decision (Yang and An, 2002; Laszlo, 2005). Education is controlled for by four dichotomous variables that are based on the number of completed years of schooling. Zero education is the benchmark category and contains about 24 percent of the rural labor force. Gender is included to control for systematic differences between male and female workers in terms of job preferences and work hours, but also to control for demand-side effects such as gender discrimination in payment schemes. Dummy variables for race/color

are included for similar reasons. A dummy variable for migrants is included, indicating whether the individual has moved to the municipality or always lived there. Migration could be an indicator of unobserved ability and risk-taking, and hence willingness to engage in the employment with highest returns for the individual. In this respect migration reflects an endogenous choice, which could lead to bias in the estimated coefficients. While we do not model the endogeneity of migration or education, such as in a two-stage least squares framework, we do explore the magnitude of the potential bias of other coefficients with several robustness tests. The remaining individual variables are used in the income analysis and are discussed in Section 2.5.

Household characteristics (H) include the number of adult household members, average education in the household (excluding individual i), and an index of household wealth. The number of adults is included to control for opportunities for employment diversification. Household labor endowment may improve the opportunities to devote some household labor to non-agricultural activities. Average education among other household members is a proxy for the household stock of human capital. Given that there are some spillover effects within the household, the higher the average education, the more likely it is that an individual undertakes employment with skill requirements (Laszlo, 2005). A proxy for household wealth is included to summarize a vector of characteristics of the domicile.⁶ Greater household wealth could increase the probability of RNAE for a number of reasons. Wealthier households are better able to finance the search and participation costs associated with RNAE. Wealth can also serve as a proxy for social capital which can facilitate access to non-agricultural jobs. Two variables are also included to indicate whether the household lived in a rural town or urban extension as opposed to a rural exclusive area. Among the household variables, the wealth and urban extension/rural town variables are the ones that are most likely to suffer from endogeneity. It is possible that causality runs in both directions between household wealth and RNAE. High-return RNAE, for example, would allow households to accumulate wealth over time. Location of

⁶ The proxy is constructed as the first principal component of the following 14 variables: ownership of domicile, ownership of land, piped water in domicile, and number of rooms, bathrooms, refrigerators, washing machines, microwaves, computers, televisions, VCRs, radios, air conditioners, and automobiles. The first principal component explains 31 percent of the variation in the original 14 variables.

residence, like migration, is also an individual (or household) decision. As with migration and education, we construct robustness tests to explore the degree to which this potential endogeneity might be biasing the estimates on the other coefficients.

Municipal-level characteristics (M) are included to assess the importance of local demand and transaction costs for the employment outcome. To estimate the local market size, we use two distance-weighted measures of aggregate income, constructed in the same spirit as Harris's (1954) market potential measure. Both measures include the total income of people in the municipality plus total income in surrounding municipalities weighted by distance, but they differ in the weighting scheme. The first variable, *Local income 1*, is defined as the sum of municipal income over all municipalities, weighted by the inverse of the distance D_{kl} from a typical rural household in the municipality of origin k to the seat of municipality l :

$$Local\ income\ 1_k = \sum_l Income_l (1/D_{kl}) \quad (5)$$

$Income_l$ refers to the sum of all income received by households in each municipality l as reported in the Census. The distance D_{kl} is the sum of two components: the estimated distance d_k from a typical rural household in municipality k to its own municipal seat and the distance d_{kl} from the seat of municipality k to the seat of municipality l . The weight for $Income_l$ in equation (5) is designed so that the size of the market – both within and outside of the municipality – is a decreasing function of distance. The second measure of market size, *Local income 2*, uses a linearly declining weight that only takes into account municipalities (l^*) within a 100-kilometer distance of a typical rural household:

$$Local\ income\ 2_k = \sum_{l \in l^*} Income_l (1 - D_{kl}/100) \quad (5')$$

In this case, the weight equals 1 for $D_{kl} = 0$ and declines to 0 for $D_{kl} \geq 100$ km. As can be seen in Table 2.4, by the large difference in means between the two variables, *Local income 1* discounts much more heavily for distance than *Local income 2*. For example, income in a municipality 50 kilometers away only gets a two-percent weight with the former, but a 50-percent weight with the latter. The weighting scheme in *Local income 2* seems more realistic in terms of potential RNAE. Analogous population variables (*Local population 1* and *Local population 2*) are constructed to check for robustness.

Chapter 2

Table 2.4. Summary statistics of variables used in the empirical analysis

Variable	Mean	Std. Dev.	Description
Dependent variables			
RNAE	0.30	0.45	Individual has RNAE as principal employment (d)
RNAE low	0.15	0.35	Individual has low-productivity RNAE (d)
RNAE high	0.15	0.35	Individual has high-productivity RNAE (d)
Non-agr income	345	1,173	Individual's earned non-agricultural income (if >0)
Individual characteristics			
Age	36.27	14.72	Individual's years of age
Male	0.71	0.45	Gender, 1 if male (d)
Black	0.07	0.26	Race – black (d)
Asian	0.002	0.05	Race – Asian (d)
Mixed	0.45	0.50	Race – mixed (d)
Indigenous	0.01	0.08	Belongs to indigenous group (d)
Education	3.57	3.24	Individual's years of education
Education 1-4	0.49	0.50	1 to 4 years of education (d)
Education 5-8	0.18	0.38	5 to 8 years of education (d)
Education 9-11	0.08	0.27	9 to 11 years of education (d)
Education 12	0.01	0.10	12 or more years of education (d)
Migrant	0.37	0.48	Individual has migrated from other municipality (d)
Formal sector	0.16	0.36	Paid employee in the formal sector (d)
Informal sector	0.25	0.43	Paid employee in the informal sector (d)
Self-employed	0.32	0.46	Self-employed (d)
Employer 1	0.005	0.07	Employer with 1–2 employees (d)
Employer 2	0.002	0.05	Employer with 3–5 employees (d)
Employer 3	0.002	0.04	Employer with 6 or more employees (d)
Unpaid	0.27	0.45	Unpaid worker (d)
Hours	42.35	15.13	Hours worked per week
Household characteristics			
HH adults	3.26	1.64	Number of adults in the household
HH education	3.64	2.73	Average years of education among other adults in the household
HH wealth	-0.65	0.74	Household wealth index
Urban extension	0.03	0.15	Residence in urban extension (d)
Rural town	0.09	0.27	Residence in rural town (d)
Rural exclusive	0.87	0.31	Residence in rural area, excl. of towns/extensions (d)
North	0.10	0.29	Residence in North (d)
Northeast	0.42	0.49	Residence in Northeast (d)
South	0.20	0.41	Residence in South (d)
Southeast	0.23	0.43	Residence in Southeast (d)
Center-West	0.05	0.22	Residence in Center-West (d)

How Important Is Economic Geography for Rural Non-agricultural Employment?

Table 2.4. (Continued)

Variable	Mean	Std. Dev.	Description
Municipal characteristics			
Urbanization	0.60	0.22	Share urban households in municipality
Telephones	0.06	0.09	Share of rural households with fixed telephone line
Electrification	0.75	0.26	Share of rural households with electric lighting
Local income 1	73.7	45.4	Distance-weighted local income, million R\$ (see eqn. 5)
Local income 2	178	531	Distance-weighted local income, million R\$ (see eqn. 5)
Local population 1	236,416	97,358	Distance-weighted local population (analogous to eqn. 5)
Local population 2	561,716	1,107,277	Distance-weighted local population (analogous to eqn. 5)
Distance 50	76	74	Distance to municipality with 50-100,000 people, kilometers
Distance 100	124	130	Distance to mun., 100-250,000 people, km
Distance 250	207	174	Distance to mun., 250-500,000 people, km
Distance 500	260	195	Distance to municipality with >500,000 people, km

Note: Weights were used to estimate population mean. Variables indicated by (d) are dichotomous variables, taking value 1 if true, 0 otherwise. The sample size is 1,724,822. For the municipal variables, the unweighted municipal-level mean is reported.

We use a collection of variables as proxies for transaction costs. As the municipality of residence may or may not be the relevant marketplace, we include measures of distance to population centers to estimate the effect of being situated away from markets of different sizes. Using D_{kl} , distances are estimated to the nearest municipality with 50–100, 100–250, 250–500, and more than 500 thousand people. The corresponding variables are labeled *Distance 50*, *Distance 100*, *Distance 250*, and *Distance 500*, respectively. Conceptually the size of the local market and the distance to markets of different sizes might both be considered as alternative proxies for demand. In contrast to using the local income variables, which emphasize the total size of the local market, we use the distance measures primarily to assess the importance of transaction costs associated with access to markets. The distance variables also enable us to capture non-linearity in the relationship between RNAE and distance to markets of different sizes. Three variables that characterize the level of infrastructure development in the municipality are also used: the shares of rural households with access to a telephone line and to electric lighting are included to capture the level of rural infrastructure in the

municipality, and the share of urban households is used to reflect the hypothesis that urbanization is correlated with infrastructure development. A greater degree of infrastructure development should lower the costs of participation in markets.

2.4.3. Estimation results of the binomial probit model

The results from the binomial probit model are provided in Tables 2.5 and 2.6. First, Table 2.5 shows coefficients from specifications in which variables are added stepwise. We compare the supply-side models to several models that include geographical variables and to a model with municipal fixed effects. Second, Table 2.6 presents the results of six robustness checks on the coefficients of the geographical and education variables. Finally, we briefly discuss several alternative specifications, which use geographical variables common in the literature.

The reported marginal effects in the tables give the estimated change in the probability of employment in the RNA sector, as opposed to agriculture, given a small change in the explanatory variable or a change from 0 to 1 in the dichotomous variables. Due to the sample size, nearly all coefficients are statistically significant at least at the one percent level. For this reason, all the tables in this chapter identify those coefficients that are *not* significant at the one-percent level.

Model (i) includes only individual variables. When household characteristics are controlled for, as in model (ii), the coefficient estimates of some individual characteristics change significantly. The marginal effects of all educational levels decrease substantially, suggesting that these variables, in part, capture the effect of the excluded household variables. Omitted-variable bias is also evident when model (ii) is compared to models (iii) – (v), which include the geographical variables. The coefficients of higher education (*Education12*), migrants, and household wealth, for example, all change significantly. Thus, failure to adequately control for the local economic geography may generate significant bias.

The results in Table 2.5 also provide insight into the extent to which local conditions matter for employment outcomes. Comparing the pseudo- R^2 from each model shows that, as a group, the locational variables explain an important share of the variance in the probability of RNAE. When household

variables are added to model (i), the explained variance increases by only 16 percent. When the household *and* locational variables are added to model (i), the explained variance increases by over 75 percent. The goodness of fit criterion also helps to choose among the geographical models. Model (iv) with the family of distance variables provides a better fit than model (iii) with the single local economic demand variable. Model (v) shows that, when both the distance and local economic demand variables are included together, the coefficient of the latter variable becomes zero. The local economic demand variable provides an interesting alternative interpretation to the distance variables, but, as discussed below, there appear to be non-linearities in the relationship between RNAE creation and municipalities of different sizes, and the family of distance variables do a better job at capturing this. We therefore focus on the coefficients in model (iv), and use this specification as a reference model for checking the robustness of our results.

Model (iv) shows that human capital is positively associated with the probability of engagement in RNAE: age has a positive and decreasing effect on the probability of non-agricultural employment, and the probability increases non-linearly with the level of educational attainment. Having one to four years of education, compared to none, is associated with an additional 5.7 percentage points in the probability of RNAE. Having five to eight or nine to 11 years of education, in contrast, is associated with increases of 18 and 36 percentage points, respectively.

Consistent with the descriptive data presented in Table 2.1, women have a substantially higher probability of engaging in RNAE. People who have moved from one municipality to another – migrants – are more likely to engage in non-agricultural activities, but the effect is quite small (2.2 percentage points). Several observations on the household variables are warranted. The positive coefficients of household wealth and education provide support for the wealth and intra-household ‘knowledge spillover’ hypotheses. Given the individual’s educational attainment, the education of other household members as well as the wealth of the household are positively correlated with RNAE outcomes. The number of household adults, in contrast, has a weak negative partial correlation with RNAE, speaking against the employment diversification hypothesis.

The reference model (iv) also shows that all but one of the proxies for demand-side effects and transaction costs are statistically significant with the

TABLE 2.5. *Empirical results: binomial probit model of RNAE*

	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Supply-side factors						
Age	0.014	0.010	0.010	0.010	0.010	0.008
Age squared	-0.000	-0.000	-0.000	-0.000	-0.000	(0.000)
Male	-0.139	-0.145	-0.150	-0.151	-0.151	-0.173
Education 1-4	0.091	0.056	0.057	0.057	0.057	0.059
Education 5-8	0.273	0.190	0.175	0.177	0.177	0.176
Education 9-11	0.486	0.361	0.359	0.363	0.363	0.383
Education 12	0.602	0.429	0.469	0.467	0.467	0.509
Migrant	0.058	0.047	0.012	0.022	0.022	0.025
HH adults		-0.010	-0.005	-0.005	-0.005	-0.005
HH education		0.013	0.010	0.010	0.010	0.009
HH wealth		0.102	0.055	0.058	0.058	0.059
Demand-side factors and transaction costs						
Local income 2 (log)			0.051		(-0.001)	
Distance 500 (log)				-0.073	-0.074	
Distance 250 (log)				-0.040	-0.040	
Distance 100 (log)				-0.011	-0.011	
Distance 50 (log)				-0.004	-0.004	
Urban extension			0.519	0.500	0.500	0.383
Rural town			0.238	0.236	0.235	0.225
Urbanization			0.118	0.099	0.098	
Telephones			0.294	0.246	0.247	
Electrification			-0.118	-0.097	-0.096	
McFadden pseudo R ²	0.112	0.130	0.190	0.198	0.198	n/a
Sample size	1,724,822	1,724,822	1,724,822	1,724,822	1,724,822	344,964

Note: The dependent variable is the binary variable *RNAE*. All coefficients are statistically significant at the one-percent level except for coefficients within parentheses, which are not significant at the 10-percent level. (log) indicates that the natural logarithm of the variable was used in the model specification. All specifications include racial control variables. Specifications (i) – (v) include macro regional control variables. Specification (vi) include municipal fixed effects. Standard errors are available from the authors.

expected sign. Living in a rural area that is an urban extension, as opposed to living in the rural exclusive category, is associated with a 50-percentage-point increase in the probability of RNAE, while residence in a rural town is associated with more than 20 additional percentage points. The degree of urbanization of the municipality also matters; the higher the share of urban households, the higher the probability of non-agricultural employment for rural residents.

The results in the reference model further suggest that distance to population centers matters for RNAE prospects. The greater the distance to large municipalities of all four size categories, the lower the probability that an individual will engage in RNAE. At the mean of 260 km, an additional standard deviation of distance (195 km) away from municipalities with more than 500,000 residents is associated with a 5.5 percentage point decline in the probability of RNAE. One measure of remoteness would be to move an additional standard deviation of distance away from each of the four classes of large municipalities. The combined effect would be a reduction of approximately 10.4 percentage points in the probability of RNAE. Municipalities of different sizes, however, have quite different impacts on the probability of RNAE. Moving 100 km away from the largest class of municipalities is associated with a change in the probability of RNAE that is five times larger than the change for municipalities in the 50–100 thousand class, and three times larger than those in the 100–250 thousand class. We suspect that it is because of these non-linearities that the distance model fits the data better than the local income model. This also suggests that proxies that only measure the distance to an urban area or state capital, without accounting for its size, miss an important part of this relationship.

The one case where we find mixed evidence of transaction costs relates to the proxies for rural infrastructure. The shares of rural households with telephones and electricity point in different directions regarding their relationship to RNAE. Telephones are associated with a higher probability of RNAE, whereas electrification is associated with a lower probability. With only six percent of rural households reporting the existence of a land line in their domicile, it is likely that this variable is highly correlated with proximity to urban areas. Thus, in addition to aiding in the flow of information, this variable serves as a proxy that complements the other locational variables. Regarding the negative coefficient of electricity, we note that the simple correlation between electricity and RNAE is 0.26, and that electricity is highly correlated

with many of the other geographical variables in the model. We have explored the possibility that municipal outliers might be driving this unexpected result and experimented individually, and jointly, with trimming the tails of the municipal variables, but in no case has this led to substantially different results. The results of a model that simultaneously removes the tails from the municipal variables electrification, telephones, and urbanization is presented in the first column of Table 2.6.⁷ The combination of exclusions reduces the number of municipalities by 1 523, and the sample by 21 percent. We conclude that the negative coefficient of electricity is not an artifact of a group of atypical municipalities. Additional research is required to better understand this result.

2.4.4. Robustness

We have performed a host of other robustness checks on the reference model (iv) in Table 5, to detect potential bias in the results. The discussion of the results focuses on the robustness of the local economic geography coefficients, and then on the education coefficients. First, the estimated effects of the individual and household characteristics could be influenced by unobserved local factors that we are unable to control for with the vector of local level variables. Instead of using a set of municipal-level variables to explore this issue, the model is estimated with municipal fixed effects and the urban extension and rural town dummies that vary by census tract. The results in column (vi) of Table 2.5 show that the coefficients of all non-municipal level variables are quite similar to the reference model. The largest differences relate to the urban extension variable, yet none of these changes are large enough to alter the interpretation of the results. We conclude that the geographical controls in the probability model are adequate.

A second set of concerns relates to the possible endogeneity of several of the regressors. The most powerful potential criticism of our results could be that unobserved individual characteristics, which have higher returns in RNAE, induce people with those characteristics to move to locations where they have a higher probability of finding RNAE. If true, the coefficients of urban extensions,

⁷ As outliers, we consider municipalities with any of the following conditions met: *Urbanization* ≥ 0.95 , *Telephones* = 0, *Telephones* ≥ 0.4 , or *Electrification* ≥ 0.99 . These exclusions reduce the number of municipalities by 234, 847, 70, and 561, respectively.

rural towns, and the family of distance variables, for example, would be biased upwards (in magnitude), because people have chosen to reside closer to where the RNA jobs exist. In order to test for this possibility, we re-estimate the reference model first without migrants, then without individuals who live in urban extensions and rural towns, and finally without either. With migrants removed from the model, column (ii) of Table 2.6 shows that the sample size drops by one third. The most notable change is that the *Distance 50* coefficient becomes statistically insignificant. The coefficients of most of the other geographical variables fall, but not by enough to change any of our conclusions regarding the importance of the local economic geography. For example, the “remoteness” exercise – which involves moving one standard deviation away from each of the four largest classes of municipalities – now leads to a decline of 8.4 (rather than 10.4) percentage points in the probability of RNAE.

By excluding towns and urban extensions, not only are we addressing the endogeneity of location of residence, but also the heterogeneity that clearly exists in relation to the exclusively rural areas. Column (iii) shows that the geographical coefficients change even less than when migrants are excluded. In the model without urban extensions, rural towns, or migrants (column iv), the sample drops by more than 35 percent, and the share with principal occupation in RNAE falls to 25 percent. Thus, while this specification eliminates the problem of endogeneity of where people choose to live, it begins to generate a sample that is no longer representative of rural Brazil. Nevertheless, column (iv) shows that the results are quite similar to when only migrants are excluded. We conclude that there is some evidence in favor of the hypothesis of endogenous sorting of the rural population, but that this does not alter the fundamental conclusions about the importance of the local economic context: distance to markets matters, as does the local infrastructure.

Columns (v) and (vi) of Table 2.6 report the results of two additional robustness tests. The question addressed here is not whether the coefficients of education and wealth might be biased due to their own endogeneity, but how much this might matter for our conclusions about the importance of the local economic geography. In both cases, we restrict the sample to be much more homogenous along these two dimensions, and explore whether any important conclusions are altered. When the sample is restricted to the middle 25 percent of individuals according to wealth, the standard deviation of the wealth variable falls by 76 percent. Other than the coefficient of the telephone variable becoming much larger, the results are largely unchanged. Similarly, when the

TABLE 2.6. *Robustness checks of the results of the binomial probit model*

	(i) No outlier municipalities	(ii) No migrants	(iii) No urban ext./rural towns	(iv) No migrants, urban ext./rural towns	(v) Homog. HH wealth	(vi) No education ≥ 5 years
Distance 500 (log)	-0.078	-0.053	-0.067	-0.048	-0.078	-0.060
Distance 250 (log)	-0.037	-0.036	-0.037	-0.034	-0.026	-0.034
Distance 100 (log)	-0.001*	-0.013	-0.009	-0.012	-0.006	-0.006
Distance 50 (log)	-0.003	(0.000)	-0.004	(-0.001)	-0.006	-0.003
Urban extension	0.484	0.492			0.442	0.492
Rural town	0.234	0.212			0.205	0.202
Urbanization	0.079	0.100	0.103	0.100	0.116	0.085
Telephones	0.471	0.131	0.195	0.109	0.637	0.241
Electrification	-0.102	-0.078	-0.089	-0.072	-0.146	-0.068
Education 1-4	0.058	0.056	0.047	0.050	0.060	0.039
Education 5-8	0.177	0.170	0.159	0.158	0.179	
Education 9-11	0.370	0.372	0.347	0.360	0.400	
Education 12	0.476	0.537	0.471	0.541	0.512	
Observed RNAE	0.297	0.280	0.263	0.252	0.314	0.229
Predicted RNAE	0.280	0.258	0.239	0.228	0.295	0.208
McFadden pseudo R ²	0.200	0.164	0.140	0.126	0.169	0.122
Sample size	1,369,849	1,097,407	1,552,654	1,005,911	431,205	1,266,379

Note: Unless otherwise stated, all specifications include the following variables: age, age squared, male, black, Asian, mixed, indigenous, educational variables, migrant, HH adults, HH education, HH wealth, and macro region. Specification (i) excludes individuals residing in "outlier" municipalities; (ii) excludes migrants; (iii) excludes individuals residing in urban extensions and rural towns; (iv) excludes migrants and individuals residing in urban extensions and rural towns; (v) includes only the middle 25 percent in the household wealth distribution; and (vi) excludes individuals with five or more years of education. All coefficients are statistically significant at the one-percent level except in the following cases: * denotes significance at 10-percent level and coefficients within parentheses are not significant at the 10-percent level. Standard errors are available from the authors.

sample is restricted to include only those individuals with 4 years or less of education (thus removing the 27 percent of the sample for whom education leads to dramatically different probabilities of RNAE), the coefficients of the economic geography variables remain quite similar to the reference model of Table 2.5. No qualitative results change, and most quantitative results remain stable.

Table 2.6 also shows how the education coefficients are affected by the robustness tests. When migration and wealth are addressed, the coefficients of the upper one or two educational dummies increase somewhat. In the test for sensitivity to municipal outliers, the education coefficients change very little. Thus, the tests conducted here point to considerable stability of the quantitative results. We conclude that education is one of the most important factors influencing the probability of RNAE, and that the coefficients in Table 2.6 provide a plausible range for these effects.

We briefly comment on alternative geographical specifications that are common in the literature. The positive coefficient of *Local income 2* in specification (iii) of Table 2.5 provides a lens for examining the importance of local demand. The coefficient of this variable indicates that a one standard deviation increase in the size of the local market is associated with a 15-percentage-point increase in the probability of RNAE. This is similar to what we find when we use the analogous *Local population 2* variable (described above). A one standard deviation increase in this variable is associated with a 12.2-percentage-point increase in the probability of RNAE. Both models are similar to the distance model in terms of removing bias from the supply-side variables.

When the population of the municipality is used instead of the population or income of the surrounding region, a few important differences emerge. The supply-side coefficients remain largely unbiased, but the signs and magnitudes of some of the other municipal variables change, the elasticity on the local population is smaller, and so is the pseudo R^2 . For these reasons, we conclude that specifications that include the surrounding income or population are preferred to those that include solely the own municipal income or population. When latitude and longitude are used in place of the distance variables, the model suffers from limitations similar to the model that uses the municipal population. Finally, a model that includes distance to the own state capital would be comparable to a model that only includes distance to municipalities with more than 500 thousand people. The estimates of the supply-side variables were almost identical, and the estimated coefficients of

the other municipal variables were similar, but the explanatory power of the full model was greater.

We conclude that the inclusion of geography in almost any form contributes to reducing bias of the supply-side coefficients. Our results also suggest that more comprehensive and precise descriptions of the local economic environment are preferred. The distance variables are preferred to the local income or local population variables, which in turn were preferred to the municipal population. Similarly, based on the pseudo R^2 , models that include a) the distance variables, b) extensions and towns, and c) municipal variables, are always preferred to models that only include one or two of these three groups.

2.4.5. Estimation results of the multinomial probit model

The results from the multinomial probit model are provided in Table 2.7. Due to computational intensity, the model is estimated with a 20-percent random sample from the data used in the estimation of binomial probit model. The results are highly consistent with the binomial model, but there are several new findings.

Even though women have a much higher probability of engaging in RNAE than men, the decomposition of RNAE into low- and high-productivity jobs shows that this “advantage” is mostly in terms of low-productivity employment, where they earn less than the mean municipal earnings of agricultural wage laborers. According to specification (ii), women are 18 percentage points more likely to be employed in low-productivity RNAE than men, but are at a slight disadvantage in the selection process into high-productivity RNAE. The results also suggest that human capital does not affect low- and high-productivity RNAE equally. Even having only one to four years of education increases the probability of high-productivity RNAE by around five percentage points, but matters little for the probability of low-productivity RNAE. Similarly, at higher levels of schooling, most if not all of the reduction in the probability of being employed in agriculture is translated into an increase in the probability of having high-, not low-, productivity RNAE.

The second specification in Table 2.7 shows that proximity to markets and factors that reduce transaction costs are generally associated with a higher probability of both low- and high-productivity RNAE. A one standard deviation

TABLE 2.7. Empirical results: multinomial probit model of rural employment outcome

	(i) Supply-side specification			(ii) Distance specification		
	Agr. employment	Low-prod. RNAE	High-prod. RNAE	Agr. employment	Low-prod. RNAE	High-prod. RNAE
Supply-side factors						
Age	-0.011	-0.005	0.016	-0.011	-0.005	0.016
Age squared	0.000	0.000	-0.000	0.000	0.000	-0.000
Male	0.147	-0.173	0.026	0.153	-0.179	0.026
Education 1-4	-0.061	0.009	0.051	-0.062	0.009	0.053
Education 5-8	-0.200	0.046	0.154	-0.187	0.035	0.152
Education 9-11	-0.382	0.050	0.332	-0.385	0.044	0.341
Education 12	-0.430	-0.053	0.483	-0.466	-0.046	0.512
Migrant	-0.047	0.029	0.018	-0.021	0.011	0.011
HH adults	0.009	-0.002	-0.007	0.004	0.001	-0.005
HH education	-0.011	0.002	0.009	-0.008	(0.000)	0.008
HH wealth	-0.094	0.015	0.079	-0.049	-0.015	0.065
Demand-side factors and transaction costs						
Distance 500 (log)				0.072	-0.036	-0.035
Distance 250 (log)				0.039	-0.010	-0.029
Distance 100 (log)				0.010	-0.009	(-0.001)
Distance 50 (log)				0.003**	(0.000)	-0.003
Urban extension				-0.515	0.265	0.250
Rural town				-0.234	0.133	0.101
Urbanization				-0.102	0.076	0.026
Telephones				-0.229	0.220	(0.008)
Electrification				0.088	-0.008*	-0.080
Sample size	345,038			345,038		

Note: The dependent variable is employment outcome (*EMP*), which is agricultural work, *RNAE low*, or *RNAE high*. The marginal effects refer to the change in probability of being in the respective employment category, given a small change in a continuous variable or a discrete change in a dichotomous variable. All coefficients are statistically significant at the one-percent level except in the following cases: ** denotes significance at five-percent level, * denotes significance at 10-percent level, and coefficients within parentheses are not significant at the 10-percent level. Standard errors are available from the authors.

move away from municipalities in all four “large” classes leads to a combined reduction of 4.4 and 5.5 percentage points in the probability of low- and high-productivity RNAE, respectively. The effect of local aggregate income – in a specification not shown here – also has a slightly larger impact on high-productivity than low-productivity RNAE. Thus, we conclude that locational factors play an important role in the selection out of agriculture and into RNAE, but they do not unambiguously favor low- or high-productivity RNAE. Gender, education, and household wealth, in contrast, help to sort across types of RNAE.

2.5. NON-AGRICULTURAL INCOME

The purpose of this section is to assess the degree to which local economic factors affect earnings opportunities in the RNA sector. Our findings suggest that geography also matters for non-agricultural income opportunities, but that the effects are not as strong as with employment outcomes.

2.5.1. Estimation method

Of the 1.7 million individuals who represented the rural labor force in the previous analysis, about 470 000 reported earned income from non-agricultural employment. The results from the probit model suggest that individual characteristics, along with demand factors and transaction costs, determine the selection process into RNAE, so that people engaged in non-agricultural activities differ systematically from people engaged in agriculture. Failure to control for this selection mechanism, and the possibility that unobserved factors influence both selection and income, would provide inconsistent coefficient estimates in an OLS regression. To adjust for the effects of censoring the sample, we apply the Heckman (1979) sample selection model.⁸

Our approach assumes that selection into remunerated RNAE is determined by a model analogous to (1) in Section 2.4. Accounting for the

⁸ A limitation of the Heckman procedure is that it relies on normality assumptions of the error terms in the selection and income equations. For alternative models, see Deaton (1997).

results of the selection process, we assume that income can be modeled as a linear function of individual, household, and locational characteristics:

$$y_i = X_{ijk} \beta_1 + H_{jk} \beta_2 + M_k \beta_3 + \gamma \lambda_{ijk} + \eta_{ijk} \quad (6)$$

where y is the logarithm of non-agricultural income of the individual. Income refers to monthly wage earnings for employees and returns to the own business for employers and the self-employed, during the month of July 2000. X , H , and M are vectors of explanatory individual, household, and municipal characteristics, λ is the inverse Mills ratio, η is the error term, assumed to be normally distributed, and β and γ are coefficients to be estimated. Most of the explanatory variables are the same as in the probit model. We add number of hours worked to the individual characteristics and variables to control for employment status: formal-sector employee, self-employed, and three groups of employers based on the number of people they hire. We interact the self-employment dummy with the household wealth index in order to control for productive assets among the self-employed.

When estimating the Heckman model, it is important to pay attention to the issue of identification of the inverse Mills ratio, λ . Identification requires having at least one variable that influences the probability of selection, but does not enter the income equation (6). We use specification (iii) of the probit model in Table 2.5 as the first-step selection equation. We believe that household size should have no influence on individual earnings. Thus, it enters the selection equation, but is excluded from the income equation. We also use the local aggregate income variable – *Local income 2* – for identification. Finally, the household wealth variable contributes, in part, to identification because it enters the selection equation for all individuals, but only enters the income equation for the self-employed.

A test of $\gamma = 0$ is a test of whether the correction for sample selection is necessary. If different from zero, this implies that there are common factors that influence both selection and income, and that the errors from these two equations are correlated. If γ is positive then there is positive selection into RNAE, that is, unobserved characteristics, which correlate positively with income, correlate positively with the probability of having RNAE. If γ is negative, the reverse is true. The inclusion of λ in the income model accounts for this correlation and enables consistent estimates of β .

2.5.2. Empirical results

Table 2.8, which provides the estimation results of five specifications of the income model, includes a supply-side specification (i), a distance specification (ii), and three specifications used for robustness checks, (iii) – (v). In all five specifications the coefficient of the Mills ratio γ is statistically significant, which suggests that correcting for sample selection is important for analyzing non-agricultural income. The negative sign indicates that the error terms in the selection and income equations are negatively correlated. Thus, unobserved factors that correlate positively with the probability of RNAE tend to decrease the earnings prospects in the RNA sector. Given the heterogeneity of the RNA sector, and the fact that nearly half of RNAE is low-productivity, we have no clear expectation of the sign of this coefficient. It is, nonetheless, important to control for the selection process.

A comparison of models (i) and (ii) shows that exclusion of geographical variables does not cause major bias in the estimates of the supply-side coefficients. Most coefficients are very similar, which is an important difference compared with the probit models in the previous section. Since all of the geographical variables in the distance specification (ii) are significant, we choose this as our reference model. The coefficients of the human capital proxies – age and education – are large and of the expected sign. There are positive and increasing returns at all four educational levels.⁹ Relative to zero education, having five to eight, or nine to 11, years of education raises non-agricultural earnings by around 23 and 46 percent, respectively. As one would expect, there is a positive premium to being self-employed (at different levels of wealth) or an employer (of different sizes) compared to being an informal employee. The estimated earnings premium for a job in the formal sector is about 27 percent. Gender and ethnicity play different roles in earnings compared to selection. Although men have a lower probability of employment in the non-agricultural sector, they have higher earnings than women in non-agricultural activities. This is most likely a result of the selection mechanism

⁹ We suspect that the magnitude of any bias on the education coefficients due to the endogeneity of the educational decision is likely to be small. Laszlo (2005) rejects the endogeneity of education with Peruvian data. Card (1999, p. 1855) writes: “The “best available” evidence from the latest studies of identical twins suggests a small upward bias (in the order of 10%) in the simple OLS estimates.”

How Important Is Economic Geography for Rural Non-agricultural Employment?

TABLE 2.8. *Empirical results: earned non-agricultural income*

	(i) Supply-side	(ii) Distance	(iii) No outlier mun.	(iv) No migrants, urban ext., rural towns	(v) Employees only
Supply-side factors					
Age	0.048	0.051	0.050	0.046	0.054
Age squared	-0.000	-0.001	-0.001	-0.000	-0.001
Male	0.476	0.457	0.459	0.505	0.446
Education 1-4	0.088	0.105	0.103	0.075	0.101
Education 5-8	0.200	0.231	0.226	0.162	0.208
Education 9-11	0.410	0.461	0.463	0.320	0.433
Education 12	0.961	1.020	1.020	0.790	1.046
Migrant	0.058	0.055	0.063		0.034
Hours (log)	0.342	0.341	0.338	0.336	0.303
Formal sector	0.275	0.268	0.272	0.275	0.274
Self-employed	0.195	0.193	0.198	0.193	
Employer 1	0.835	0.839	0.835	0.812	
Employer 2	1.145	1.143	1.192	1.043	
Employer 3	1.377	1.380	1.398	1.280	
Self-empl*HH wealth	0.331	0.339	0.340	0.360	
HH education	0.034	0.034	0.034	0.028	0.035
Demand-side factors and transaction costs					
Distance 500 (log)		-0.010	-0.009	(-0.003)	-0.011
Distance 250 (log)		-0.013	-0.013	-0.020	-0.019
Distance 100 (log)		-0.004	(0.002)	(-0.002)	-0.006
Distance 50 (log)		0.006	0.005	0.010	0.006
Urban extension		-0.029	-0.049		-0.015*
Rural town		-0.038	-0.035		-0.038
Urbanization		-0.040	-0.039	-0.115	-0.052
Telephones		0.599	0.734	0.580	0.576
Electrification		-0.126	-0.139	-0.077	-0.180
Constant	2.647	2.687	2.677	2.961	2.933
Mills ratio	-0.20	-0.12	-0.12	-0.25	-0.18
Wald χ^2	233,487	242,507	185,050	111,975	197,415
Sample size	1,724,822	1,724,822	1,369,849	1,005,911	1,724,822
Uncensored observations	469,667	469,667	365,296	231,487	340,931

Note: The dependent variable is log of earned non-agricultural income. All coefficients are statistically significant at the one-percent level except in the following cases: * denotes significance at 10-percent level; coefficients within parentheses are not significant at the 10-percent level. All specifications include racial and macro regional control variables. Standard errors are available from the authors.

discussed in the previous section: women are more likely to engage in the low-paid forms of non-agricultural work. There is some evidence of racial earnings differentials. While there is not much difference in the probabilities of blacks and people of mixed origin participating in the RNA sector, both groups earned between eight and 10 percent less than whites, controlling for all other observables in model (ii).

The results suggest that local characteristics tend to affect employment outcomes and income prospects in different ways. Whereas nearly all the locational variables have the expected relationship with employment, the results are more mixed when the dependent variable is earnings. Three of the four distance coefficients are negative and statistically significant, as expected, but one is positive. All four coefficients are quite small. Unexpectedly, earnings appear to fall slightly with residence in an urban extension or rural town, and with urbanization. A possible explanation for the lack of a strong positive relationship between earnings and location relates to an excess supply of labor for RNA jobs, which prevents wages from rising. Thus, while non-agricultural employment prospects improve for those rural residents who live close to more urban locations, competition with the urban residents and unemployment may imply that there is no clear earnings premium associated with residence in these locations. Although some locational variables affect RNA earnings positively, and others negatively, perhaps the most important finding is that the magnitude of the effects is substantially smaller for earnings than for employment. Residence in an urban extension or rural town, for example, is associated with a 20 to 50 percentage point increase in the probability of RNAE. The corresponding figures for earnings are only three to four percent.

As with the probability model, we perform multiple robustness checks on the income model, three of which are reported in Table 2.8. In column (iii) the sample is trimmed to exclude municipal outliers (similar to specification (i) in Table 2.6) in order to find out whether these cause some of the unexpected results in the reference model. With the exception of the coefficient of telephones, the quantitative changes are small. One of the distance coefficients is statistically insignificant. Endogeneity of the decision to migrate across municipalities, or to live in an urban extension or rural town, could bias the results in the same way as in the employment model. Specification (iv) jointly excludes migrants and individuals who live in rural towns or urban extensions. This reduces the number of uncensored observations by more than 50 percent and renders two of the distance coefficients insignificant, but does not cause

any qualitative changes in the coefficients. Finally, in specification (v) we reduce the sample to employees only (excluding the self-employed and employers) to obtain a more homogeneous sample and to account for the possible problem of income measurement for non-wage earners. This narrowing of the sample does not generate any important changes in the coefficient estimates. The principal conclusion, that the local economic geography matters much more for the probability of employment than for earnings, is robust to the tests in Table 2.8.

2.6. CONCLUSION

With 30 percent of the rural labor force in Brazil having their principal source of earned income in RNAE, it is clear that non-agricultural activities take place far beyond the urban periphery. In this chapter we seek to expand the understanding of the rural non-agricultural sector by empirically testing the extent to which the employment and income opportunities in the sector relate to the economic geography. The empirical analysis shows that demand side factors, such as local market size, play an important role in shaping an individual's probability of having RNAE. Proxies for transaction costs, such as distance to markets, correlate negatively with RNAE. This does not mean that supply-side factors are unimportant for employment outcomes. Even when controlling for the local context, the coefficients of education, gender, and other individual characteristics are statistically and economically significant. Individual characteristics also play a key role in sorting people across low- and high-productivity RNAE. In contrast to the probability of employment, however, our results suggest that the local economic context is considerably less important for shaping earnings.

The implications for the poverty alleviation potential of the RNA sector are mixed. Among those who participate in the RNA sector, poverty is lower. But, given that the empirical results suggest that the local economic context and personal characteristics jointly shape employment and earnings prospects in the rural economy, RNAE is unlikely to be a feasible pathway out of poverty for the majority of the rural poor. On the one hand, RNAE opportunities are lowest in locations where poverty is highest. On the other, access to well-remunerated non-agricultural jobs depends on assets – such as human capital – that the poor are most likely to lack. The question of access, and thus of education and

training, is especially important for women who have a much higher probability than men of finding RNA jobs that pay even less than the average local wages in agriculture. While these jobs may help to diversify household income risk, they do not appear to provide movement up the occupational ladder.

Policies that are aimed at supporting rural non-agricultural employment should be designed with the role of location in mind. It is evident that the rural non-agricultural sector is viable, diverse, and important, but its potential to improve the living standards of rural households is conditioned by distance to larger markets, infrastructure, and the level of local aggregate demand. The benefits of geographical concentration of economic activities become increasingly important as agriculture absorbs less and less of the rural labor force. Therefore, in addition to programs that support specific types of RNA activities, such as tourism or agricultural processing, promotion of RNAE should constitute one component of a strategy aimed at developing small and medium-sized cities. These locations may provide an attractive alternative to migration to metropolitan areas.

REFERENCES

- Abdulai, A. & Delgado, C. L. (1999). Determinants of Nonfarm Earnings of Farm-Based Husbands and Wives in Northern Ghana. *American Journal of Agricultural Economics*, 81 (1), 117–130.
- Baltenweck, I. & Staal, S. (2007). Beyond One-Size-Fits-All: Differentiating Market Access Measures for Commodity Systems in the Kenyan Highlands. *Journal of Agricultural Economics*, 58 (3), 536–548.
- Card, D. (1999). The Causal Effect of Education on Earnings. In O. Ashenfelter, & D. Card (Eds.), *Handbook of Labor Economics, Volume 3A* (pp. 1801–1863). Amsterdam: Elsevier Science.
- Corral, L. & Reardon, T. (2001). Rural Nonfarm Incomes in Nicaragua. *World Development*, 29 (3), 427–442.
- de Janvry, A. & Sadoulet, E. (1993). Rural Development in Latin America: Relinking Poverty Reduction to Growth. In M. Lipton, & J. van der Gaag (Eds.), *Including the Poor: Proceedings of a Symposium Organized by the World Bank and the International Food Policy Research Institute* (pp. 249–277). Washington D.C.: World Bank.
- de Janvry, A. & Sadoulet, E. (2001). Income Strategies among Rural Households in Mexico: The Role of Off-farm Activities. *World Development*, 29 (3), 467–480.
- Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*. Baltimore: The Johns Hopkins University Press for the World Bank.
- Dirven, M. (2004). Rural Non-farm Employment and Rural Diversity in Latin America. *CEPAL Review*, 83 (August), 47–65.
- Echeverría, R. G. (2000). Options for Rural Poverty Reduction in Latin America and the Caribbean. *CEPAL Review*, 70 (April), 151–164.
- Ellis, F. (2000). *Rural Livelihoods and Diversity in Developing Countries*. Oxford: Oxford University Press.
- Escobal, J. (2001). The Determinants of Nonfarm Income Diversification in Rural Peru. *World Development*, 29 (3), 497–508.
- Ferreira, F. & Lanjouw, P. (2001). Rural Nonfarm Activities and Poverty in the Brazilian Northeast. *World Development*, 29 (3), 509–528.
- Graziano da Silva, J. & del Grossi, M. E. (2001). Rural Nonfarm Employment and Incomes in Brazil: Patterns and Evolution. *World Development*, 29 (3), 443–453.

- Haggblade, S., Hazell, P., & Reardon, T. (2007). Introduction. In S. Haggblade, P. Hazell, & T. Reardon (Eds.), *Transforming the Rural Nonfarm Economy: Opportunities and Threats in the Developing World* (pp. 3–24). Baltimore: The John Hopkins University Press.
- Harris, C. D. (1954). The Market as a Factor in the Localization of Industry in the United States. *Annals of the Association of American Geographers*, 44 (4), 315–348.
- Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47 (1), 153–161.
- Isgut, A. E. (2004). Non-farm Income and Employment in Rural Honduras: Assessing the Role of Locational Factors. *Journal of Development Studies*, 40 (3), 59–86.
- Kay, C. (2005). Reflections on Rural Poverty in Latin America. *European Journal of Development Research*, 17 (2), 317–346.
- Lanjouw, J. O. & Lanjouw, P. (2001). The Rural Non-farm Sector: Issues and Evidence from Developing Countries. *Agricultural Economics*, 26 (1), 1–23.
- Lanjouw, P. (2001). Nonfarm Employment and Poverty in Rural El Salvador. *World Development*, 29 (3), 529–547.
- Laszlo, S. (2005). Self-employment Earnings and Returns to Education in Rural Peru. *Journal of Development Studies*, 41 (7), 1247–1287.
- Mellor, J. W. (1976). *The New Economics of Growth: A Strategy for India and the Developing World*. Ithaca: Cornell University Press.
- Ney, M. G. & Hoffmann, R. (2007). Educação, Atividades Nao-Agrícolas e Desigualdade de Renda no Brasil Rural. Paper presented at the XLV Congresso da Sociedade Brasileira de Economia, Sociologia, e Administração Rural, July 22–25, Londrina, Brazil.
- Quijandría, B., Monares, A., & de Peña Montenegro, R. U. (2001). *Assessment of Rural Poverty: Latin America and the Caribbean*. Santiago: IFAD, Latin America and Caribbean Division.
- Reardon, T., Berdegue, J., & Escobar, G. (2001). Rural Nonfarm Employment and Incomes in Latin America: Overview and Policy Implications. *World Development*, 29 (3), 395–409.
- Tomich, T. P., Kilby, P., & Johnston, B. F. (1995). *Transforming Agrarian Economics: Opportunities Seized, Opportunities Missed*. Ithaca: Cornell University Press.

- van de Walle, D. & Cratty, D. (2004). Is the Emerging Non-farm Market Economy the Route out of Poverty in Vietnam? *Economics of Transition*, 12 (2), 237–274.
- World Bank (2003). *Rural Poverty Alleviation in Brazil: Toward an Integrated Strategy*. Washington D.C.: World Bank.
- World Bank (2007). *World Development Report 2008: Agriculture for Development*. Washington D.C.: World Bank.
- Yang, D. T. & An, M. Y. (2002). Human Capital, Entrepreneurship, and Farm Household Earnings. *Journal of Development Studies*, 68 (1), 65–88.

Acknowledgements: This chapter is based, in part, upon work supported by a grant from the United States Agency for International Development (AID) through a program called BASIS/CRSP. We are thankful for valuable comments received from Sonja Opper, Fredrik Wilhelmsson, Juliano Assunção, Johan Eklund, Niclas Berggren, Andreas Bergh, and from five anonymous referees of *World Development*. We also thank Eustáquio Reis, Marcia Pimentel, and the Applied Economics Research Institute (IPEA) for assistance in constructing some of the key geographical data used in our empirical analysis. The opinions are solely those of the authors.

APPENDIX: CORRELATION MATRIX

TABLE 2.A1. *Correlation matrix of variables used in empirical analysis (part 1 of 4)*

	RNAE	RNAE low	RNAE high	N-A inc	Age	Male	Black	Asian	Mixed	Indigenous	Education
RNAE	1.00										
RNAE low	0.64	1.00									
RNAE high	0.65	-0.17	1.00								
Non-agr inc	0.23	0.02	0.27	1.00							
Age	-0.11	-0.11	-0.04	0.01	1.00						
Male	-0.17	-0.21	-0.01	0.00	0.01	1.00					
Black	-0.02	0.00	-0.03	-0.01	0.02	0.02	1.00				
Asian	0.00	0.00	0.01	0.01	0.02	0.00	-0.01	1.00			
Mixed	-0.03	0.00	-0.04	-0.04	-0.05	0.04	-0.24	-0.04	1.00		
Indigenous	0.00	0.01	-0.01	0.00	-0.01	0.00	-0.02	0.00	-0.07	1.00	
Education	0.31	0.10	0.29	0.14	-0.31	-0.12	-0.10	0.03	-0.19	-0.03	1.00
Edu1-4	-0.10	-0.04	-0.09	-0.04	0.05	0.02	-0.02	-0.01	-0.01	-0.02	-0.20
Edu5-8	0.11	0.08	0.07	0.03	-0.25	-0.02	-0.04	0.00	-0.07	-0.01	0.41
Edu9-11	0.22	0.07	0.21	0.08	-0.16	-0.09	-0.04	0.02	-0.08	-0.01	0.61
Edu12	0.12	-0.01	0.17	0.14	-0.01	-0.05	-0.02	0.04	-0.06	0.00	0.35
Migrant	0.05	0.03	0.04	0.04	0.15	0.00	0.00	0.02	-0.02	-0.03	0.03
Formal	0.25	0.07	0.26	0.07	-0.07	0.03	0.00	0.00	-0.06	-0.01	0.17
Self-emp	-0.09	-0.06	-0.05	0.00	0.22	0.19	-0.02	0.01	-0.05	-0.01	-0.08
Emp1	0.01	-0.02	0.03	0.04	0.05	0.03	-0.01	0.02	-0.02	0.00	0.04
Emp2	0.01	-0.01	0.03	0.06	0.03	0.02	-0.01	0.02	-0.02	0.00	0.03
Emp3	0.02	-0.01	0.04	0.14	0.03	0.02	-0.01	0.03	-0.02	0.00	0.05
Hours	-0.02	-0.08	0.05	0.04	0.02	0.23	-0.01	0.01	-0.10	-0.03	0.03
HH adults	-0.02	0.00	-0.03	-0.02	-0.11	0.01	0.01	0.00	0.06	-0.01	0.01
HH edu	0.20	0.04	0.23	0.13	-0.01	-0.02	-0.09	0.03	-0.20	-0.04	0.48
HH wealth	0.20	0.03	0.22	0.16	0.07	-0.05	-0.11	0.06	-0.33	-0.07	0.46
Rural town	0.21	0.13	0.15	0.08	-0.02	-0.02	0.01	0.00	-0.01	-0.01	0.12
Urban ext	0.13	0.08	0.09	0.02	-0.01	-0.01	0.02	-0.01	0.08	0.01	0.00
Rural excl	-0.23	-0.13	-0.16	-0.06	0.03	0.02	-0.02	0.00	-0.07	-0.01	-0.07
North	0.01	0.04	-0.02	0.00	-0.03	0.04	0.01	0.00	0.13	0.09	-0.05
Northeast	-0.06	-0.06	-0.02	-0.06	-0.01	0.00	0.07	-0.02	0.30	-0.03	-0.28
Southeast	0.09	0.05	0.06	0.05	-0.01	0.02	0.01	0.02	-0.11	-0.03	0.14
Southeast	-0.03	-0.03	-0.01	0.01	0.04	-0.07	-0.09	0.00	-0.34	-0.02	0.21
Center-West	0.00	0.03	-0.03	0.01	0.01	0.05	-0.01	0.01	0.01	0.04	0.04
Urbanization	0.14	0.10	0.09	0.06	0.01	0.04	-0.01	0.03	-0.08	-0.02	0.14
Telephones	0.19	0.11	0.13	0.10	0.02	-0.04	-0.06	0.03	-0.24	-0.03	0.27
Electrification	0.12	0.06	0.09	0.06	0.03	-0.04	-0.08	0.01	-0.30	-0.06	0.28
Local inc 1	0.23	0.13	0.16	0.10	-0.02	-0.02	-0.01	0.02	-0.13	-0.03	0.19
Local inc 2	0.24	0.14	0.17	0.10	-0.02	-0.03	0.00	0.02	-0.05	-0.01	0.14
Local pop 1	0.23	0.13	0.17	0.10	-0.02	-0.02	0.00	0.02	-0.11	-0.04	0.17
Local pop 2	0.25	0.14	0.18	0.10	-0.02	-0.03	0.00	0.02	-0.04	-0.02	0.13
Dist50	-0.07	-0.02	-0.07	-0.02	-0.01	0.04	0.01	0.00	0.09	0.17	-0.07
Dist100	-0.06	-0.01	-0.07	-0.03	-0.02	0.03	0.01	0.00	0.13	0.14	-0.10
Dist250	-0.11	-0.05	-0.10	-0.04	-0.01	0.03	0.04	-0.01	0.16	0.11	-0.14
Dist500	-0.13	-0.06	-0.11	-0.04	-0.01	0.03	-0.01	-0.01	0.05	0.10	-0.05

How Important Is Economic Geography for Rural Non-agricultural Employment?

TABLE 2.A1. (part 2 of 4)

	Edu1-4	Edu5-8	Edu9-11	Edu12	Migrant	Formal	Self-emp	Emp1	Emp2	Emp3	Hours
Edu1-4	1.00										
Edu5-8	-0.47	1.00									
Edu9-11	-0.29	-0.13	1.00								
Edu12	-0.11	-0.05	-0.03	1.00							
Migrant	-0.01	0.01	0.00	0.04	1.00						
Formal	-0.04	0.07	0.12	0.06	0.10	1.00					
Self-emp	0.06	-0.05	-0.07	-0.03	0.00	-0.28	1.00				
Emp1	-0.01	0.00	0.02	0.04	0.02	-0.03	-0.05	1.00			
Emp2	-0.01	0.00	0.02	0.04	0.01	-0.02	-0.03	0.00	1.00		
Emp3	-0.02	0.00	0.02	0.07	0.02	-0.02	-0.03	0.00	0.00	1.00	
Hours	0.05	0.02	-0.02	-0.02	0.08	0.13	0.07	0.03	0.02	0.02	1.00
HH adults	-0.01	0.01	0.03	-0.02	-0.13	-0.05	-0.08	-0.02	-0.01	-0.01	-0.05
HH edu	-0.05	0.18	0.26	0.21	0.11	0.15	0.05	0.06	0.05	0.06	0.11
HH wealth	-0.03	0.14	0.24	0.21	0.09	0.19	0.03	0.08	0.07	0.09	0.16
Rural town	-0.06	0.06	0.09	0.03	0.09	0.13	-0.03	0.00	0.00	0.00	0.03
Urban ext	-0.03	0.01	0.02	0.00	0.03	-0.01	0.00	0.00	0.00	0.00	-0.04
Rural excl	0.06	-0.04	-0.07	-0.02	-0.08	-0.07	0.03	0.00	0.00	0.00	0.02
North	0.00	-0.01	-0.03	-0.02	0.08	-0.08	0.07	0.00	0.00	0.00	-0.03
Northeast	-0.05	-0.13	-0.09	-0.05	-0.17	-0.16	-0.04	-0.02	-0.01	-0.01	-0.23
Southeast	0.03	0.04	0.06	0.04	0.02	0.18	-0.08	0.02	0.01	0.01	0.13
Southeast	0.04	0.10	0.06	0.03	0.05	0.03	0.09	0.00	0.00	0.00	0.12
Center-West	-0.01	0.04	0.01	0.02	0.14	0.05	-0.01	0.02	0.01	0.01	0.08
Urbanization	-0.02	0.07	0.07	0.05	0.12	0.23	-0.06	0.02	0.01	0.02	0.14
Telephones	-0.03	0.12	0.13	0.08	0.13	0.21	0.00	0.02	0.01	0.02	0.16
Electrification	0.02	0.12	0.12	0.06	0.10	0.20	-0.04	0.02	0.01	0.02	0.17
Local inc 1	-0.03	0.09	0.10	0.05	0.09	0.21	-0.05	0.00	0.01	0.01	0.10
Local inc 2	-0.05	0.07	0.09	0.04	0.09	0.17	-0.05	0.00	0.00	0.01	0.05
Local pop 1	-0.03	0.07	0.10	0.05	0.06	0.21	-0.07	0.00	0.01	0.01	0.08
Local pop 2	-0.05	0.07	0.09	0.04	0.08	0.17	-0.05	0.00	0.00	0.01	0.04
Dist50	-0.01	-0.02	-0.04	-0.02	0.05	-0.07	0.02	0.00	0.00	-0.01	-0.02
Dist100	-0.01	-0.03	-0.05	-0.03	-0.01	-0.11	0.04	0.00	-0.01	-0.01	-0.05
Dist250	-0.02	-0.05	-0.06	-0.03	0.03	-0.15	0.07	-0.01	-0.01	-0.01	-0.09
Dist500	0.00	-0.01	-0.03	-0.03	0.12	-0.12	0.06	0.00	-0.01	-0.01	-0.02

Chapter 2

TABLE 2.A1. (part 3 of 4)

	HH adults	HH edu	HH wealth	Rural town	Urban ext	Rural excl	North	Northeast	Southeast	Southeast	Center-W.
HH adults	1.00										
HH edu	-0.01	1.00									
HH wealth	0.02	0.50	1.00								
Rural town	-0.04	0.13	0.14	1.00							
Urban ext	-0.01	0.01	-0.03	-0.04	1.00						
Rural excl	0.03	-0.08	-0.05	-0.43	-0.85	1.00					
North	0.00	-0.06	-0.23	-0.02	0.03	-0.02	1.00				
Northeast	0.10	-0.29	-0.43	-0.07	0.16	-0.11	-0.26	1.00			
Southeast	0.00	0.14	0.29	0.14	-0.08	0.00	-0.18	-0.47	1.00		
Southeast	-0.08	0.22	0.36	-0.05	-0.12	0.13	-0.16	-0.43	-0.29	1.00	
Center-West	-0.07	0.05	0.04	-0.02	-0.02	0.02	-0.07	-0.19	-0.13	-0.12	1.00
Urbanization	-0.06	0.16	0.24	0.16	-0.05	-0.04	-0.09	-0.23	0.30	-0.04	0.12
Telephones	-0.08	0.30	0.46	0.34	-0.08	-0.10	-0.12	-0.40	0.18	0.37	0.01
Electrification	-0.06	0.30	0.54	0.13	-0.04	-0.03	-0.40	-0.37	0.33	0.38	-0.02
Local inc 1	-0.04	0.20	0.33	0.56	-0.07	-0.21	-0.17	-0.30	0.47	0.02	-0.05
Local inc 2	-0.04	0.15	0.22	0.56	-0.03	-0.25	-0.07	-0.16	0.30	-0.05	-0.04
Local pop 1	-0.03	0.18	0.32	0.54	-0.06	-0.21	-0.25	-0.21	0.50	-0.04	-0.09
Local pop 2	-0.03	0.14	0.21	0.56	-0.03	-0.25	-0.09	-0.12	0.29	-0.07	-0.06
Dist50	-0.02	-0.08	-0.20	-0.07	0.00	0.04	0.33	-0.11	-0.10	-0.14	0.27
Dist100	0.01	-0.11	-0.27	-0.09	0.02	0.02	0.48	-0.06	-0.18	-0.16	0.16
Dist250	-0.01	-0.15	-0.35	-0.10	0.10	-0.04	0.28	0.21	-0.30	-0.19	0.11
Dist500	-0.01	-0.05	-0.16	-0.13	-0.01	0.08	0.44	-0.14	-0.19	0.01	0.11

TABLE 2.A1. (part 4 of 4)

	Urbaniz.	Telephones	Electrif.	Local inc 1	Local inc 2	Local pop 1	Local pop 2	Dist50	Dist100	Dist250	Dist500
Urbanization	1.00										
Telephones	0.31	1.00									
Electrif.	0.33	0.46	1.00								
Local inc 1	0.34	0.53	0.35	1.00							
Local inc 2	0.29	0.49	0.22	0.94	1.00						
Local pop 1	0.35	0.49	0.38	0.99	0.92	1.00					
Local pop 2	0.29	0.47	0.25	0.92	0.99	0.92	1.00				
Dist50	-0.11	-0.18	-0.41	-0.19	-0.14	-0.26	-0.18	1.00			
Dist100	-0.14	-0.21	-0.46	-0.25	-0.17	-0.32	-0.20	0.47	1.00		
Dist250	-0.22	-0.28	-0.54	-0.32	-0.23	-0.39	-0.27	0.47	0.42	1.00	
Dist500	-0.15	-0.19	-0.36	-0.31	-0.25	-0.40	-0.30	0.54	0.57	0.46	1.00

Note: The sample size is 1,724,822. All correlations (including municipal variables) are on individual level. See Table 2.4 for definitions and further details.