Locational Determinants of Rural Non-agricultural Employment: Evidence From Brazil[*]

ERIK JONASSON
Lund University, Sweden
erik.jonasson@nek.lu.se

and

STEVEN M. HELFAND
University of California, Riverside
steven.helfand@ucr.edu

Summary: By paying particular attention to the local economic context, this paper analyzes the determinants of non-agricultural employment and earnings in non-agricultural jobs. The empirical analysis is based on the Brazilian Demographic Census, allowing for disaggregated controls for the local economy. Education stands out as one of the key determinants of employment outcome and earnings potential. Failure to control for locational effects, however, can lead to biased estimation of the importance of individual and household-specific characteristics. The empirical results show that local market size and distance to population centers have a significant impact on non-agricultural employment prospects and earnings.

Key Words: Rural non-agricultural employment, economic geography, Latin America, Brazil.

I. INTRODUCTION

Rural non-agricultural activities have 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, emphasizing that the rural economy consists of much more than just agriculture (Reardon et al., 2001). Among the roles of the rural non-agricultural (RNA) sector are its potential 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). These roles of the RNA sector, and particularly the potential to be a pathway out of poverty for rural landless households and land constrained family farmers, have been recognized in rural development strategies during the past two decades (de Janvry and Sadoulet, 1993; Echeverría, 2000; Quijandría et al., 2001; World Bank 2003 and 2008).

The extent to which rural non-agricultural employment (RNAE) is able to reduce poverty ultimately depends on rural households' access to such employment and the income prospects in these activities. The accessibility and the income prospects, in turn, are likely to depend jointly on supply-side effects (individual and household characteristics), demand-side effects (characteristics of the relevant labor markets), and market participation costs. Household asset endowments on their own will not generate upward income mobility if there is insufficient demand for labor, or if market participation is very costly due to physical distance to markets and underdeveloped infrastructure that obstruct the mobility of people, capital, goods, and information. Employment prospects and earnings potential as a function of location is an argument in line with Harris's (1954) market potential analysis of industry localization, further developed by Krugman (1991) and Fujita et al. (1999).

In this paper we seek to assess the importance of supply, demand, and participation cost effects on an individual's probability of engaging in RNAE and on earned income in the RNA sector. The previous empirical literature on this topic has been concerned mainly with the supply-side determinants of RNAE. Even though there is a consensus that location does matter 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] Similarly, the Brazilian literature on RNAE has been based almost exclusively on the national household surveys (PNAD).[2] PNAD is only representative at the state level, thus providing little insight into locational determinants of employment and income. Besides controlling for broad geographical differences, with regional dummy variables, previous studies have utilized various indicators to capture local economic conditions. These include rural sub-category (Ferreira and Lanjouw, 2001), distance to nearest health center (Corral and Reardon, 2001), neighborhood average household income, local urbanization and electricity (Isgut, 2004), and the number of urban and rural population centers within one hour's commuting distance (de Janvry and Sadoulet, 2001). Failure to control adequately for local determinants may lead to biased estimates of the role of individual or household-specific characteristics when drawing inference about rural household labor allocation and livelihood strategies.

To reach a deeper understanding of demand-side effects and the role of market-participation costs, our study utilizes a fuller set of variables than previous studies to describe the local economic geography. By utilizing data from the Brazilian Demographic Census of year 2000, we are able to test for the role of municipal-level economic factors such as local market size and distance to population centers. As expected, our results show that personal as well as household characteristics do matter for employment outcomes as well as for income earnings potential. Demand-side factors and proxies for participation costs, however, also have strong effects on the probability of being engaged in RNAE. Market size and the degree of urbanization increase RNAE 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. These variables have a much smaller, and less consistent, impact on earnings. Our conclusions about the importance of the local economic geography stand up to a number of robustness checks that seek to address endogeneity concerns related to where people with different characteristics choose to reside.

The next section of the paper gives a brief overview of rural employment and the RNA sector in Brazil. Section three contains the empirical analysis of determinants of RNAE. The fourth section deals with earnings in rural areas. Following Heckman (1979), it corrects the earnings regressions for possible sample selection. Section five concludes.

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

Due to Brazil's size, locational factors are of particular importance for the viability of the RNA sector and the potential of RNAE to alleviate rural poverty. With only 19 percent of its population residing in rural areas, Brazil is a highly urbanized country.[3] Whereas 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). The North and Northeast regions of Brazil, where 30 percent of the population lives in rural areas, are the least urbanized. In the densely populated Southeast, only nine percent of households are rural. 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. The Census classifies the census tracts where rural households live 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 live in rural towns or agglomerations, and only three percent are found in areas considered as urban extensions. As Table 1 indicates, rural remoteness goes hand in hand with poverty. Rural poverty was above 70 percent in the North and Northeast, and below 45 percent in the other three regions. Poverty rates within each region also increase with rural sub-category, rising from 42 percent in rural areas classified as extensions to 62 percent in rural areas exclusive of agglomerations.[4]

[TABLE 1]

The rural non-agricultural sector

Of the rural labor force, about 70 percent had their principal employment in agriculture (cultivation, animal rearing, and forestry). The remaining 30 percent had principal employment in rural non-agricultural activities. Empirical evidence shows that this share has increased over time. Graziano da Silva and del Grossi (2001) note that employment in the rural non-agricultural sector grew 1.2 percent annually between 1981 and 1992, rising to 2.5 percent per year between 1992 and 1997. The corresponding figures for employment in agriculture were 0.4 and -2.2 percent, respectively. As shown in Table 2, there are regional variations in the composition of the rural labor force. The Northeast is not only the poorest region, but also the region with the lowest share of the rural labor force in the non-agricultural sector (25 percent). Rural non-agricultural employment was greatest in the highly urbanized Southeast region (39 percent). Table 2 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. Most people in the rural labor force can be divided into three broad categories: wage laborers, self-employed, and unpaid workers (working either as unpaid household members, trainees, or in subsistence agriculture). Each of these groups constitutes about one third of the rural labor force. Half of the wage laborers are engaged in RNAE, a fourth of the self-employed, and only a very small share of the unpaid workers. This implies that the majority of those who work in the RNA sector are wage laborers. In fact, 68 percent of RNAE takes the form of wage labor, 27 percent is self-employment, and the rest is non-remunerated. The share of RNAE is considerably higher among women (42 percent) than men (25 percent).

[TABLE 2]

Traditionally, the rural non-agricultural 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 local linkages are in Brazil relative to countries with smaller farms, lower levels of technology, and weaker links to the world market. In this spirit, Graziano da Silva and del Grossi (2001) argue that the composition of the rural non-agricultural 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. This view is supported by the maps in Figures 1 and 2 of the Brazilian Northeast and Southeast. The maps depict the share of the rural labor force whose principal occupation is in RNAE in each municipality. It is evident that non-agricultural activities are more prevalent in the proximity of capital cities and highly urbanized areas. The pattern is most pronounced in the densely populated areas surrounding São Paulo, Rio de Janeiro, and Belo Horizonte in Figure 2.[5]

[FIGURES 1 and 2]

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 3 shows that the five largest RNA sectors employ almost 70 percent of the non-agricultural labor force. Manufacturing employs a considerably higher 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, domestic services is the largest sector of non-agricultural employment. The most noticeable difference between male and female non-agricultural work is that women dominate the jobs classified as domestic services and education. Men are to a higher extent engaged in traditional male-dominated activities such as construction and transportation.

[TABLE 3]

Rural non-agricultural income

On average, people earned higher incomes in the rural non-agricultural sectors than in agriculture. This was true for wage laborers and self-employed, as well as for men and women. Table 4 shows average monthly earnings in the five 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 were the only major RNA sector in which average earnings were lower than in agriculture. The self-employed earned more than wage laborers, and in all sectors men earned more than women.

Even though average earnings in most of the RNA sectors were higher than in agriculture, there were also many low-paid non-agricultural jobs. We divided RNAE into two groups depending on earnings relative to agriculture. If an individual was engaged in RNAE and had earnings below the average municipal earnings of wage laborers in agriculture, we considered the individual as being engaged in low-productivity RNAE. Those who earned above this average were classified as being engaged in high-productivity RNAE. With this categorization, although average earnings in RNAE were 25 percent higher than in agriculture, only 53 percent of the non-agricultural labor force was engaged in high-productivity RNAE. As indicated by the last column in Table 4, in the educational sector more than two-thirds of the labor force had high-productivity jobs. In domestic services, in contrast, only one-fifth of employment was high productivity.[6]

[TABLE 4]

Non-agricultural activities are sometimes considered primarily as a means of income diversification among rural households (Barrett et al., 2001; Ellis, 2000). For households in rural Brazil, however, using RNAE for this purpose does not appear to be a deliberate strategy of the majority of households. We defined households as specialized in agriculture if they derived 90 percent or more of their earned income from agriculture, specialized in non-agriculture if they derived 90 percent or more from RNAE, and pluriactive otherwise. Only 14 percent of rural households were considered pluriactive by this definition.[7] Noticeable in terms of specialization is that richer households were to a larger extent engaged in RNAE than poorer households. In the lowest income quintile of rural households, 21 percent of the households were specialized in non-agriculture, whereas the share was almost twice as high (37 percent) in the highest income quintile. The positive correlation between household income and RNAE is consistent with several other country studies in Latin America (Ecuador – Lanjouw, 1998; Mexico – de Janvry and Sadoulet, 2001; and Peru – Escobal, 2001).

Differences in average earnings suggest that the rural non-agricultural sector could potentially offer a pathway out of rural poverty. To assess this potential, we need to understand the joint forces of demand-side factors, location, and labor market participation costs, alongside individual characteristics, in shaping non-agricultural opportunities in rural areas. In the following two sections we analyze the importance of these supply and demand side effects, first by assessing what influences the probability that people in the rural labor force engage in non-agricultural activities, and second by examining what determines their earnings.

III. DETERMINANTS OF RURAL NON-AGRICULTURAL EMPLOYMENT

In this section we report the results of a probability analysis of engagement in rural non-agricultural employment. The analysis first focuses on the probability that a rural worker is engaged in any kind of RNAE. We estimated a binomial probit model in which the dependent variable indicates whether the individual was engaged in RNAE as opposed to agriculture. Due to the heterogeneity of the RNA sector, we then used a multinomial probit model to estimate jointly the probabilities of engaging in high- and low-productivity RNAE in comparison with agriculture.

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_i. If V_i exceeds a certain threshold value, V_i*, the individual is engaged in RNAE; otherwise, he or she is engaged in agriculture. The latent variable V can be thought of, in this case, as the rural worker's expected earnings if participating in the rural non-agricultural sector. The threshold V* could be the shadow wage for agricultural work on the own farm or the wage rate on the agricultural labor market.[8] Thus, the probability of individual i being engaged in RNAE, P_i, is modeled as the probability that V_i exceeds V_i*:

$$P_i = PROB_i\left(RNAE_i = 1 \middle| X_{ijk}, H_{jk}, M_k\right) = PROB(V_i \geq V_i^*) \qquad (1)$$

in which X_{ijk}, H_{jk}, and M_k denote vectors of variables to characterize, respectively, individual i, household j to which the individual belongs, and municipality k in which the household is situated. Let v_i denote the difference V_i–V_i*, which is the expected net benefit of RNAE. This net gain is modeled as a log-linear function of X, H, and M:

$$v_i = X_{ijk}\beta_1 + H_{jk}\beta_2 + M_k\beta_3 + \varepsilon_{ijk} \qquad (1')$$

where the βs 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 $F(.)$ be the standard normal cumulative distribution function of ε; then individual i's probability of engaging in RNAE can be estimated as:

$$P_i = PROB\left(X_{ijk}\beta_1 + H_{jk}\beta_2 + M_k\beta_3 \geq -\varepsilon_{ijk}\right) = F\left(X_{ijk}\beta_1 + H_{jk}\beta_2 + M_k\beta_3\right) \qquad (1'')$$

In the second approach, involving the estimation of a multinomial probit model, we applied the distinction between low- and high-productivity RNAE that we introduced in the previous section. The benchmark that separates the two employment types is the average agricultural earnings of wage laborers in each municipality. To estimate the effect of individual, household, and local characteristics on the probability of having a certain type of employment, the model was specified as:

$$P_i^e = PROB_i\left(EMP_i = e \middle| X_{ijk}, H_{jk}, M_k\right) = F\left(X_{ijk}\beta_1^e + H_{jk}\beta_2^e + M_k\beta_3^e\right) \qquad (2)$$

in which P^e is the probability that individual i has employment e; e being either i) agricultural work, ii) low-productivity RNAE, or iii) high-productivity RNAE. As in the previous probability models, P^e is modeled as a log-linear function of X, H, and M.

Variables and data source

The Demographic Census data are based on a sample of more than 20 million observations (12 percent of the population), constructed to be representative at the municipal level. There were 5,507 municipalities defined for the 2000 Census, with an average population of approximately 30,000 people.[9] In the empirical analysis, we used the rural adult labor force as the base sample, which included 1.7 million individual observations. Adults were defined as everyone age 15 years or older. Anyone reporting an occupation was considered as a participant in the labor force, including unpaid workers. Descriptive statistics of the variables included in the regression analysis are provided in Table 5.

The endogenous RNAE variable was based on reported principal occupation. The individual characteristics included in the vector X include age, gender, race, education, and migrant-status. Age, age-squared, 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 was controlled for by four dichotomous variables, which are based on the number of completed years of schooling. Zero education is the benchmark category and contains almost 30 percent of the rural labor force. Gender was included to control for systematic differences between male and female workers in terms of job preferences, work hours, but also demand-side effects such as gender discrimination in payment schemes. Race, controlling for black, mixed, Asian and indigenous groups, was included for similar reasons. A dummy variable for migrants was included, indicating whether the individual has moved to the municipality rather than always lived there. People who have moved could have a lower opportunity cost of staying on the farm. Migration could also be an indicator of unobserved ability and risk-taking, and hence willingness to engage in the employment with the highest returns for the individual. The remaining individual variables are used in the income analysis and will be discussed below.

[TABLE 5]

Household characteristics (*H*) include the number of adult household members, average education in the household (excluding individual *i*'s education), and an index of household wealth. The number of adults was included to control for opportunities for employment diversification: the larger the labor supply in the household, the more the opportunities to devote some household labor to non-agricultural activities. Average years of schooling among other household members are 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. A proxy for household wealth was constructed that summarizes a vector of characteristics of the domicile.[10] Household wealth should 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 were included indicating whether the household was situated in a rural town or urban extension as opposed to a rural exclusive area.

Municipal-level characteristics were included to assess the importance of local demand and market participation costs for the employment outcome. To capture local market size, we used two distance-weighted measures of local aggregate income, in the same spirit as Harris's (1954) market potential analysis.[11] Both measures include the total income of people in the municipality plus total income in surrounding municipalities weighted by distance, but differ in the weighting scheme. The first variable, *Inc1d*, was defined as the sum over all municipalities of municipal income, weighted by the inverse of the distance D_l from a typical rural household in the municipality of origin k to the seat of municipality l. Income refers to the sum of all income received by households in each municipality as reported in the Demographic Census. The distance D_l aggregates two components: a) the estimated distance d_k from a typical rural household in municipality k to its own municipal seat, and b) the distance from the seat of municipality k to the seat of municipality l. Distance to the own municipal seat was estimated by assuming that the municipality was circle shaped, with the municipal seat in the center, and with the average rural household located at a distance equal to one half the radius from the seat. Thus, $d_k = 0.5\sqrt{A/\pi}$, where A is the area of the municipality in square-kilometers. When k=l the distance between municipalities equals zero, and D_l equals the intra-municipal distance d_k. This implies that the size of the market – both within and outside of one's own municipality – is a decreasing function of distance. The second measure of market size, *Inc100d*, uses a linearly declining weight that only takes into account municipalities (l*) within a 100-kilometer distance of a typical rural household (weight=1 at zero km, weight=0 at 100km). Formally, the two local aggregate income measures were defined as:

$$Inc1d_k = \sum_l Income_l \left(1/D_l\right) \qquad (4)$$

$$Inc100d_k = \sum_{l \in l^*} Income_l \left(1 - D_l/100\right) \qquad (4')$$

As can be seen in Table 5 by the large difference in means between the two variables, *Inc1d* discounts much more heavily for distance than *Inc100d*. For example, income in a municipality at 50 kilometers of distance only gets a 2-percent weight with *Inc1d,* but a 50-percent weight with *Inc100d*. While both variables have the advantage of taking into account the size of the market within and outside of the own municipality, the weighting scheme in *Inc100d* seems more realistic in terms of potential RNAE.[12]

We used a collection of variables as proxies for market participation costs. The shares of rural households with access to a telephone line and to electric lighting were included to capture the level of municipal rural infrastructure. The share of households in the municipality that was classified as urban was used to reflect the hypothesis that urbanization is correlated with infrastructural development which, in turn, should lower the costs of participation in input and output markets. The own municipality may or may not be the relevant marketplace. Therefore measures of distance to population centers were included as an alternative proxy for access to markets. Using D_l, distances were estimated to the nearest municipality with 50–100, 100–250, 250–500, and more than 500 thousand people, respectively. In contrast to the local demand variables *Inc1d* and *Inc100d*, which emphasize the total size of the local market, the distance measures focus on the importance of transactions costs associated with access to markets. The distance variables also permit capturing non-linearities in the relationship between RNAE and distance to markets of different sizes.

Estimation results

The results from the binomial probit model are provided in Table 6. The variables were added stepwise to determine their explanatory power and their effect on other coefficient estimates. The marginal effects reported in the table 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 for the dichotomous variables. Due to the sample size, nearly all coefficients are statistically significant at least at the one percent level and the standard errors are quite small. For this reason, all tables identify those coefficients that are not significant at the 1-percent level. Standard errors are available from the authors.

Model (i) includes only individual variables. When household characteristics are controlled for in model (ii), the coefficient estimates on some individual characteristics change significantly. The marginal effects of all four educational levels, for example, decrease substantially when household characteristics are added, suggesting that the educational variables were, in part, capturing the effect of the excluded household variables. Omitted variables bias is also evident when model (ii) is compared to models (iii) and (iv) that include the geographical variables. The coefficients on migrants and on household wealth, for example, both decline significantly. Thus, failure to adequately control for the local economic geography can generate significant bias. For this reason, we focus most of the discussion below on models (iii) and (iv).

Table 6 shows that human capital affects positively 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, has little impact on the probability of RNAE, whereas having five to eight years of education increases the probability by about 15 percentage points, while nine to 11 years of education increases the probability by around 30 percentage points. Consistent with the descriptive data presented in Table 2, women have a higher probability of engaging in RNAE. Compared to whites, Asians have a lower probability of engaging in RNAE while indigenous people have a higher probability. For blacks and people of mixed origin, the probability deviates little from whites. People who have moved from one municipality to another – migrants – are more likely to engage in non-agricultural activities.

[TABLE 6]

Several observations are warranted on the effects of the household variables. The positive coefficients on 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 also influence employment outcomes. The effect of wealth is quantitatively more important. Based on the coefficient in model (v), a one standard deviation increase in the wealth index is associated with a 4.9 percentage point increase in the probability of RNAE. The number of household adults, in contrast, has a small and negative effect on RNAE, speaking against the employment diversification hypothesis. Even though larger households have more labor hours to allocate to various activities, this does not substantially affect an individual's probability of having RNAE as a principal occupation.

Specifications (iii) and (iv) provide insight into the extent to which local conditions matter for employment outcomes. Comparing the pseudo R^2 in columns (i)-(iv) of Table 6 shows that, as a group, the locational variables explain more of the variance in the probability of RNAE than do the household variables. After controlling for supply side factors, the locational variables explain near two thirds as much variance as the individual characteristics do.

The positive coefficient on the local aggregate income variable in column (iii) – *Inc100d* – suggests that RNAE opportunities improve with the amount of local aggregate demand. In fact, 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. The magnitude of this increase is, perhaps, easier to interpret in terms of population. The mean municipal population was approximately 27,000. When we constructed a local population variable – *Pop100d* – that was analogous to *Inc100d*, the mean local population was about 560,000. A one standard deviation increase in the local population is associated with a 12.2 percentage point increase in the probability of RNAE.

Specifications (iii) and (iv) show that all but one of the proxies for participation costs are statistically significant of the 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.[13] Residence in a rural town increases the probability by more than 20 percentage points. These point estimates are considerably larger than those found by Ferreira and Lanjouw (2001), but are consistent with the descriptive data in Table 2. Possible explanations for this difference are that their study is restricted to the Northeast of Brazil, and is based on the PNAD household survey rather than the Census. Controlling for other locational factors, 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 column (*iv*) also suggest that distance to population centers matters for RNAE prospects. The greater the distance to large municipalities of all four size categories, the lower is the probability that an individual will engage in RNAE. At the mean of 260 km, an additional standard deviation of distance – or 195 km – away from municipalities with greater 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.

We also observe that, although distance from municipalities with larger populations has a greater absolute impact on the probability of RNAE, the impact rises less than proportionately with the size of these municipalities. For example, if we compare the impact of moving 100 km closer to municipalities in the largest class (over 500,000 people) with those in the 50-100 thousand class, we observe increases in the probability of RNAE of 2.8 and 0.5 percentage points respectively. The ratio of impacts (2.8/0.5 = 5.4) is substantially smaller than the ratio of mean populations (21.8) in these two groups, confirming that the impact does not rise proportionately with the population of the municipality. The ratio of impacts divided by the ratio of populations yields a statistic of 0.25. Similar calculations for the 100-250 and 250-500 thousand classes in relation to the 50-100 thousand class yield statistics of 0.76 and 0.73. Thus, on a per capita basis, municipalities with 50-100 thousand residents appear to be more successful at generating non-agricultural employment for rural residents than municipalities with 100-500 thousand residents, and substantially more so than the 31 largest municipalities.

The one case where we find mixed evidence for participation costs relates to the proxies for rural infrastructure. The shares of rural households with telephones and electricity, respectively, point in different directions regarding their effect on RNAE. Telephones are associated with a higher probability of RNAE, whereas electricity is associated with a lower probability. With only 6 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 complements the other locational variables. Regarding the negative coefficient on electricity, we note that the simple correlation between electricity and RNAE is positive 0.26, and that electricity is highly correlated with many of the other variables in models (iii) and (iv).[14] After controlling for all of the different dimensions of location and local market size that are correlated with this variable, electricity in rural areas does not appear to have an independent positive effect on the probability of RNAE.

We performed three types of robustness checks to detect potential bias in our results. First, the estimated effects of the individual and household characteristics could be influenced by unobserved local factors that we were unable to control for with our vector of local level variables. In order to explore this issue, instead of using a set of municipal level variables, the model was estimated with municipal fixed effects and the urban extension / rural town dummies that vary by census tract. The results in column (*v*) show that the coefficients on all non-municipal level variables are quite similar to specifications (iii) and (iv) that include municipal characteristics. The largest differences relate to the race dummies, and the urban extension variable, yet none of the qualitative results are altered. We conclude that the estimated coefficients in the probability model are not altered dramatically by the failure to include additional municipal level variables in the model.

A second concern relates to the possible endogeneity of several of the locational variables. It could be, for example, that unobserved individual characteristics that have a higher return in RNAE induce people with those characteristics to move to locations where they have a higher probability of finding RNAE. If true, the coefficients on urban extensions, rural towns, and the family of distance variables, for example, could 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-estimated model (iv) first without migrants, then without individuals who lived in urban extensions and rural towns, and finally without both groups. When migrants were removed from the model, the coefficients on extensions and towns fell by only one and 10 percent respectively. Similarly, when towns and extensions were removed, the coefficient on migrants only declined from 0.025 to 0.021. In the model without extensions, towns or migrants, the sample drops from 1.7 million down to 1.1 million observations. Nonetheless, there was not a single case where a coefficient changed signs. The magnitude of some of the locational variables declined, and in other cases remained steady. Among the distance variables, the coefficients on dist500, dist250500, and dist100250 fell by 36, 18, and zero percent. We conclude that although there is some evidence in favor of the hypothesis of endogenous sorting of the rural population, this does not alter any of the qualitative results. A final robustness check was to replace, *Inc100d* in specification (iii) with, consecutively, *Inc1d, Pop100d,* and *Pop1d.* The results were similar in all cases.

The results from the multinomial probit model are provided in Table 7. Due to computational intensity, the model was estimated with a 5 percent random sample from the data used to estimate the binomial models. The base specification in Table 7 includes personal and household characteristics, whereas the distance specification includes municipal characteristics and the family of distance variables. The multinomial results are highly consistent with the binomial results. There are a number of results from the multinomial model that are not captured in the binomial model. 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. 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-productive RNAE. The results also suggest that human capital matters for the probability of having non-agricultural employment among the rural labor force, but as in the case of gender, it does not affect low- and high-productivity RNAE equally. Even having only one through four years of education, relative to zero, increases the probability of high-productivity RNAE by around two percentage points, but has no statistically significant impact on the probability of low-productivity RNAE. Similarly, at higher levels of schooling where the impact on RNAE is much larger, most 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. Household wealth also increases the probability of high-productivity RNAE, but in this case it leads to a lower probability of both other types of employment.

[TABLE 7]

The second specification in Table 7 shows that the locational variables generally increase the probability of both low-and high-productivity RNAE. The participation cost variables appear to have a slightly larger impact on the probability of low-productivity RNAE. As for the role of distance, the results show that a one standard deviation move away from municipalities in all four “large” classes leads to a combined reduction of 4.4 and 5.7 percentage points in the probability of low- and high-productivity RNAE, respectively. The impact of local aggregate income – not shown here due to space limitations – also has a slightly larger impact on high-productivity RNAE. Thus, we conclude that locational factors have a strong impact on 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, play key roles in sorting across types of RNAE.

IV. DETERMINANTS OF EARNED INCOME

To identify the key determinants of earned income in rural areas, and to assess the extent to which the structural coefficients differ between agricultural and non-agricultural workers, we estimated income regressions separately for the agricultural and non-agricultural labor force. Many of the same explanatory variables were included as in the probit analysis. In the probit model we assumed that these variables determined the potential net gain of non-agricultural work (v). In this case, we estimated their effect on earnings.

Estimation method

Due to censoring of the data, we applied the Heckman (1979) sample selection model. In the non-agricultural income model, unpaid and agricultural employees are censored because they report zero non-agricultural income, and in the agricultural income model, unpaid and non-agricultural employees are censored. For simplicity, we describe the non-agricultural (NA) model here. The results from the probit model in the previous section suggest that individual characteristics and other factors determine the selection process into RNAE, so that non-agricultural employees differ systematically from agricultural employees. 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. Our approach assumes that selection into paid RNAE is determined by a model analogous to (1) in the previous section. The only difference is that we now focus on *paid* RNAE, rather than *any*

RNAE. Then, accounting for the results of the selection process, we assume that income can be modeled as a log-linear function of individual, household, and regional characteristics:

$$y_{ijk}^{NA} = X_{ijk}\beta_{1}^{NA} + H_{jk}\beta_{2}^{NA} + M_{k}\beta_{3}^{NA} + \gamma^{NA}\lambda_{ijk}^{NA} + \eta_{ijk}^{NA} \qquad (6)$$

in which y^{NA} is the logarithm of non-agricultural income of the individual, *X, Z*, and *M* are vectors containing explanatory individual, household, and municipal characteristics, λ is the inverse Mills ratio, η is the error term, and β and γ are coefficients to be estimated. A test of $\gamma^{NA} = 0$ 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. The inclusion of λ in the income model accounts for this correlation and permits obtaining consistent estimates of β. The approach for estimating the agricultural model is identical. One simply needs to replace (NA) with (A) in the model above. A complete analysis of agricultural income determination would include a farm production function, which takes productive assets other than household labor into account. Given that our primary purpose is to compare the extent to which the effects of the explanatory variables differ for agricultural and non-agricultural income, and given that the Demographic Census does not contain data on productive assets, we applied the same specifications for the two models. We did, however, include two proxies for productive assets. We interacted the self-employment dummy with our household wealth index and, following Ney and Hoffmann (2007), we divided employers into three groups based on the number of people they employed.[15]

Results

Table 8 provides estimation results of three model specifications for earnings in RNAE and in agriculture. The base specification includes only supply side variables, the income specification adds locational variables with local aggregate income, and the distance specification adds locational variables along with the distance variables. 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). In our context, we believe that family size should have no influence on individual earnings. Thus, it enters the selection equation, but is excluded from the income equation. We also use all of the locational variables from the distance specification to help identify the base model, the four distance variables to help identify the income model, and the Inc100d variable to help identify the distance model. Finally, the 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. In model (i) of Table 8, the coefficient on λ is significantly different from zero in both specifications. In models (ii) and (iii), which include locational variables, λ remains significant in the non-agricultural specifications, but becomes very small or statistically insignificant in the agricultural specifications. Thus, we conclude that correcting for sample selection is important for RNAE, but much less so, or not at all, in agriculture.

Specifications (ii) and (iii) in Table 8, reveal that only some of the factors that affect the selection process positively into rural non-agricultural employment also affect earned income positively. The human capital proxies – age and education – are large and of the expected sign. There are positive returns at all four educational levels in both sectors, yet the returns to education are substantially higher in non-agriculture than in agriculture. Relative to zero education, having four to eight, or nine to 11, years of education raises non-agricultural earnings by around 16 and 37 percent, respectively. For agriculture, the effect is about 12 and 22 percent. Ney and Hoffman (2007) argue that differences in human capital explain the largest share of the variance of earnings in non-agricultural employment, whereas due to the importance of land as an asset for agricultural production, differences in physical capital explain more of the variance of earnings in agriculture. We find a similar result here. Relative to an informal employee, the impact of being self-employed (at different levels of wealth), or being an employer (of different sizes), are more important in agriculture.

In contrast to human and physical capital, gender and ethnicity play different roles in earnings than in selection. Although men had a lower probability of employment in the non-agricultural sector, they have higher earnings than women in both agriculture and non-agriculture. In fact, the gender earnings gap is larger in the non-agricultural sector. This is most likely a result of the selection mechanism 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 discrimination as well. While there was not much difference in the probabilities of blacks and people of mixed origin participating in the RNA sector, both groups earn between 8 percent and 10 percent less than whites, with little difference across sectors.

The results in specifications (*ii*) and (*iii*) also suggest that local characteristics tend to affect employment outcomes and income prospects in different ways. Whereas nearly all locational variables had the expected effect on RNAE, the results are much more mixed when the dependent variable is earnings. For example, non-agricultural earnings rise slightly with local aggregate income (Inc100d) and proximity to municipalities with at least 100,000 residents, yet appear to fall slightly with residence in an urban extension or rural town, and with urbanization. A possible explanation for the lack of any 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 – implies that there is no clear earnings premium associated with residence in these locations.

Although some locational variables affect RNA earnings positively, and others negatively, an important result is that the magnitude of the impacts is substantially smaller for earnings than for employment. Residence in an urban extension or rural town, for example, was associated with a 20 to 50 percentage point increase in the probability of RNAE. The impact on earnings is only in the range of three to four percent. The same general conclusions apply to agricultural earnings: the impact of locational variables is mixed, and the magnitude of the impacts tends to be quite small.

Another important finding shown in Table 8 is that the point estimates on some of the supply-side factors change considerably when locational variables are included in the model. The returns to education in RNAE, for example, are estimated to be as much as 32 percent higher in models (ii) and (iii) than in the base model. In agriculture, the inclusion of locational variables increases the estimated returns to education for people with 5-11 years of education by 36 percent to 47 percent. As in the probit models, these examples suggest that the estimated coefficients on some key variables, such as education, are likely to be biased if the local economic environment is not controlled for appropriately.

[TABLE 8]

V. 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. We have claimed in this paper that the prospects for RNAE depend jointly on supply-side factors, demand-side factors, and the magnitude of market participation costs. The empirical analysis shows that, when holding individual and household characteristics constant, demand side factors such as local market size have a strong impact on an individual's probability of having RNAE. Proxies for transactions costs, such as distance to markets, correlate negatively with RNAE. This does not mean that supply-side effects are unimportant for employment outcomes. Even when controlling for local factors, the effects 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. Market access and market size mattered for non-agricultural earnings, but much less so than personal and household-specific characteristics.

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 with the local economic context and personal characteristics jointly determining 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 and HH wealth – 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 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 support the RNA sector should be designed with the role of location in mind. It is evident that the rural non-agricultural sector is viable and important, but the sector's potential 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, and to the extent that farm households are unable to escape poverty solely with agricultural income. Therefore, rather than focusing on the RNA sector as such, promotion of RNA activities should be an ingredient in strategies aimed at developing viable rural economic centers, that is, small and medium sized cities that are "growth motors" (Reardon et al., 2001) in themselves or well-connected with the broader urban economy. These rural growth motors could provide an attractive alternative to migration to metropolitan areas. Relative to large cities, they could also serve as places that offer lower costs of living for their residents and lower costs of production for their businesses.

REFERENCES

Atlas of Human Development (2003). Database developed by ESM Consultoria; data by PNUD.

Barrett, C. B., T. Reardon, and P. Webb (2001), “Nonfarm Income Diversification and Household Livelihood Strategies in Rural Africa: Concepts, Dynamics, and Policy Implications”, *Food Policy*, 26(4), pp. 315–331.

Corral, L. and T. Reardon (2001), “Rural Nonfarm Incomes in Nicaragua”, *World Development*, 29(3), pp. 427–442.

Campanhola, C. and J. Graziano da Silva (eds), *O Novo Rural Brasileiro: Uma Análise Nacional e Regional,* São Paulo: Emprapa, 2000.

de Janvry, A. and E. Sadoulet (1993), “Rural Development in Latin America: Relinking Poverty Reduction to Growth”, in Lipton and van der Gaag (eds), *Including the Poor.* World Bank, Washington D.C.

de Janvry, A. and E. Sadoulet (2001), ”Income Strategies Among Rural Households in Mexico: The Role of Off-farm Activities”, *World Development*, 29(3), pp. 467–480.

Dirven, M. (2004), “Rural Non-farm Employment and Rural Diversity in Latin America”, *CEPAL Review*, vol. 83, August.

Echeverría, R. G. (2000), “Options for Rural Poverty Reduction in Latin America and the Caribbean”, *CEPAL Review,* vol. 70, April.

Ellis, F. (2000) *Rural Livelihoods and Diversity in Developing Countries.* Oxford University Press, New York.

Escobal, J. (2001), “The Determinants of Nonfarm Income Diversification in Rural Peru”, *World Development*, 29(3), pp. 497–508.

Ferreira, F. and P. Lanjouw (2001), “Rural Nonfarm Activities and Poverty in the Brazilian Northeast”, *World Development*, 29(3), pp. 509–528.

Figueiredo, F., S. Helfand, and E. Levine (2007) “Income versus Consumption Measures of Rural Poverty and Inequality in Brazil”, paper presented at *Pobreza Rural no Brasil: O Papel as Políticas Públicas*, conference in Brasilia, DF, April 17-18,.

Fujita, M., P. Krugman and A. J. Venables (1999), *The Spatial Economy - Cities, Regions, and International Trade.* The MIT Press, Cambridge.

Graziano da Silva, J. and M. E. del Grossi (2001), “Rural Nonfarm Employment and Incomes in Brazil: Patterns and Evolution”, *World Development*, 29(3), pp. 443–453.

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, pp. 315-348.

Heckman, J. J. (1979), “Sample Selection Bias as a Specification Error”, *Econometrica*, 47(1), pp. 153–161.

Helfand, S. and E. Levine (2004), “Changes in Brazilian Rural Poverty from 1991 to 2000: A Preliminary Analysis of Determinants, and an Assessment of Alternative Levels of Household Aggregation”. Report prepared for OECD, 2004.

Isgut, A. E. (2004), “Non-farm Income and Employment in Rural Honduras: Assessing the Role of Locational Factors”, *Journal of Development Studies*, 40(3), pp. 59–86.

Kay, C. (2005), “Reflections on Rural Poverty in Latin America”, *European Journal of Development Research*, 17(2), pp. 317–346.

Krugman, P. (1991), "Increasing Returns and Economic Geography", *Journal of Political Economy*, 99(3), pp. 483-499.

Lanjouw, J. O. and P. Lanjouw (2001), “The Rural Non-farm Sector: Issues and Evidence from Developing Countries”, *Agricultural Economics*, vol. 26, pp. 1–23.

Lanjouw, P. (1998), “Rural Nonagricultural Employment and Poverty in Ecuador”, *Economic Development and Cultural Change*, 48(1), pp. 91–122.

Laszlo, S. (2005), “Self-employment Earnings and Returns to Education in Rural Peru”, *Journal of Development Studies*, 41(7), pp. 1247–1287.

Mellor, J. W. (1976), *The New Economics of Growth. A Strategy for India and the Developing World.* Cornell University Press, Ithaca.

Ney, M. G., and R. Hoffmann (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.

OECD (2005), *OECD Review of Agricultural Policies: Brazil*.

Quijandría, B., Monares, A., and de Peña Montenegro, R. U. (2001), *Assessment of Rural Poverty. Latin America and the Caribbean*. IFAD, Latin America and Caribbean Division, Santiago.

Reardon, T., J. Berdegue and G. Escobar (2001), “Rural Nonfarm Employment and Incomes in Latin America: Overview and Policy Implications”, *World Development*, 29(3), pp. 395–409.

Tomich, T. P., P. Kilby, and B. F. Johnston (1995), *Transforming Agrarian Economics. Opportunities Seized, Opportunities Missed.* Cornell University Press, Ithaca.

World Bank (2003), *Rural Poverty Alleviation in Brazil – Toward an Integrated Strategy*. World Bank, Washington D.C.

World Bank (2008), *World Development Report 2008: Agriculture for Development*. World Bank, Washington D.C.

Yang, D. T. and M. Y. An (2002), “Human Capital, Entrepreneurship, and Farm Household Earnings”, *Journal of Development Studies,* 68(1), pp. 65–88.

TABLE 1: Population and Poverty

	Brazil	*North*	*North-east*	*South-east*	*South*	*Center-West*
Population, total (millions)	169.8	12.9	47.7	72.4	25.1	11.6
Population, rural	31.8	3.9	14.8	6.9	4.8	1.5
Share of population that is rural	0.19	0.30	0.31	0.09	0.19	0.13
Poverty, headcount ratio						
Brazil	0.32	0.48	0.55	0.19	0.19	0.24
Urban	0.25	0.39	0.45	0.16	0.16	0.21
Rural	0.61	0.70	0.77	0.42	0.35	0.43
Urban extensions	0.42	0.43	0.57	0.24	0.26	0.30
Isolated rural agglomerations or towns	0.58	0.66	0.72	0.43	0.33	0.42
Rural areas exclusive of agglomerations	0.62	0.72	0.79	0.45	0.36	0.44

Source: Demographic Census 2000, authors' calculation.

TABLE 2: Percentage of Rural Labor Force by Sector of Principal Occupation

	Cultivation	*Animal rearing*	*Forestry*	*Non-agriculture*	*Total*
Region					
Brazil	0.56	0.12	0.02	0.30	*1.00*
North	0.52	0.12	0.04	0.32	*1.00*
Northeast	0.66	0.07	0.03	0.25	*1.00*
Southeast	0.43	0.16	0.01	0.39	*1.00*
South	0.56	0.15	0.02	0.27	*1.00*
Center-West	0.27	0.41	0.02	0.30	*1.00*
Rural sub-category					
Urban extension	0.08	0.02	0.00	0.90	*1.00*
Rural towns	0.38	0.06	0.02	0.54	*1.00*
Rural exclusive	0.60	0.13	0.02	0.25	*1.00*
Employment status					
Wage labor	0.31	0.15	0.02	0.52	*1.00*
Self-employed	0.60	0.11	0.03	0.26	*1.00*
Unpaid	0.83	0.10	0.02	0.05	*1.00*
Gender					
Men	0.59	0.14	0.02	0.25	*1.00*
Women	0.48	0.07	0.03	0.42	*1.00*

Source: Demographic Census 2000, authors' calculation.

TABLE 3: Rural Non-Agricultural Employment by Sub-Sector (% with Primary Occupation)

	Region						*Employment*		*Gender*	
	Brazil	*North*	*North-east*	*South*	*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
Total	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*

Source: Demographic Census 2000, authors' calculation.

TABLE 4: Rural Non-Agricultural Income by Sector (R$ per Month, 2000)

Sector	*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	346	294	479	416	236	0.53
Agriculture	280	198	346	296	170	n/a

Note: The exchange rate R$/US$, August 2000, was 1.81.
Source: Demographic Census 2000, authors' calculation.

TABLE 5: Summary Statistics of Variables Used in the Empirical Analysis

Variable	*Mean*	*Std. Dev.*	*Description*
Dependent variables			
RNAE	0.30	0.451	Has RNAE as principal employment (d)
RNAE low	0.15	0.348	Low-productivity RNAE (d)
RNAE high	0.15	0.350	High-productivity RNAE (d)
Y^{NA}	346	1170	Earned non-agricultural income, for Y^{NA} >0
Y^{AGR}	280	1331	Earned agricultural income, for Y^{AGR} >0
Individual characteristics			
Age	36	14.712	Individual's age
Male	0.71	0.452	Gender, 1 if male (d)
Black	0.07	0.257	Race – black (d)
Asian	0.002	0.048	Race – Asian (d)
Mixed	0.44	0.496	Race – mixed (d)
Indigenous	0.01	0.086	Belongs to indigenous group (d)
Education	3.55	3.240	Individual's years of education
Edu14	0.50	0.50	1 to 4 years of education (d)
Edu58	0.12	0.32	5 to 8 years of education (d)
Edu911	0.08	0.265	9 to 11 years of education (d)
Edu12	0.01	0.104	12 or more years of education (d)
Migrant	0.37	0.482	Has migrated from other municipality (d)
Formal sector	0.16	0.356	Paid employee in the formal sector (d)
Inform. sector	0.25	0.434	Paid employee in the informal sector (d)
Self-employed	0.32	0.464	Self-employed (d)
Employer 1	0.005	0.072	Employer with 1–2 employees (d)
Employer 2	0.002	0.047	Employer with 3–5 employees (d)
Employer 3	0.002	0.042	Employer with 6 or more employees (d)
Unpaid	0.27	0.447	Unpaid worker (d)
Hours	42	15.13	Hours worked per week
Household characteristics			
HH adults	3.32	1.636	Number of adults in the household
HH edu	3.96	3.140	Av. years of education among adults in HH, excluding ind.
HH wealth	-0.65	0.744	Household wealth index
Urban ext.	0.03	0.145	Residence in urban extension (d)
Rural town	0.09	0.268	Residence in rural town (d)
Rural excl.	0.87	0.306	Residence in rural area, exclusive of towns/extensions (d)
North	0.10	0.287	Residence in North (d)
Northeast	0.42	0.491	Residence in Northeast (d)
South	0.23	0.427	Residence in South (d)
Southeast	0.20	0.408	Residence in Southeast (d)
Center-West	0.05	0.220	Residence in Center-West (d)
Municipal characteristics			
Mun urban	0.60	0.22	Share of urban households in municipality
Mun tel	0.06	0.09	Share of rural households with fixed telephone
Mun el	0.75	0.26	Share of rural households with electric lighting
Inc1d	73.7	45.40	Distance-weighted local income, million R$ (see eqn. 4)
Inc100d	178	531	Distance-weighted local income, million R$ (see eqn. 4')
Pop1d	561,716	1,107,277	Distance-weighted local population
Pop100d	236,416	97,358	Distance-weighted local population
Dist500	260	195	Distance to municipality with >500,000 people, km
Dist250500	207	174	Distance to mun., 250-500,000 people, km
Dist100250	124	130	Distance to mun., 100-250,000 people, km
Dist50100	76	74	Distance to mun., 50-100,000 people, km

Note: Variables indicated by (d) are dichotomous variables, taking value 1 if true, 0 otherwise. Continuous variables are converted into log-form in the model estimations. Sample size is 1,724,822. For the municipal variables, the unweighted municipal-level mean is given.

TABLE 6: Results of the binomial probit model estimation: probability of any RNAE

	Marginal effects on probability				
	(i)	*(ii)*	*(iii)*	*(iv)*	*(v)*
Supply-side factors					
Age	0.013	0.009	0.009	0.009	0.008
Age2	-0.000	-0.000	-0.000	-0.000	0.000
Male	-0.141	-0.148	-0.152	-0.153	-0.184
Black	-0.003††	0.038	0.013	0.009	(0.000)
Asian	-0.095	-0.138	-0.146	-0.144	-0.093
Mixed	(0.000)	0.029	0.016	0.013	0.002††
Indigenous	(0.006)	0.074	0.100	0.103	0.042
Edu14	0.037	0.015	0.022	0.022	0.026
Edu58	0.232	0.153	0.147	0.149	0.153
Edu911	0.422	0.301	0.306	0.310	0.323
Edu12	0.563	0.369	0.420	0.419	0.448
Migrant	0.061	0.048	0.012	0.023	0.025
Hh adults		-0.011	-0.006	-0.006	-0.005
Hh edu		0.014	0.011	0.011	0.010
Hh wealth		0.106	0.058	0.061	0.066
Demand-side factors					
Inc100d			0.051		
Participation costs					
Urban ext.			0.520	0.502	0.378
Rural town			0.239	0.237	0.221
Mun_urban			0.119	0.099	
Mun_tel			0.298	0.250	
Mun_el			-0.114	-0.093	
Dist500				-0.074	
Dist250500				-0.039	
Dist100250				-0.011	
Dist50100				-0.004	
Macro-regional controls	Yes	Yes	Yes	Yes	No
Municipal fixed effects	No	No	No	No	Yes
Pseudo-R^2	0.108	0.127	0.189	0.197	
Observations	1,724,822	1,724,822	1,724,822	1,724,822	1,724,822

Note: The dependent variable is *RNAE* (dichotomous variable to indicate RNAE as opposed to agricultural employment). Marginal effects refer to the change in probability of engaging in RNAE, given a small change in a continuous variable or a discrete change in a dichotomous variable. All coefficients are statistically significant at the 1-percent level other than in the following cases: †† Denotes significance at 5-percent level and coefficients within parentheses are not significant at the 10-percent level. Standard errors are available from the authors.

TABLE 7: Results of the multinomial probit model estimation: probability of employment category

	i) base			*ii) distance*		
	Agricultural employment	*Low-prod. RNAE*	*High-prod. RNAE*	*Agricultural employment*	*Low-prod. RNAE*	*High-prod. RNAE*
Supply-side factors						
Age	-0.010	-0.005	0.016	-0.010	-0.006	0.016
Age2	0.000	0.000	-0.000	0.000	0.000	-0.000
Male	0.152	-0.177	0.024	0.158	-0.182	0.024
Black	-0.026	0.030	(-0.004)	(0.006)	0.011†	-0.017
Asian	0.081††	(-0.015)	-0.066	0.099	(-0.026)	-0.073
Mixed	-0.029	0.021	0.007††	-0.011††	0.012	(-0.001)
Indigenous	-0.072	0.069	(0.003)	-0.087	0.074	(0.013)
Edu14	-0.015	(-0.005)	0.021	-0.022	(-0.001)	0.023
Edu58	-0.154	0.036	0.118	-0.150	0.031	0.118
Edu911	-0.293	0.047	0.246	-0.302	0.046	0.255
Edu12	-0.349	-0.058	0.407	-0.401	-0.046	0.447
Migrant	-0.045	0.028	0.017	-0.020	0.009	0.010
Hh adults	0.010	-0.003	-0.008	0.005	(0.001)	-0.006
Hh edu	-0.014	0.004	0.011	-0.011	(0.001)	0.010
Hh wealth	-0.102	0.017	0.084	-0.056	-0.013	0.070
Participation costs						
Urban ext.				-0.513	0.293	0.220
Rural town				-0.238	0.130	0.107
Mun_urban				-0.106	0.088	0.018
Mun_tel				-0.220	0.194	(0.026)
Mun_el				0.084	(-0.004)	-0.080
Dist500				0.074	-0.039	-0.035
Dist250500				0.042	-0.012	-0.030
Dist100250				0.008	-0.008	(0.000)
Dist50100				(0.001)	0.004††	-0.005
Macro-regional controls	Yes			Yes		
Observations	86,231			86,231		
Wald chi-square	11,758			13,884		

Note: The dependent variable is employment category, *e*, where *e* is i) agricultural work, ii) $RNAE^{LOW}$, or iii) $RNAE^{HIGH}$. For each independent variable, the marginal effect refers to the change in probability of being in employment category *e*, given a small change in a continuous variable or a discrete change in a dichotomous variable. All coefficients are statistically significant at the 1-percent level other than in the following cases: †† Denotes significance at 5-percent level, † denotes significance at 10-percent level; coefficients within parentheses are not significant at the 10-percent level. Standard errors are available from the authors.

Table 8: Regression results of income model: agricultural and non-agricultural earned income

	i) base		*ii) income*		*iii) distance*	
	Non-ag.	*Ag.*	*Non-ag.*	*Ag.*	*Non-ag.*	*Ag.*
Supply-side factors						
Age	0.048	0.031	0.050	0.025	0.050	0.026
Age2	-0.000	-0.000	-0.001	-0.000	-0.001	-0.000
Male	0.475	0.489	0.458	0.350	0.458	0.363
Black	-0.105	-0.084	-0.103	-0.083	-0.103	-0.081
Asian	0.114	0.183	0.077	0.121	0.078	0.125
Mixed	-0.082	-0.091	-0.079	-0.078	-0.080	-0.079
Indigenous	-0.029†	-0.134	(-0.021)	-0.060	(-0.021)	-0.062
Edu14	0.025	0.040	0.033	0.044	0.033	0.044
Edu58	0.145	0.087	0.165	0.121	0.165	0.118
Edu911	0.334	0.152	0.369	0.224	0.369	0.217
Edu12	0.882	0.600	0.924	0.718	0.923	0.708
Migrant	0.058	0.106	0.053	0.093	0.055	0.091
Hours	0.342	0.351	0.341	0.342	0.341	0.343
Formal sector	0.277	0.380	0.270	0.361	0.270	0.364
Self-employed	0.197	0.394	0.195	0.389	0.195	0.389
Employer 1	0.839	1.067	0.843	1.065	0.843	1.066
Employer 2	1.144	1.404	1.143	1.408	1.143	1.411
Employer 3	1.378	1.899	1.380	1.907	1.380	1.910
Self-empl*HH wealth	0.337	0.411	0.345	0.402	0.344	0.395
HH education	0.035	0.034	0.035	0.035	0.035	0.035
Demand-side factors						
Inc100d			0.008	-0.002†		
Participation costs						
Urban extension			-0.031	(0.022)	-0.036	(0.009)
Rural town			-0.043	-0.036	-0.043	-0.033
Mun_urban			-0.034	0.042	-0.043	0.039
Mun_tel			0.600	0.881	0.590	0.868
Mun_el			-0.128	0.028	-0.120	0.011†
Dist500					-0.010	0.035
Dist250500					-0.012	-0.029
Dist100250					-0.003††	-0.017
Dist50100					0.006	0.009
Constant	2.737	2.315	2.554	2.667	2.791	2.603
Macro-regional controls	Yes	Yes	Yes	Yes	Yes	Yes
Lambda	-0.21	0.19	-0.14	-0.03	-0.14	(-0.01)
Observations	1,724,822	1,724,822	1,724,822	1,724,822	1,724,822	1,724,822
Censored obs.	1,255,155	950,456	1,255,155	950,456	1,255,155	950,456
Wald chi-square	232,103	270,577	239,840	299,628	240,554	301,521

Note: The dependent variables are log of earned income from principal employment. All coefficients are statistically significant at the 1-percent level other than in the following cases: †† Denotes significance at 5-percent level, † denotes significance at 10-percent level; coefficients within parentheses are not significant at the 10-percent level. Standard errors are available from the authors.

Figure 1: Rural Non-Agricultural Employment in the Brazilian Northeast

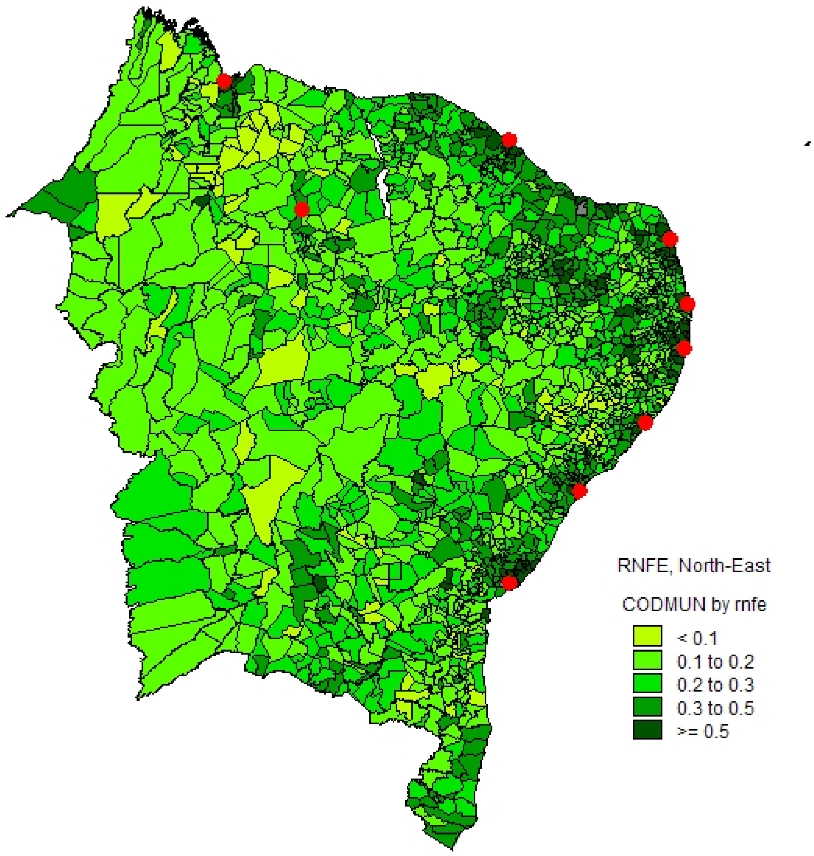
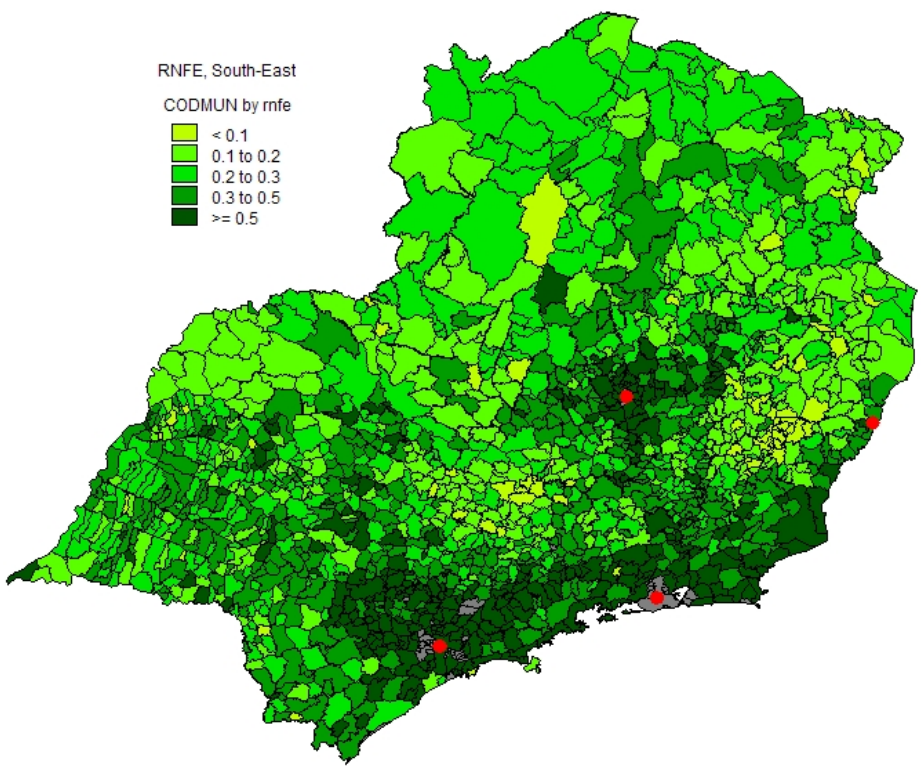


Figure 2: Rural Non-Agricultural Employment in the Brazilian Southeast



* This paper is based, in part, upon work supported by a grant from the United States Agency for International Development (AID) through a subcontract from BASIS/CRSP at the University of Wisconsin, Madison awarded to the Regents of the University of California, Riverside. The opinions, findings, conclusions, and recommendations are solely those of the authors. We are thankful for valuable comments received from participants at BASIS seminars in Rio de Janeiro and Brasilia, as well as from participants in seminars at IPEA-RJ, and UFF.

[1] In her recent 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 is still scant, but there is no doubt that distance and the transaction costs that ensue play a role both directly and indirectly…"

[2] Ney and Hoffmann (2007) is the one exception that we are aware of. They also utilize the 2000 Demographic Census. In contrast to our paper, they do not estimate models of employment in the RNA sector, nor do they prioritize locational effects. Their paper focuses on the relative importance of human and physical capital for explaining earnings.

[3] There is considerable debate in Brazil about the appropriate definition of "rural" vs urban "areas". In this paper, 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 almost 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] The poverty headcount ratio reported in this paper uses a poverty line set at 75 Reais per month, which corresponds to half the minimum wage of August 2000. This poverty line was also used by the Atlas of Human Development, IPEA/UNDP. The Demographic Census, and the national household surveys (PNAD), only contain information on monetary income. The poverty rates reported here are similar to income-based poverty measures reported by OECD (2005) based on the Census, and by the World Bank (2003) based on the 1996 PNAD survey. For a detailed analysis of the differences between income- and expenditure-based poverty measures in rural Brazil, see Figueiredo, Helfand, and Levine (2007).

[5] Given that rural households that live close to densely populated areas are within commuting distance to urban non-agricultural jobs, average regional shares tend to overstate the extent of rural non-agricultural employment available to most rural residents. In fact, 65 percent of the rural population lives in municipalities with less than 30 percent RNAE. Only 16 percent live in municipalities with over 50 percent RNAE (the darkest color in Figures 1 and 2).

[6] Ferreira and Lanjouw (2001) classify entire sub-sectors as high- or low-productivity based on average income in relation to the poverty line. Our approach, which classifies individuals rather than sub-sectors, and uses the local agricultural wage rather than the poverty line, is consistent with the assumptions underlying the probability model that is described in the following section.

[7] Our approach is different from what is commonly used in Brazil. The *Rurbano* project, for example, defines pluriactivity based on the primary and secondary occupations of household members, rather than on income shares (see Campanhola and Graziano da Silva (2000) for details). They also consider as pluriactive households that work on-farm and in agricultural wage labor. In contrast to the annual household surveys (PNAD), the Demographic Census does not permit identifying whether secondary occupations are agricultural or non-agricultural. For this reason, and because secondary occupations only accounted for 2 percent of rural income in 2000, our approach seemed satisfactory. It does, however, lead to a slightly lower estimate of the share of pluriactive households. Graziano da Silva and del Grossi (2001, Table 3) report 19.4 percent for the year 1997.

[8] In a broader sense, V could also be interpreted as a subjective utility measure of the individual, so that RNAE is chosen if the expected utility of RNAE is higher than the expected utility (V^*) of agricultural work.

[9] We excluded 61 municipalities from our analysis because they did not have rural areas.

[10] The proxy was 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.

[11] We thank Eustáquio Reis, Marcia Pimentel, and the Applied Economics Research Institute (IPEA) for assistance with the construction of the local market size and distance variables.

[12] Analogous population variables—*Pop1d* and *Pop100d*—were constructed to check for robustness.

[13] We recognize that the point estimates reflect correlation, not causation. Especially in the case of location of residence, endogeneity could be a serious concern. We address this issue with several robustness tests below.

[14] The simple correlations between electricity and telephones, local population, and local income are all between 0.40 and 0.51. There is also a strong negative correlation, ranging between –0.42 and –0.62, between electricity and the family of distance variables in model (iv).

[15] Small, medium, and large employers were identified as those that employed 1–2, 3–5, and 6 or more employees, respectively.