Labor Markets in Transformation: Case Studies of Latin America

Jonasson, Erik

2009

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Labor Markets in Transformation: Case Studies of Latin America

Erik Jonasson

Lund Economic Studies Number 155

Distributed by the Department of Economics
Lund University
P. O. Box 7082
220 07 Lund
Sweden
Telephone: +46 (0)46 222 00 00
Fax: +46 (0)46 222 46 13
www.nek.lu.se

ISSN 0460-0029

Printed in Sweden by KFS i Lund AB, Lund 2009
Cover design: Lund University, 1966

erik.jonasson@europe.com

To Åsa and my parents

Contents

ACKNOWLEDGEMENTS

I would probably not be the first to compare the completion of a PhD degree with travelling a dirt road after the end of a rainy season. The journey begins with a certain sense of anxiety and not until you have reached the very final destination do you know whether you will make it. Along the road there are water-filled pot holes that you don't know whether they are shallow and harmless or two-foot deep and hazardous. There might be fallen trees over the road, and never can you rule out the risk of finding yourself standing at a rim, where a large piece of the roadway has been washed away down into a deep ravine. With good guidance and a fuel tank of inspiration and motivation the odds improve a lot of making it all the way.

For excellent guidance, I want to thank my dissertation supervisor **Sonja Opper**, who has provided the directing support that I have needed to reach the goal, with her quick, constructive, and helpful feedback on my work. At some moments along the road, I have wondered whether my guide has really got it right this time. Why take a detour over a hill when there is a flat and straight road beside the hill? Once past such hills and detours, however, I have almost always realized that my guide was right. Always alert, thoughtful, and able to see the wider horizon, Sonja has been able to foresee collapsed bridges, fallen trees, and roadblocks on the seemingly perfect road paths that I otherwise would have chosen.

For many wise and helpful words along the path, I also want to thank **Steven Helfand**, my 'friend, co-author, and pseudo advisor' at University of California, Riverside (UCR). From the first lecture of ECON265 to the re-submission of our paper, he has been a great source of guidance, inspiration, and motivation. The same gratefulness I feel to **David Fairris**, also at UCR. He invited me to his office in the fall 2002 to hear my research interests, trusted me as his research assistant two years later and as a co-author shortly thereafter. During my three-month visit to UCR in 2008, not only did he lend me his Sproul office, but also offered his generous encouragement and guidance in my work with my last dissertation paper while I was there.

I got guidance also in the form of excellent teaching during my two years of coursework at UCR, 2002–2004. For this I want to thank Taradas Bandyopadhyay, Stephen Cullenberg, Gloria Gonzalez-Rivera, Keith Griffin, J-T Guo, Aziz Khan, Tae-Hwy Lee, Victor Lippit, Prasanta Pattanaik, Robert Russell, and Aman Ullah. This coursework became even more comprehensible and stimulating thanks to many of my fellow graduate students. Thanks to Jodi Marvel for all our hard and disciplined work together that enabled me to pass the comprehensive exams after the first tough year of studies. Thanks also to Francisco, Kazuo, Mark, Glen, Xiaoyu, Weiqiang, Huiyu, Ted, Anirban, Jaehee, Leon, Heather, Juyoung, Amita, and Jean for making the Riverside years filled of joy. The financial support that enabled me to study at UCR was provided by the Jan Wallander and Tom Hedelius Foundation. Vetenskapssociteten in Lund financially supported my visit to UCR in 2008.

The Atlantic Ocean was more than a pot hole to overcome. The culture shock of going from California to Lund seemed just as big as going from Jönköping to California. The warm and friendly informal welcome to Lund by Pernilla Johansson and Therese Nilsson (who curiously knocked the Alfa 4088 office door to see who their three new colleagues were) mitigated this shock. Göte Hansson generously welcomed and supported me as dissertation supervisor during my first year in Lund. Many of my colleagues created a pleasant and productive work environment during the years in Lund. Fredrik Wilhelmsson and Maria Persson, together with Pernilla and Therese, have been the cornerstones in creating this great work environment, not the least during periods of particular frustration and sudden bumps and holes in the road. Anna A, Alia, Fredrik N G, Ola A, Mariana, Andreas B, Yves, Margareta, Fredrik G, Fredrik H, Pontus, Wolfgang, Per, Nils, Peter K, Marcus, Henrik L, Carl Hampus, Birger, Martin, Lennart, Robin, Eric, Peter S, Victor, Susanna, Mårten, and Erik W have all been great elements of this stimulating work atmosphere too.

The people who inspired me to engage in graduate studies in Economics, in the first place, were several of my friends, colleagues, and teachers at my undergraduate college, Jönköping International Business School. Without the friendship of Martin Andersson my career as an economist would probably have stopped with my undergraduate degree in 2001, if not earlier. He is the one who made Economics fun to study and to work with – be it marginal utility curves at 5 am in a study room of the Jönköping university library, or bootstrapping at 1 am, ten years later in my Paris flat. Bo Södersten trusted me to become his full-time research assistant in 2001 and later author in two of his edited books. He generously supported me in my application for graduate studies in California, and has ever since then given me confidence in the process of becoming a research economist. Scott Hacker was not only my inspiring and excellent teacher, but also generously supported me during the application procedures in 2002 for my graduate studies. Always being a role model in carefulness, thoughtfulness, and genuine helpfulness, Scott seven years later served as discussant in my final seminar, giving me detailed and helpful feedback of the three papers in this dissertation. Others in Jönköping who helped me to believe in myself as a future economist were Per-Olof Bjuggren, Abdulnasser Hatemi-J, Börje Johansson, Charlie Karlsson, and Lasse Pettersson.

The final parts of the dissertation were written in Paris. Even though this has been largely a matter of evening and weekend work, my superb colleagues at the Development Division, at the OECD Trade & Agriculture Directorate, have with their energizing kindness and cheerfulness given me the strength and motivation to be productive with my dissertation research. Special thanks to Jonathan Brooks, my closest colleague.

We usually assume in Economics that consuming sensible combinations of goods are preferable to extreme quantities of only one good. The same principle ought to apply between spending time on dissertation work and spending time on other activities. In this respect, it has been difficult to practice what I have preached. Thankfully, my good old friends outside academia – Magnus Ericsson, Christian Friis, Andreas Hilner, Martin Humble, Magnus Juvén, Solo Kirppo, Anna Rehnberg, Olof Sidén, and Karin Wallin – have helped me a lot to keep my life balanced and fun during my years as a graduate student.

I thank my family – Helena, Anita, and David – in Värmland for their completely unconditional supportiveness, despite that my visits to Bäckelid and Nordhult have tended to be infrequent and far too short throughout my years in Lund.

Åsa, lastly: along the road I also met you! You are the one who brightens my day every day and creates that pleasant feeling of harmony, warmth, and completeness in my life.

Paris, April 2009

Erik Jonasson

Chapter 1

Introduction

Today's low- and middle-income countries are following economic development trajectories that have several similarities to what Western high-income countries experienced in the 19^{th} and 20^{th} centuries. While the transformation from agriculture-based to service-based economies took almost two centuries for many Western countries, the countries that undergo this transformation today might be able to take several shortcuts. These could consist of adopting technology or learning from policy experiences from already industrialized countries. In other respects they might find themselves taking detours through civil war, corrupt government, or inefficient spending of public resources.

A central part of economic development is the process by which people gradually leave low-productivity jobs to engage in higher-productivity employment. This process involves migration of labor, mobility of people from low-value added to high-value added industrial sectors, as well as fundamental changes to the way labor markets function. Given the close relationship between employment and income, the smoothness and speed of this process have direct implications for the living standards of people and the poverty prevalence in a country.

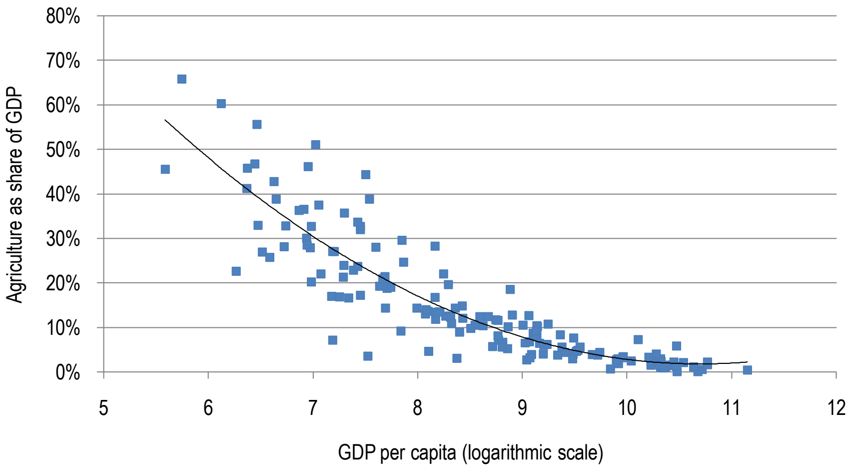
The three independent studies in this dissertation concern two elements of particular relevance for labor markets in developing countries: non-agricultural employment in the rural economy and informal employment in the urban economy. This introductory chapter discusses the context and rationale of these studies and ends with an overview of the main findings of each study.

1.1. ECONOMIC DEVELOPMENT ALONG THREE DIMENSIONS OF TRANSFORMATION

Irrespective of which shortcuts and detours countries take along their economic development path, there are three dimensions of transformation that they tend to undergo: *sectoral*, *spatial*, and *institutional*. Several other dimensions that characterize economic development may be thought of, but these three capture important empirical regularities of economic development that are of particular relevance for the analysis in this dissertation.

In similar fashion to the economic-historical experience of today's high-income countries, developing countries go through *sectoral transformation* from agriculture-dominated to manufacturing and service-dominated economies. This pattern is evident by looking at cross-country correlations of the share of agriculture in GDP and GDP per capita. Figure 1.1 provides a scatter plot of these two indicators for 180 countries in 2005. For countries with a GDP per capita of 2 000 US dollars or less (approximately 7.5 on the logarithmic scale), it is still not uncommon for agriculture to constitute 30 percent or more of the economy. As per capita income rises above 10 000 dollars, practically no country has an agricultural sector that accounts for more

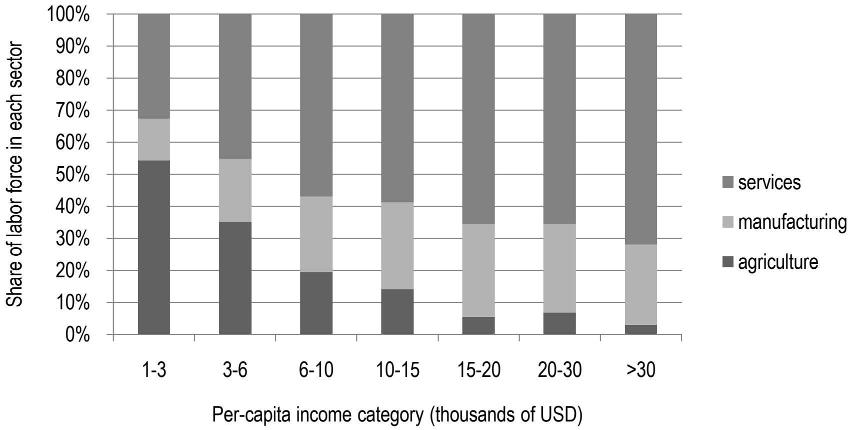
Figure 1.1. *Share of agriculture in GDP and per-capita GDP*



Note: 180 countries; GDP per capita refers to 2005 PPP USD. *Source*: World Development Indicators, 2009.

than 10 percent of GDP. The composition of the labor force largely mirrors this transformation. Figure 1.2 shows the average shares of employment in agriculture, manufacturing, and services for 120 countries, divided into seven income categories. On average, half of the labor force is occupied in agriculture in the poorest countries. This share quickly falls with income, and for countries that have a 15 000-dollar or higher per-capita income, the service sector generally occupies two-thirds or more of the labor force, manufacturing most of the remainder, and agriculture just a few percent.

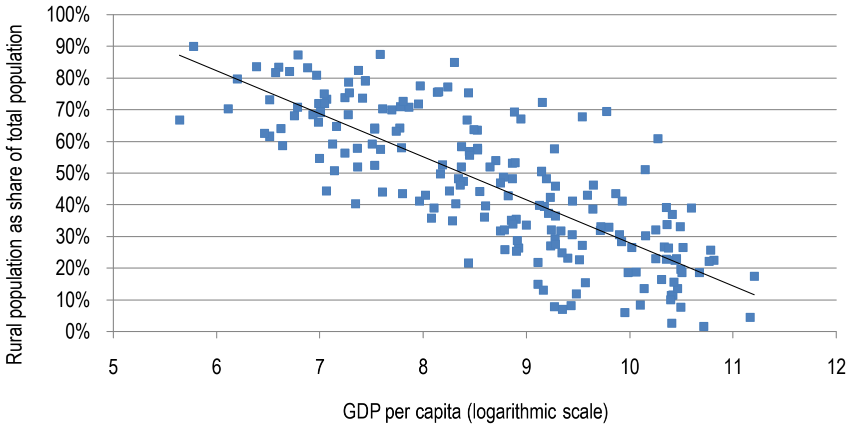
Figure1.2. *Employment shares in agriculture, manufacturing, and services in countries of different income level*



Note: 120 countries; income categories are based on GDP per capita 2005, PPP USD.
Source: Author's calculations based on World Development Indicators, 2009.

Along the development path, countries also undergo *spatial transformation* from rural to predominantly urbanized economies. While the transformation along this dimension is not as uniform as in the case of sectoral transformation, Figure 1.3 shows that a majority of the countries with less than 5 000 dollars (approximately 8.5 on the logarithmic scale) have more than 50 percent of their population in rural areas. On average, this share declines to 25 percent when countries reach an income of 20 000 dollars. Urbanization – the process in which the share of people living in urban areas increases – may occur both as a

Figure1.3. *Share of population that is rural and GDP per capita*



Note: 190 countries; GDP per capita refers to 2005 PPP USD. *Source*: World Development Indicators, 2009.

result of higher birth rates in urban areas compared to rural areas and as a result of rural-to-urban migration.[1] In China a majority (about 56 percent) of the population is still rural, but rapid migration from rural areas might soon change this situation. In 1983 the cumulative number of rural migrants was about 2 million in China. This number had increased to about 78 million in the year 2000. Six years later, in 2006, the estimated cumulative number of rural migrants was 132 million (OECD, 2009). In India, which has the largest rural population in the World (approximately 800 million), it is estimated that rural-to-urban migration accounts for about 30 percent of the urbanization (Mitra and Murayama, 2008). In Brazil, rapid migration from rural areas increased the share of the population in urban areas from 15 percent in 1940 to 56 percent in 1970, and to more than 80 percent in the year 2000 (Wagner and Ward, 1980; Brazilian Demographic Census 2000).

The third dimension of change that developing countries tend to go through is *institutional transformation.* This is understood here as the

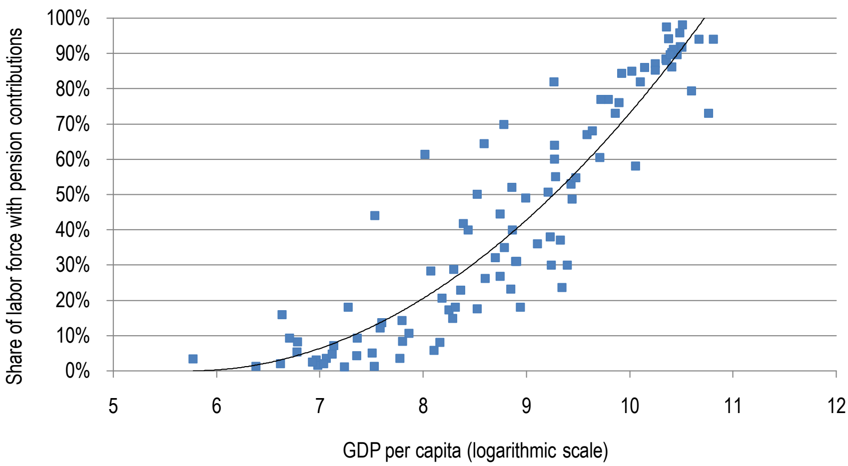
[1] Rural areas may grow "urban" if they reach the population threshold that defines an urban area. Thus, the rate of urbanization depends to a certain degree on how urban and rural areas are defined. Usually, population agglomerations of 5 000 people constitute the lower threshold for what is officially defined as an urban area (Haggblade et al., 2007).

transformation from an economy based largely on informal rules and procedures to one based on a framework of formal legislation – in short, the transformation from *informal* to *formal* institutions. Institutions are the 'rules of the game' that shape and guide human behavior (North, 1990). The distinction between formal and informal lies largely in the enforcement mechanism. While formal institutions are usually enforced by official entities (such as police, bureaucrats, and courts), informal institutions are socially sanctioned norms of behavior that rely primarily on self-enforcement mechanisms of obligation, expectations of reciprocity, and internalized norm adherence (de Soysa and Jütting, 2008). In the absence of formal rules that effectively regulate employment, property ownership, or land use, various types of informal rules and procedures are usually applied instead. In an agrarian economy, sharecropping as a means to overcome moral hazard situations in the farmer-laborer relation and "squatter's rights", which regulate access to land, are two such examples.

The institutional transformation analyzed in this dissertation is the formalization of the labor market. This refers to the process where jobs in the "informal" sector are replaced by jobs in the "formal" sector. The defining characteristic of jobs in the informal sector is usually that they are outside the scope of labor regulation and do not make the worker eligible for social security benefits. They are usually perceived as the jobs at the bottom of the wage distribution, associated with the poorest working conditions (Jütting and de Laiglesia, 2009; Perry et al., 2007). Along with growing average incomes, the labor market (as well as other markets) tends to become more formalized. This transformation is illustrated in Figure 1.4, which shows a strong positive correlation between the share of the labor force employed in the formal sector and per-capita income for 100 countries. The indicator used here to represent the formalization of the labor market is the share of the labor force that makes pension contributions. In low-income countries this share is 20 percent or lower, while in countries with an income of 20 000 dollars or more the share is usually 85 percent or larger.

This illustration of economic transformation – from agrarian to service-dominated, from rural to urban, and from informal to formal – does not imply that these phenomena are necessary sources of economic growth. The relationship between economic growth and these changes may probably best be understood as one of circular causation. In terms of urbanization, for

Figure1.4. *Formalization of the labor market and GDP per capita*



Note: 100 countries; GDP per capita refers to 2005 PPP USD. *Source*: World Development Indicators, 2009.

example, the opportunities to earn a higher income motivate people to migrate to urban areas. With increased spatial concentration of people and economic activity, in turn, follow increased agglomeration economies, which have been identified as an important determinant of economic growth (World Bank, 2009; Fujita et al., 1999). Thus, although the cross-country correlations outlined above are results of complex and mutually reinforcing processes, they give insight into the trajectories that today's low-income countries will be likely to follow as they develop and grow richer.

1.2. FROM FARMER TO FORMAL: EVOLUTION OF EMPLOYMENT FORMS IN DEVELOPING COUNTRIES

Economic and political actors face a broad range of challenges during this three-dimensional transformation. People generally strive to improve their standard of living, whether they are poor or not, and poverty and well-being can be conceptualized in more dimensions than income (Sen, 1999). Still, income tends to be closely related to a person's standard of living even if measured in broader terms of capabilities or freedoms from constraints. A basic assumption in much of economic theory is that individuals respond to incentives to make themselves better off (Easterly, 2002). This assumption, which even underlies Adam Smith's notion of the invisible hand, has also been the basis of several theories of economic transformation, migration, and labor market outcomes. Lewis (1954), for example, assumed in his model that a source of expansion of the modern manufacturing sector was the pool of underemployed agricultural labor that would be willing to undertake non-agricultural employment for practically any wage rate above zero. Harris and Todaro (1970) assumed that rural-to-urban migration would occur until the point where expected income in the urban sector was equal to rural income. In labor market theories of compensating wage differentials, workers are assumed to weigh costs and benefits attached to different jobs that are available and to choose the employment outcome that gives the highest expected utility (Rosen, 1986).

The studies in this dissertation are concerned with some of the employment choices that people make during this transformation process. A simple chart of employment forms in developing countries can help to set the scene and place these employment outcomes in a broader context. Three dichotomies, based on the dimensions of transformation discussed above, are used to categorize the different employment forms. The first dichotomy is the distinction between rural and urban areas. Although this is a strong simplification of the economic geography (which is also sensitive to definition), it helps us to generalize the spatial allocation of people and economic activities.[2] Urban areas are characterized by a certain population density, which usually allows for larger and more specialized product markets, greater degree of agglomeration economies, and lower transaction costs than in rural areas. These have far-reaching implications for the structure of labor markets in urban and rural areas. The second dichotomy is the distinction between the agricultural and non-agricultural sectors in the rural economy. The rural economy in developing countries is generally characterized, somewhat misleadingly, as being largely agricultural. Such a characterization overlooks

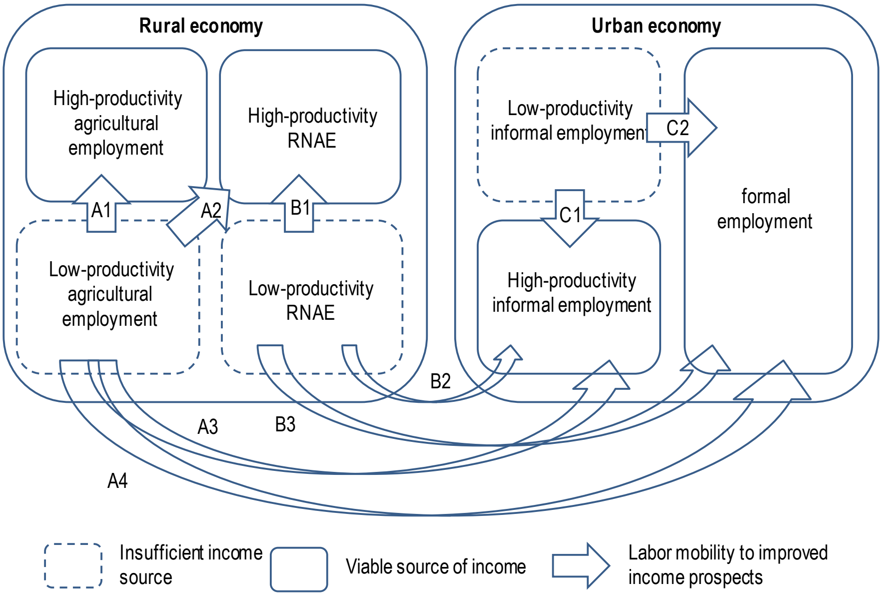
[2] The rural/urban dimension may also be considered as a continuum of agglomerations, ranging from large-city metropolitan areas to medium-sized cities, small cities, rural towns, villages, and dispersed rural settlements. The World Bank (2009) proposes an agglomeration index to distinguish between rural and urban areas, based on population density and access/proximity to a major population center.

the fact that a sizeable share of rural income stems from non-agricultural sources, primarily non-agricultural work. It is estimated that rural non-agricultural employment (RNAE) generates 40 to 60 percent of rural household incomes in Latin America, Asia, and Africa (Davis et al., 2009). In an opening remark of their book *Transforming the Rural Nonfarm Economy*, Haggblade and his co-authors note that the RNA sector has grown "too large to ignore" (p. 3). The third dichotomy relates to the formalization of the urban labor market and distinguishes between formal and informal employment.

Figure 1.5 represents employment forms based on these three ways of dividing the labor force. It depicts the rural economy as consisting of the agricultural and non-agricultural sectors, and the urban economy as being divided into the formal and informal sectors. Earnings potential is added as a fourth dimension in the figure: the agricultural, rural non-agricultural, and urban informal sectors are divided into low-productivity and high-productivity employment. The threshold between the two can be thought of as the poverty line or some other level of income that separates people with a decent standard of living from those below this standard. Thus, having high-productivity employment here means having a job that enables the worker to secure a certain standard of living, but it does not say anything about which industrial sector this job is in or what kind of production technology is involved.

The arrows suggest the main directions of labor mobility that will enable improved income prospects during the path of economic development. In the least developed countries (which are also the least urbanized), the majority of the poor are occupied in low-productivity agriculture as farmers or wage laborers. A major challenge for these countries is to find viable income sources for their still growing rural populations. The potential for agriculture to absorb more labor is limited by a fixed amount of land and by technology improvements that tend to make agriculture less, rather than more, labor intensive (World Bank, 2007). Poverty alleviation strategies usually point at three potential pathways out of poverty for these people. One is to identify viable farmers and support them to increase their productivity by adopting improved technology or altering their crop mix (pathway A1). A second pathway out of poverty is to remain in the rural economy but engage in high-productivity non-agricultural activities (A2). A third pathway is migration to urban areas, where income prospects might be better than in agriculture, at least in the long run (A3, A4 – even though this path might go through low-productivity informal employment in the short run).

Figure 1.5. *Employment transition and pathways out of poverty*



Source: The author.

Earnings in the rural non-agricultural sector are on average higher than in agriculture, yet there are poor people in this sector as well. Landlessness prevents these poor from engaging in own farming, while the agricultural labor market might be too thin during parts of the year to offer a stable and sufficient labor income. Lacking other immediate options, these people are "pushed" into low-productivity non-agricultural activities (Reardon et al., 2001). To escape poverty they need access to productive assets to gain increased returns in the RNA sector (B1) or might, just like poor farmers, benefit most from migrating to urban regions, where there exists a larger demand for non-agricultural labor (B2, B3).

Low-income countries that have reached a high degree of urbanization – like many of the Latin American countries – still tend to have high poverty rates in rural areas to take care of. In absolute numbers, however, the majority of the poor in these countries live in urban areas. The elements of poverty alleviation strategies for the urban poor differ from strategies aimed at the rural poor. Access to land is not an issue here, nor is unfavorable location the main binding constraint. Alleviating urban poverty is a mixture of allowing the poor to invest in their human capital (by improved health and education), creating an inclusive institutional framework that is designed to meet the needs of the poor, and ensuring that public services and infrastructure reach the poor so they can benefit from the agglomeration economies that the urbanized economy can offer (de Soto, 2001; Smith, 2005; World Bank, 2009). This will enhance the opportunities of people in low-productivity informal employment to gradually engage in higher-productivity activities (C1), and eventually be integrated with the formal economy (C2).

The studies in this dissertation analyze pieces of this framework of employment transition. The underlying question of Chapters 2 and 3 is to what extent the rural non-agricultural sector is a viable pathway out of poverty. Chapter 4 seeks to identify the key factors that provide access to formal-sector jobs in the urban economy. A brief overview of each chapter is provided below.

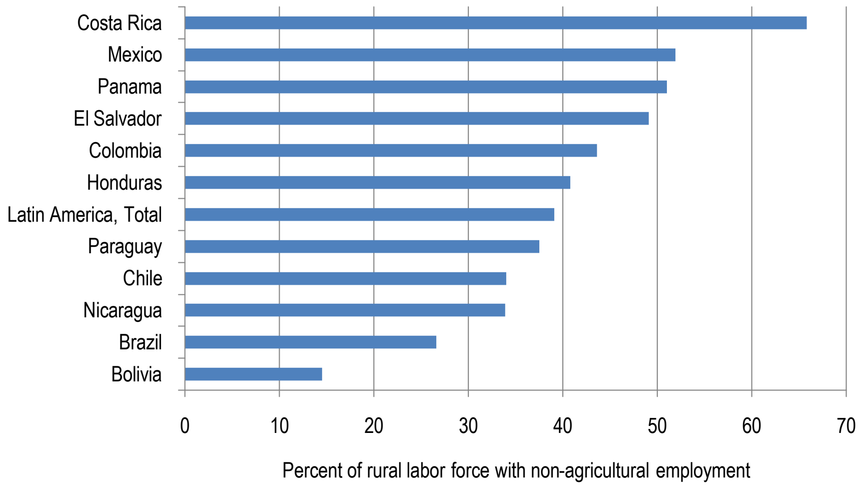
1.3. SUMMARIES OF THE STUDIES

How important is economic geography for rural non-agricultural employment? (Chapter 2)

As a residual concept, the rural non-agricultural sector contains a wide variety of activities, ranging from agro-food processing to public-school teaching. By definition, it includes everything done in the rural economy, except agriculture (cultivation, livestock production, and forestry). The RNA sector has attracted a considerable amount of interest during the last two decades, not only because of its relative importance for rural income (Figure 1.6), but also for its potential to serve as a stepping stone out of rural poverty. Some recent empirical evidence even suggests that countries that have had a sizeable RNA sector and secondary towns – as a middle step in their transition to urbanized economies – have been more successful in alleviating poverty than countries that have had a rapid urbanization, without this middle step (Christiaensen and Todo, 2008).
What determines rural non-agricultural employment (RNAE) opportunities, and to what extent is RNAE able to reduce poverty and improve living standards for rural households? The general hypothesis posed in this chapter is that RNAE opportunities are determined jointly by individual and

household characteristics (supply-side effects), labor market characteristics (demand-side effects), and by the transaction costs of participating in markets. Previous empirical studies concerned with determinants of RNAE have tended to focus mainly on the supply-side factors. The principal contribution of the chapter is to analyze, in greater detail than in previous studies, the role of demand-side effects and transactions cost, which we refer to as the "economic geography".

Figure 1.6. *Rural non-agricultural employment in Latin America*



Source: Dirven (2004), Table 3.

By utilizing data from the Brazilian Demographic Census of year 2000, we are able to test for the role of economic factors, such as local market size and distance to population centers, at a disaggregated level. Our empirical results show that the economic geography can explain a considerable share of the variation in RNAE. Failure to control for such factors may exaggerate the importance of personal characteristics, such as education, in the prediction of a person's employment outcome. Our findings on the relationship between geographical factors and non-agricultural income are less conclusive. Remoteness from markets, for example, does not appear to have the same negative effect on earnings as on opportunities to find employment in the non-agricultural sector.

The implications of the results are mixed regarding RNAE as a potential pathway out of rural poverty. The average income is indeed higher among people who are engaged in RNAE, but, to gain access to high-productivity RNAE, the empirical results suggest that both personal characteristics and the economic geography matter. Thus, for RNAE to serve as a poverty exit path, it seems contingent on favorable location, within a certain proximity to markets, as well as on investment in human capital for the worker.

Earnings differentials in the rural labor market (Chapter 3)

It is commonly observed that rural non-agricultural employment on average provides higher incomes than agricultural work in developing countries (World Bank, 2007; Lanjouw, 2007). This signals a potential of the RNA sector, not only to provide an alternative means of employment where agricultural labor markets are thin, but to substantially increase the income prospects of rural households. This potential, however, largely depends on whether the observed differences in labor compensation between the sectors can be explained by worker characteristics (such as skill), or whether significant differences in earnings remain after holding such characteristics constant. Chapter 2 analyzes factors that affect the employment and income opportunities in the RNA sector. The empirical question raised in Chapter 3 is: When such factors are controlled for, are the income prospects significantly better in non-agriculture than in agriculture?

A basic farm household model is used to place this issue into a theoretical context. The empirical study is based on the Peruvian Living Standard Measurement Study (LSMS) of 1994. There is little support in the empirical results for the notion that an unskilled worker would earn a higher income in RNAE than in agriculture. The results do suggest, however, that returns to education are higher in RNAE and hence that skilled people tend to do better in RNAE than in agriculture, which is consistent with the predictions of the theoretical model.

The policy implications of the results are in line with those of Chapter 2. It is unlikely that RNAE will offer better income prospects unless the worker has the skills that allow high-income RNA activities. While the role of the economic geography is not evaluated in any great detail in this study, it is likely to play the same important role in Peru as in Brazil. Encouraging the rural labor force to exit farming for rural non-agricultural activities might offer little income improvement in the short run. Viable RNAE opportunities are likely to come as a by-product of rural development strategies that aim at strengthening the human capital of the rural population as well as creating viable market places.

Regional variation in informal employment (Chapter 4)

In Chapter 4 the focus shifts from the rural to the urban economy. As described above, a sizeable share of the urban labor force in developing countries tends to be employed in the informal sector. Brazil is no exception in this respect. According to the Demographic Census of 2000, about 44 percent of the adult urban labor force has some form of informal employment as their principal occupation. A comparison across municipalities also reveals great regional variation in the degree of informal employment, ranging from below 20 percent of the urban labor force in some municipalities to 80 percent or more in others.

There are studies that explain differences in informality on a cross-country level and point to the role of the burden of business regulation, labor regulation, taxes, governance, corruption, and institutional quality (Schneider and Enste, 2000). There are also micro-level studies that explain formal or informal employment as an outcome of worker characteristics, such as education, gender, and household position (Funkhouser, 1996). This study adds to a small branch of studies that lies between these two approaches – those analyzing informality at the regional, within-country level. The question is: When country-specific factors that affect informality are held constant, what sources are left *within* the country to cause regional variation in informality? This is the topic of Chapter 4, in which a theoretical model is developed and then empirically evaluated.

The theoretical model identifies worker characteristics as well as institutional characteristics as sources of regional variation in informal employment. On the worker side, skill endowment is assumed to be the key factor that determines which sector a worker participates in. On the institutional side, the local government is assumed to play a role by its level of *government effectiveness* in its political decision process, bureaucratic functioning, and public goods provision. It can also affect the expected labor income by labor taxation in the formal sector and enforcement of tax and labor. The role of informal institutions is taken into account in an extended version of the model. While formal rules and regulations, and their enforcement, affect the incentives faced by the individual, it could also be that informal rules – social norms – influence the incentives to work in one sector or the other. In particular, if there is a strong social norm to obey tax and labor regulations, there could be a large non-pecuniary cost of violating those regulations. This potential "moral cost" of informal employment is taken into account in the model.

The empirical results are largely consistent with the theoretical model: informality is more prevalent where average education is lower, where governance is less effective, and possibly where social norms are weaker regarding tax compliance. On the one hand, the results complement empirical cross-country studies by holding the national formal-institutional framework constant and emphasizing the role of governance and policy implementation at the local level. On the other hand, they complement the micro-level studies by showing that local characteristics affect the individual's sector of employment, when controlling for a series of worker characteristics.

Policies that aim to transform informal jobs into formal jobs are likely to contain certain trade-offs. Getting rid of informal jobs, by increased enforcement of labor regulations, may cause increased unemployment and poverty in the short run. It could be that the informal sector is an important, and almost inevitable, "middle step" in the transformation from a rural to an urban economy – just as the rural non-agricultural sector is referred to by Christiaensen and Todo (2008). If we want to believe the implications of the theoretical model and the empirical results, then the degree of formalization will increase as workers' perceived returns to formal sector participation increase. Increasing the access and expected returns to work in the formal sector requires, to some extent, the same actions as those for fighting urban poverty discussed above: letting the poor invest in their human capital and creating an inclusive institutional framework that supports the poor in central issues such as property ownership, access to the financial system, and public-goods provision.

REFERENCES

Christiaensen, L. & Todo, Y. (2008). Poverty Reduction during the Rural-Urban Transformation: The Role of the Missing Middle. Mimeo. World Bank.

Davis, B., Winters, P., Reardon, T., & Stamoulis, K. (2009). Rural Nonfarm Employment and Earnings: Household-Level Linkages. *Agricultural Economics, 40* (2), 119–123.

de Soto, H. (2001). *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else.* London: Black Swan.

de Soysa, I. & Jütting, J. (2007). Informal Institutions and Development: How They Matter and What Makes Them Change. In J. Jütting, D. Drechsler, S. Bartsch, & I. de Soysa (Eds.), *Informal Institutions: How Social Norms Help or Hinder Development* (pp. 29–44). Paris: OECD Development Centre.

Dirven, M. (2004). Rural Non-farm Employment and Rural Diversity in Latin America. *CEPAL Review*, *83* (August), 47–65.

Easterly, W. (2002). *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*. Cambridge, MA: The MIT Press.

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

Funkhouser, E. (1996). The Urban Informal Sector in Latin America: Household Survey Evidence. *World Development, 24* (11), 1737–1751.

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, J. R. & Todaro, M. P. (1970). Migration, Unemployment and Development: A Two-Sector Analysis. *American Economic Review, 60* (1), 126–142.

Jütting, J. & de Laiglesia, J. R. (2009). Employment, Poverty Reduction and Development: What's New? In J. Jütting & J. R. de Laiglesia (Eds.), *Is Informal Normal?: Towards More and Better Jobs in Developing Countries* (pp. 17–26). Paris: OECD Development Centre.

Lanjouw, P. (2007). Does the Rural Nonfarm Economy Contribute to Poverty Reduction? In S. Haggblade, P. Hazell, & T. Reardon (Eds.), *Transforming the Rural Nonfarm Economy: Opportunities and Threats in the Developing World* (pp. 55–79). Baltimore: The John Hopkins University Press.

Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labour. *The Machester School, 22* (2), 139–191.

Mitra, A. & Murayama, M. (2008). Rural to Urban Migration: A District Level Analysis for India. IDE Discussion Paper No. 137. Institute of Developing Economies.

North, D. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge, MA: Cambridge University Press.

OECD (2009). *OECD Rural Policy Reviews: China*. Paris: OECD.

Perry, G., Maloney, W., Arias, O., Fajnzylber, P., & Mason, A. (2007). *Informality: Exit and Exclusion.* Washington, D.C.: World Bank.

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

Rosen, S. (1986). The Theory of Equalizing Differences. In O. Ashenfelter & R. Layard (Eds.), *Handbook of Labor Economics, Volume 1* (pp. 641–692). Amsterdam: Elsevier Science.

Schneider, F. & Enste, D. (2000). Shadow Economies: Size, Causes, and Consequences. *Journal of Economic Literature*, *38* (1), 77–114.

Sen. A. (1999). *Development as Freedom*. New York: Knopf.

Smith, S. C. (2005). *Ending Global Poverty: A Guide to What Works*. New York: Palgrave MacMillan.

Wagner, F. E. & Ward, J. O. (1980). Urbanization and Migration in Brazil. *American Journal of Economics and Sociology, 39* (3), 249–259.

Worldbank (2007). *World Development Report 2008: Agriculture for Development*. Washington, D.C.: World Bank.

Worldbank (2009). *World Development Report 2009: Reshaping Economic Geography*. Washington, D.C.: World Bank.

Acknowledgements: I am grateful to Jonathan Brooks, Steven Helfand, Aziz Khan, and Sonja Opper, who have provided helpful comments on this chapter.

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.

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

	Agriculture			*Non-agriculture*
	Cultivation	Animal rearing	Forestry	
Region				
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
Rural sub-category				
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
Employment status				
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
Gender				
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*

	Region						Employment		Gender	
	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
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' 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.

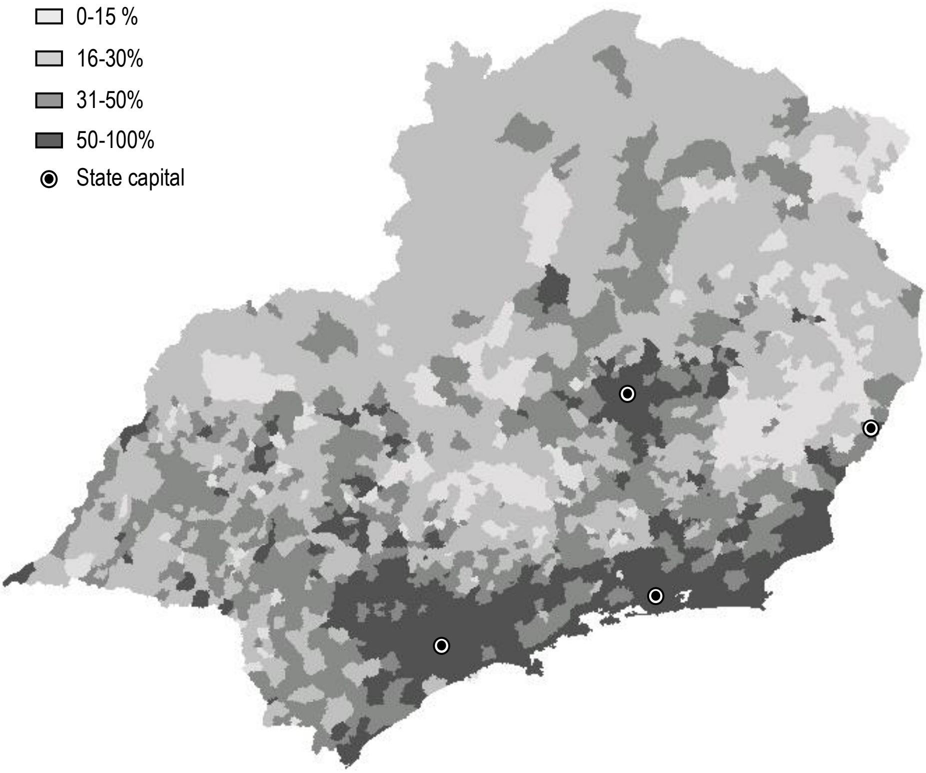
Even though average earnings in most of the RNA sectors are higher than in agriculture, there are also many low-paid non-agricultural jobs. We divide individuals with RNAE into two groups depending on earnings relative to agriculture. If an individual is engaged in RNAE and has earnings below the average 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
Agriculture	*280*	*198*	*346*	*296*	*170*	*n/a*

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

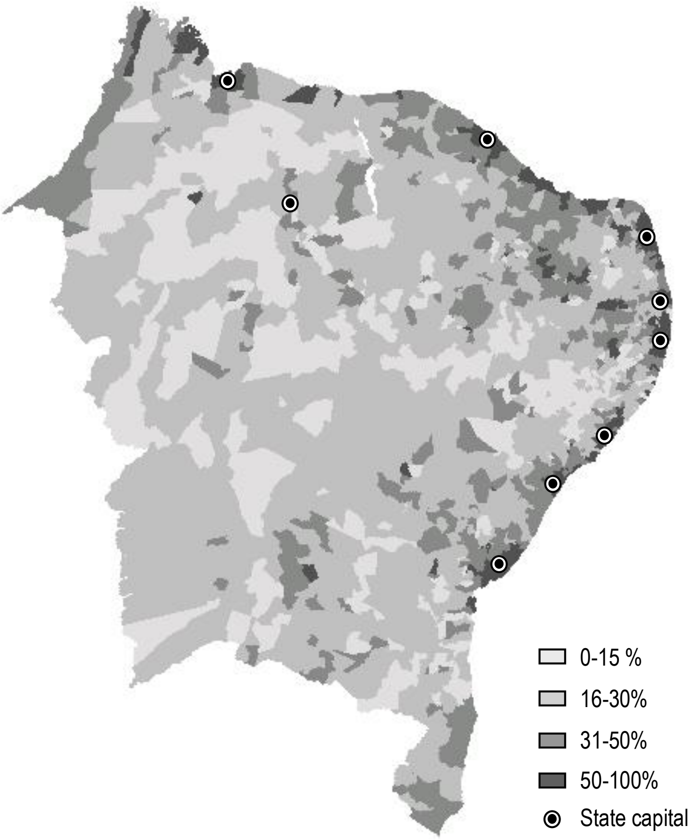
FIGURE 2.1A. *Rural non-agricultural employment in the Brazilian Southeast*



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:

$$P_i = prob\left(RNAE_i = 1 \middle| X_{ijk}, H_{jk}, M_k\right) = prob(v_i \geq 0) \qquad (1)$$

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:

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

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. The individual's probability of engaging in RNAE is estimated as:

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

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:

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

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:

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

$Income_l$ 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:

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

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.

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)

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* $\geq$ 0.95, *Telephones* = 0, *Telephones* $\geq$ 0.4, or *Electrification* $\geq$ 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:

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

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."

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 Microeconometric 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., Berdegué, 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

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

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.

Chapter 3

Earnings Differentials in the Rural Labor Market: Does Non-agricultural Employment Pay Better?

3.1. INTRODUCTION

Even though agricultural development has traditionally been the main ingredient in rural development strategies, scholars have for long emphasized the need for diversified approaches to fighting rural poverty in order to take the heterogeneity of the rural population into account. The message is that efforts to improve agricultural productivity should be concentrated to viable farm households, and that alternative paths out of poverty should be stimulated for landless or non-viable farm households. These alternative paths could be migration for some and participation in the rural non-agricultural (RNA) sector for others.

An attractive feature of rural non-agricultural employment (RNAE) is that it may provide a source of income for some of the rural landless and for those who cannot secure their income from agricultural wage labor. It also constitutes a source of complementary income for farm households. Diversifying into non-agricultural activities could be a response to insufficient farm income or a means to decrease the vulnerability associated with volatile agricultural incomes. Although migration to urban areas might be the most appropriate route out of poverty for some groups, RNAE may also have the potential to slow down rural-to-urban migration and the process of rural poor merely becoming urban poor (Lanjouw and Lanjouw, 2001). For the most vulnerable rural poor, poverty alleviation will require assistance through social transfers (de Janvry and Sadoulet, 2000; Echeverría, 2000). The need for heterogeneous efforts to alleviate rural poverty in Latin America is well represented in recent strategy formulations (World Bank, 2003 and 2007; de Ferranti et al., 2005).

Even though average earnings in the RNA sector are higher than in agriculture, it is unclear whether income prospects are systematically better in non-agricultural activities than in agriculture (Lanjouw, 2007). In particular, is there a systematic earnings differential between RNAE and agricultural work, when controlling for other factors that are likely to determine the earnings potential of an individual? Whether such an earnings differential exists is relevant, from a policy and strategic point of view, to determine whether RNAE should be included as a general element of rural development strategies or be promoted under certain conditions only. This chapter adds to the RNAE literature by empirically testing for such a sectoral earnings differential between agriculture and non-agriculture. A household model with a dualistic rural labor market is introduced to guide the empirical analysis.

There are few studies that explore earnings differentials between agricultural employment and RNAE. The empirical literature on RNAE focuses mainly on the determinants of participation in the RNA sector, with that participation considered as being either an occupational choice of the individual (Ferreira and Lanjouw, 2001; Lanjouw, 2001) or part of a household income diversification strategy (Barrett et al., 2001; Reardon et al., 2000). Studies are also concerned with the determinants of the income of those who participate in the RNA sector (de Janvry and Sadoulet, 2001; Isgut, 2004).

A reason that few studies examine income differentials, or wage gaps, between agricultural and non-agricultural employment is the empirical challenge of isolating the sector effect from unobserved factors that influence income and sector choice simultaneously. Dabalen et al. (2004) estimate returns to participation in the RNA sector relative to the agricultural sector in Rwanda. They use the method of propensity score matching to test whether people with similar attributes, but in different sectors, earn different incomes, and find that the self-employed in the non-agricultural sector earn significantly more than farm workers. McCulloch et al. (2007), in a study on pathways out of rural poverty in Indonesia, use panel data to trace the income changes of people switching from agriculture to non-agricultural activities. They find that increased engagement of rural farmers in non-agricultural businesses has been the most promising path out of rural poverty. It is unclear, however, to what extent these findings may be generalized to the context in Latin-America, which differs from South-East Asia and Sub-Saharan Africa by its higher level of per-capita income, its much lower population density, and its high degree of wealth and income inequality.

To broaden the empirical evidence, this study shifts focus to Peru as one of the poorest Latin-American economies. The analysis is undertaken using the 1994 Peruvian *Encuesta Nacional de Hogares Sobre Medición de Vida*, which is the survey source for the World Bank Living Standard Measurement Study (LSMS) for Peru that year. Ordinary least squares income regressions, in which sector of employment is treated as an exogenous choice, serve as the basis for the empirical approach. This OLS approach is complemented with an instrumental-variable approach to adjust for bias of OLS coefficient estimates, which could arise due to the potential endogeneity of sector choice. There is little support in the results for the notion that an unskilled worker would earn a higher income in RNAE than in agriculture. Still, the results do suggest that returns to education are higher in RNAE and hence that skilled people tend to do better in RNAE than in agriculture. This finding is robust across most of the regression specifications, including the instrumental-variable approach.

The next section provides an overview of rural poverty and employment in rural Peru. Section 3.3 introduces the theoretical model, followed by the empirical analysis in Sections 3.4–3.5. Section 3.6 concludes.

3.2. A PROFILE OF RURAL POVERTY AND EMPLOYMENT IN PERU

Peru, with 27 million citizens, is the fifth largest country in Latin America in terms of population. According to the demographic census of 2005, 26 percent of the population lives in rural areas, which is close to average for the region. About 50 percent of the population lives in the coastal region (*Costa*), including Lima; about 37 percent live in the highlands (*Sierra*); and the remaining 10–15 percent of the population live in the lowland jungle of the Amazon basin (*Selva*).

One of the biggest economic and social challenges for Peru is the large share of the rural population that lives in poverty. The Peruvian National Institute of Statistics and Informatics (INEI) estimates that 72.5 percent of the rural population fell below the national poverty line (defined as twice the cost of a daily food basket) in 2004. With a poverty rate similar to that of the 1980s, the evolution of rural poverty is discouraging. Although economic growth led to a poverty decline in the 1990s, poverty increased again with the economic recession at the turn of the century. Programs specifically targeting rural poverty amount to 450 million US dollars per year but have not shown positive long-term results. Escobal (2004) notes that, even though this is a large commitment of resources, the majority of the programs consist of safety nets and temporary relief, and that little is spent to overcome the structural causes of poverty. Table 3.1 shows that the poverty rate is highest in the less developed Sierra and Selva regions, where people fall lowest beneath the poverty line and to which most of the anti-poverty resources are directed.

It is evident that the path out of poverty for most rural households will have to be accompanied by continued political efforts to invest in rural infrastructure and to promote institutional change to the advantage of the poor (World Bank, 2003). The question is: What opportunities do the rural households have on their own to improve their income prospects? In particular, is it likely that the household will increase its income through diversifying its sources of income by engaging in RNAE? Agriculture is still the main sector of employment in rural Peru. According to the 1994 survey *Encuesta Nacional de Hogares Sobre Medición de Vida* (henceforth referred to as the Peruvian LSMS 1994), more than 70 percent of rural household labor was engaged in agriculture and the remaining 30 percent in non-agricultural activities. More recent statistical sources suggest that the general structure of the rural labor market was stable during the 1990s and the early 2000s.[1]

Although some 28 percent of household labor hours are spent on RNAE, the share of the rural labor force with RNAE as the principal form of employment is only 20 percent. Thus, many households have RNAE as a secondary source of earned income. The degree of diversification of income sources in rural households is generally high. If employment specialization is defined as spending 90 percent or more of labor time in one sector, then about

[1] Escobal (2001), using the LSMS of 1994 and 1997, estimates the shares of the rural labor force engaged in non-agricultural employment at 31.6 and 30.5 percent for these two years respectively. The World Bank (2005) reports that 72.8 percent of household labor hours are spent on agricultural work and 27.2 percent on non-agricultural work. For Latin America as a whole, about 40 percent of the rural labor force is involved in non-agricultural activities, which is an increase by 5–10 percentage points since the early 1990s (Dirven, 2004).

TABLE 3.1. *Poverty indicators, Peru 2004*

	Poverty (head count ratio)	Poverty gap	Extreme poverty
National	51.6	18.0	19.2
Regions:			
Urban	40.3	12.4	7.9
Rural	72.5	28.3	40.3
Urban Costa	37.1	10.6	6.2
Metropolitan Lima	36.6	10.4	3.4
Rural Costa	53.5	16.4	14.6
Sierra	67.7	27.2	36.5
Selva	59.5	19.7	26.4

Note: The national poverty line 2004 was PEN 202.5 (1 PEN = 0.30 USD, June 1, 2004). The regional poverty lines vary from 170 in rural areas to 273 in Metropolitan Lima. The national extreme poverty line was 113 and was estimated as the cost of a daily minimum food basket. *Source*: INEI (2006).

36 percent of rural households are specialized in agriculture, whereas only 5 percent are specialized in non-agricultural activities. The remaining 59 percent are pluriactive households. Table 3.2 shows that RNAE is most prevalent in the coastal region, which is the more developed region in terms of average income, infrastructure, and labor market participation. More than 37 percent of the labor force in this region is engaged in the non-agricultural sector. The survey data, however, do not tell us whether a rural resident also works in a rural area. Some non-agricultural workers are likely to be rural residents who commute to urban areas.[2] This causes some over-estimation of the size of the 'truly' rural non-agricultural labor force, particularly in the more densely populated coastal region. Self-employment, as opposed to wage labor, is the dominant form of employment in both agriculture and non-agriculture, with the exception of the non-agricultural sector in the coastal region.

The World Bank (2005) estimates that the poverty rate in Peru is 80 percent among people employed in the agricultural sector and 60 percent among people employed in the RNA sector. Wage workers in agriculture are those who are most likely to be poor, followed by farmers (Lopéz and della Maggiora, 2000). This is confirmed by the Peruvian LSMS 1994. Without controlling for other factors, there is a statistically significant difference in

[2] Urban areas are defined as all towns and cities with 2 000 or more inhabitants.

TABLE 3.2. *Rural household labor allocation by region, percent of weekly labor hours*

	Costa	Sierra	Selva	Rural, total
Agriculture	*62.7*	*74.3*	*70.2*	*71.8*
Wage labor	16.5	6.6	8.9	8.6
Self-employment	46.2	67.7	61.3	63.2
Non-agriculture	*37.3*	*25.6*	*29.8*	*28.2*
Wage labor	23.4	11.8	11.2	13.5
Self-employment	13.9	13.8	18.6	14.7

Source: Author's calculation based on the Peruvian LSMS, 1994.

average earnings between the agricultural and non-agricultural sectors, with earnings being higher in the latter. Table 3.3, which reports monthly income from principal employment, shows that this difference is driven mainly by different average wages in the wage labor market. Average earnings for the self-employed are not notably different between the two sectors.[3] Neither is there any difference in average earnings between the self-employed and wage laborers within the non-agricultural sector.

In this study, RNAE includes all activities other than agricultural work at a farm.[4] A considerable share of the RNA economy consists of activities closely related to agriculture, such as food processing, transportation, and marketing of agricultural goods. It also includes activities such as mining, construction, domestic services, and tourist-related services, with little or no connection to agriculture. Table 3.4 shows the composition of employment in the RNA sector. Commerce and manufacturing stand out as the biggest sectors in terms of employment, absorbing more than 55 percent of RNAE. The most

[3] Income estimates for the self-employed are subject to a higher degree of measurement error than estimates for wage labor. Two sources of possible measurement error are the volatility in income flows for the self-employed and the fact that earned income is not adjusted for expenditure related to these business activities.

[4] There is no consensus in the literature on whether to include auxiliary farm activities, such as fishing and hunting in RNAE. Saith (1992), for example, considers these activities as non-agricultural since they do not fall under the constraint of land use. Non-agricultural work should not be confused with *off-farm* work, which is a broader concept used to denote all work (agricultural or non-agricultural) performed outside one's own farm.

TABLE 3.3. *Earned income in the rural workforce, 1994, local currency*

	Wage laborer	Self-employed	All
Agriculture	125 (4.9)	202 (9.2)	183 (7.1)
Non-agriculture	199 (7.9)	200 (11.9)	200 (7.6)
Total	161 (4.9)	201 (7.4)	188 (5.4)

Note: Local currency units (PEN). Standard error is within parentheses. Unpaid family members are excluded. *Source:* Author's calculations based on the Peruvian LSMS, 1994.

common manufacturing activities are food processing and textile work. Among the self-employed in non-agriculture, almost 80 percent are engaged in commerce and manufacturing. Wage labor is not heavily concentrated in any particular sector, but more than 30 percent is found in education and other forms of public-sector employment. Manufacturing, domestic services, and construction are the most important private non-agricultural sectors for wage laborers.

As Peruvian rural households get wealthier, they tend to spend less labor time on agriculture and more time on non-agricultural activities. Table 3.5 shows that households in the lowest quintile spend around 20 percent of their labor on RNAE, whereas households in the highest income quintile spend

TABLE 3.4. *Rural non-agricultural employment by sector*

	Percentage of RNA labor force	Percentage of self-employed	Percentage of wage labor
Commerce	34.4	51.1	5.9
Manufacturing	22.7	28.2	15.4
Construction	7.0	5.0	12.1
Education	6.5	<1	17.6
Domestic services	5.9	1.8	13.5
Transportation	5.6	4.4	8.4
Public administration	4.4	<1	11.5
Hotels and restaurants	4.0	3.1	4.5
Fishing	3.9	3.2	1.5
Other public and private services	1.7	1.8	2.3
Social services	1.6	<1	3.6
Real estate and business services	1.1	<1	2.7
Other	1.0	<1	1.0
Non-agricultural sectors, total	100	100	100

Note: Principal occupations only. Twenty percent of the paid rural labor force has a non-agricultural principal occupation. *Source*: Author's calculations based on the Peruvian LSMS, 1994.

about 35 percent on RNAE. One can think of reasons for positive as well as negative correlations between household wealth and engagement in RNAE. Reardon et al. (2000) suggest that wealthier households are likely to possess the assets that make non-agricultural employment profitable (the necessary capital for business start up or education for well-paid employment), giving them 'pull' factors to increase their share of RNAE as their wealth increases. To the extent that poor people are 'pushed' into RNAE, as an income source of last resort and as a backup for low agricultural incomes, one would predict a negative relationship between wealth and RNAE.

TABLE 3.5. *Rural household labor allocation by income quintile (percent of weekly labor hours)*

	Lowest	Second	Third	Fourth	Highest
Agriculture	*78.4*	*75.6*	*71.6*	*68.1*	*64.6*
Wage labor	8.8	11.6	8.6	7.1	6.8
Self-employment	69.6	64.0	63.0	61.0	57.8
Non-agriculture	*21.6*	*24.4*	*28.5*	*31.9*	*35.5*
Wage labor	12.6	13.8	13.4	15.0	12.8
Self-employment	9.0	10.6	15.1	16.9	22.7

Source: Author's calculation based on the Peruvian LSMS, 1994.

According to Ellis (2000), a negative relationship between wealth and degree of multi-activity is observed in regions where land ownership distinguishes the well-off from the poor. A positive relationship tends to be observed in regions where livestock and human capital are the main assets distinguishing the better off from the poor. Reardon et al. (2000) also discuss the possibility of a U-shaped relationship between household wealth and degree of multi-activity. They explain this relationship in terms of push factors forcing the landless poor to undertake a high degree of non-agricultural employment, whereas the middle-income households are well off enough to survive on farm production or agricultural wage labor alone. High-income households are able to engage in well-paid non-agricultural activities due to high skills or asset holdings. In Peru pull incentives appear to dominate push incentives for participation in RNAE, as suggested by the increasing share of RNAE from low-income to high-income household quintiles.

3.3. A HOUSEHOLD MODEL WITH DUALISTIC LABOR MARKETS

This section proposes a basic farm household model in order to put the livelihood strategy of rural households into a theoretical framework. The model is chosen deliberately to capture some of the characteristics of rural households described in the previous section; the dominant source of employment and income for rural households in Peru is agriculture, and agricultural work most often takes the form of self-employment (peasant farming) rather than wage labor. The model extends the static farm household model developed by Bardhan and Udry (1999). The innovation is to take a dualistic feature of the rural labor market into account. In particular, the model assumes that there exists an agricultural labor market, in which labor is treated as homogenous, and a non-agricultural labor market, in which workers are compensated according to their skills.

The model assumes that households have their own farm, but that they can supply all or parts of their labor in the competitive agricultural and non-agricultural labor markets. Households differ by their level of skill, which determines their labor productivity in the non-agricultural labor market, but not in the agricultural labor market. Let δ denote a household-specific skill parameter, which is normalized to 1 for unskilled households. Provided that they supply their labor in the non-agricultural labor market, skilled households are able to supply a multiple δ of "unskilled-labor equivalents". The price of one unit of unskilled labor is equal in the two markets.

In the absence of market failures, the farm household's consumption and production decisions are separable and can be made independently of each other (Taylor and Adelman, 2003). The profit-maximizing level of output determines the amount of labor hired and the amount of capital rented. The constrained optimization problem that the household faces consists of maximizing household composite utility with respect to consumption and leisure, given its total income from farm production and off-farm labor:[5]

$$\max_{c,l} U(c,l) \tag{1}$$

subject to the following constraints:

[5] See Bardhan and Udry (1999) for the full exposition of the original model.

$$p_c c + wL_A^h + rK^h \leq p_F F(K, L_A) + w(L_A^m + \delta L_N^m) + rK^m \quad (2a)$$

$$L_A = L_A^f + L_A^h \quad (2b)$$

$$K = K^f + K^h \quad (2c)$$

$$E^L = L_A^f + L_A^m + L_N^m + l \quad (2d)$$

$$E^K = K^f + K^m \quad (2e)$$

$$c, l, L_A^f, L_A^m, L_N^m, K^f, K^m \geq 0;\ \delta \geq 1 \quad (2f)$$

In the utility function (1), c and l denote the household's composite consumption and leisure. In the budget constraint (2a) prices of consumption goods, farm production, labor, and capital are denoted p_C, p_F, w, and r, respectively. $F(K, L_A)$ is the household farm production function with capital (K) and agricultural labor (L_A) as inputs. For simplicity, capital includes productive land as well as physical equipment (non-agricultural productive capital is abstracted from). The budget constraint states that expenditures on consumption, hired labor, and capital cannot exceed the revenues from production and from marketed labor (L^m) and capital (K^m). The subscripts A and N on marketed labor distinguish labor supplied in the agricultural labor market from labor supplied in the rural non-agricultural labor market. Identities (2b) and (2c) state that the household itself provides labor and capital in farm production (f) or hires in the factor markets (h). Household labor and capital endowments are given by E^L and E^K and are allocated according to identities (2d) and (2e). The household can devote its time to four activities: work on the own farm, marketed off-farm wage labor in the agricultural or non-agricultural labor markets, and leisure. Non-negativity constraints are listed in (2f).

The separability assumption allows for profits from farm production be maximized independently of household preferences:

$$\pi^*(p_f, w, r) = \max_{L_A, K}[p_f F(K, L_A) - wL_A - rK] \quad (3)$$

The household labor allocation decision depends on preferences for leisure and on labor productivity, δ. For the unskilled household, for which δ equals one, the opportunity cost of leisure is the wage rate w. Since the marginal revenue product of labor in own farm production equals the going wage rate, the unskilled household is indifferent to supplying its labor to farm production or in the agricultural or non-agricultural labor markets. The skilled household (δ>1) faces an opportunity cost of leisure equal to δw, provided that it can supply all its labor in the non-agricultural labor market. A skilled household will therefore neither allocate its labor to its own farm nor participate in the agricultural labor market, in which returns to labor are no higher than w, even for skilled labor. Thus, for skilled households the labor allocation identity (2d) reduces to $E^L = L_N^m + l$. Re-arranging (2a) and substituting in (3), the full-income constraint for the unskilled and skilled households become, respectively:

$$p_c c + wl \leq \pi^*(p_f, w, r) + wE^L + rE^K \tag{4a}$$

and

$$p_c c + w\delta l \leq \pi^*(p_f, w, r) + w\delta E^L + rE^K \tag{4b}$$

Two implications of particular interest follow from the model. First, in the absence of market failures, no household has an incentive to diversify its income sources by participating in both labor markets. Second, provided that the two labor markets are in equilibrium, an unskilled agricultural worker will not earn a higher income by switching to non-agricultural work, as both sectors have the same competitive wage rate, w. Thus while earnings in the non-agricultural sector are higher due to higher average labor productivity, there is no "unconditional" earnings differential between the sectors once worker skills are taken into account – unconditional in the sense of not being conditional on having a certain level of education. This hypothesis is subject to empirical evaluation in the next section. Before turning to the empirical analysis, the limitations of this "naive" model are briefly elaborated on.

Under missing or imperfect markets the situation will be different, and the household problem cannot be represented by equations (1) – (3). Various forms of market failures are commonly observed in product and factor markets in the rural areas of developing countries (Stiglitz, 1998). The roots of limited market access are usually high transaction costs (de Janvry et al., 1991: Key, et al., 2000; Sadoulet et al., 1998). Incomplete information usually causes inefficiency in factor markets (Stiglitz, 1988) and poorly defined property rights often add to the problem (de Soto, 2001). If there are barriers to entering the land and capital markets, the production possibility of the household will be largely determined by its factor endowments. In the extreme case of complete factor market inaccessibility, the factor input identities (2b) and (2c) reduce to $L_A = L_A^f = (E^L - l)$ and $K = K^f$, respectively.[6] If a household is landless and cannot gain access to productive land through land rental or other arrangements, the agricultural production function is no longer part of the budget. The household is then constrained to whatever income it can earn in the labor market. Similarly, when access to financial or productive capital other than land is limited, chances to engage in non-agricultural business are limited. If the household lacks endowment of, and access to, productive land, then the budget constraint (2a) reduces to $p_c c \leq w(L_A^m + \delta L_N^m)$. Still worse, if unemployment emerges in the agricultural labor market in slack seasons, and if access to the non-agricultural labor market is obstructed by entry barriers, not even labor income will be a secure means of income for households lacking productive assets. Thus, as soon as we allow for imperfections in markets, it is no longer obvious how households will allocate their labor time.

3.4. DATA AND EMPIRICAL METHOD

Given that average earnings in the rural non-agricultural sector are higher, and that wealthier households tend to devote more of their labor time to non-agricultural activities, is the rural non-agricultural sector a potential pathway out of rural poverty? Little can be said with descriptive statistics. If skilled and highly educated labor is systematically drawn to the RNA sector in search of the highest returns to labor, then the earnings differential is just a skill-compensating wage differential. As suggested in the theoretical model, such a case would give unskilled wage labor in the agricultural sector small chances to increase income by switching to non-agricultural employment. This section empirically assesses the extent to which the sectoral income differential between agricultural and non-agricultural employment is observed when a range of other factors, which might determine the individual's earnings potential, are controlled for.

[6] In this case, household-specific shadow prices of inputs and outputs jointly determine consumption and production, i.e. separability is no longer maintained.

The empirical analysis is based on data from the Peruvian LSMS of 1994. A comparison with more recent data (ENAHO 1997 and 2002) suggests that the general structure of the rural labor market has remained largely unchanged since the survey year (see footnote 1). The survey sample is nationally representative and consists of 3 623 households, of which 1 336 resided in rural areas. The 1994 LSMS also includes a community survey, which provides information on 204 population centers where the household survey was carried out. Most of these population centers were small villages of 100 households or less. This information is used in the analysis to control for local characteristics that are likely to influence the employment outcome of rural households, yet not directly affect their income. The unit of analysis is the rural worker, and after the exclusion of unpaid family members and individuals below the age of 12, there are 1 680 individuals in the sample. The sample is a fair representation of the rural labor force, with the exception that the exclusion of unpaid employees increases the share of people in RNAE from 20 to 35 percent and the share of males in the labor force from 55 to over 70 percent. Table 3.6 provides descriptive statistics for the variables included in the analysis, and covers the sample used in the regression analysis.

A standard log-linear income equation serves as the basis for testing the presence of earnings differentials between agricultural and rural non-agricultural employment. This approach treats sector of employment as an exogenous choice. Income regressions are estimated according to the following setup:

$$y_{ij} = \beta_0 + \beta_1 RNAE_{ij} + x_{ij}\beta_x + h_j\beta_h + \varepsilon_{ij} \qquad (5)$$

where y_{ij} is the logarithm of earned income from the principal employment of individual i in household j. The variable of main interest is *RNAE*, the binary variable that distinguishes non-agricultural employment from agricultural employment. Its corresponding coefficient β_1 is the estimate of the sectoral earnings differential. Individual and household characteristics are controlled for by the vectors of variables x_{ij} and h_j, described below. The residual ε has expected zero mean and constant variance σ^2. Model (5) is estimated by OLS and is used as a benchmark for comparison with alternative model specifications. Various interaction effects are tested for and the exogeneity assumption of *RNAE* is relaxed by using a two-stage least square approach.

TABLE 3.6. *Summary statistics of regression sample*

Variable	Sample mean	Standard deviation	Definition
Dependent variable			
Earned income	217	272	individual's earned income from principal employment (log), dep. variable
Independent variables			
RNAE	0.35	0.48	has principal occupation in the non-agricultural sector (d)
Age	39.3	15.7	individual's age, in years
Male	0.71	0.45	gender, 1 for male, 0 for female (d)
Education	5.0	3.67	Individual's years of education
Non-Spanish	0.31	0.46	mother tongue other than Spanish (d)
Work migrant	0.13	0.34	individual has migrated for work (d)
Self-employed	0.67	0.47	self-employed in principal occupation (d)
Land ownership	11.5	64.3	hectares of owned land (log)
Livestock	3.22	5.29	owned livestock, cow equivalents, (log)
Equipment	418	3463	value of owned farm equipment, local currency (log)
Costa	0.26	0.44	individual lives in the coastal region (d)
Sierra	0.51	0.50	individual lives in the highlands (d)
Selva	0.23	0.42	individual lives in the jungle region (d)
Instrumental variables			
Household head	0.61	0.49	Individual is head of the household (d)
Household size	5.97	2.75	number of members in the household
Population size	117	144	number of households in nearest population center (log)
Paved road	0.46	0.50	paved or improved road in local population centre (d)
Distance to market	10.7	19.3	Distance (km) to the nearest market from local population center (log)

Note: All variables were generated from the Peruvian LSMS, 1994. The number of observations is 1 680. Dummy variables are indicated by (d), taking value 1 if true, 0 otherwise. (log) indicates that the variable enters in log form in regressions.

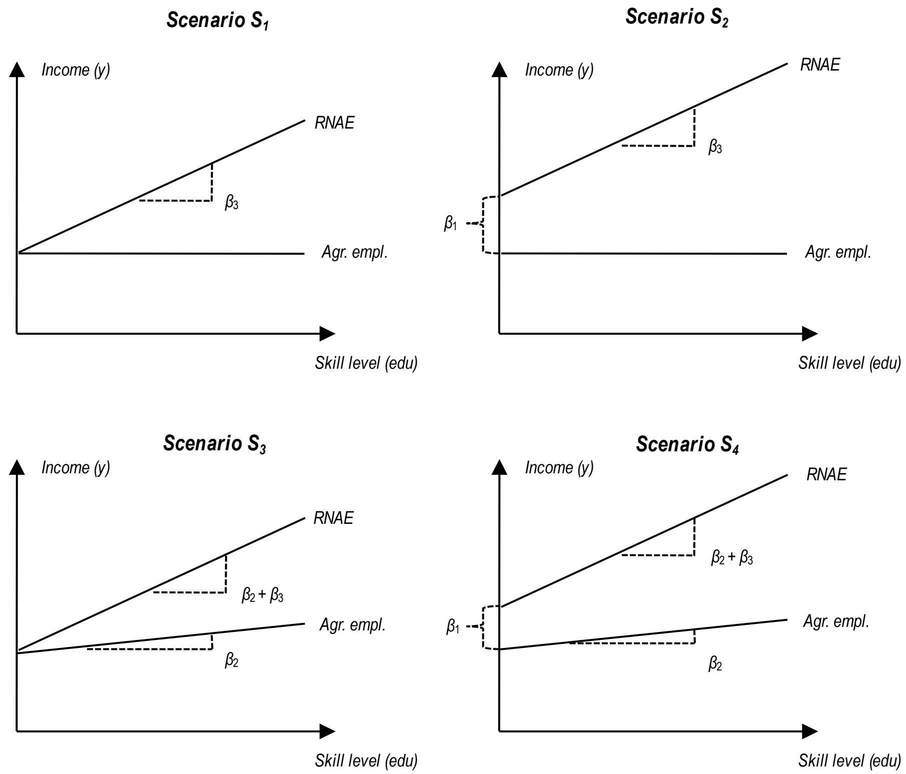
Among the individual variables in the vector x_{ij}, age, age-squared, and years of education are included as proxies for work experience and human capital acquirement. Gender is included due to the large earnings differential between men and women. This earnings gap is likely due to a combination of self-selection into low-paid employment, lower average work hours, and gender discrimination in the labor market (Abramo and Valenzuela, 2005). A dummy variable for non-Spanish mother tongue controls for ethnicity. The Quechua and Aymara languages are common among the indigenous people, who are most likely to be poor. Economic marginalization of indigenous groups is most likely a result of several factors, such as language barriers, ethnic discrimination, and adverse geographical location (Escobal, 2004). To capture mobility and the ability of an individual to respond to economic opportunities, the analysis includes a variable that controls for whether the individual has migrated for work. A binary variable separating the self-employed from laborers controls for the different conditions these two groups might face. In the labor market, earnings depend on the going wage rate, whereas for the self-employed a wider spectrum of factors will determine earnings. In combination with other productive assets, self-employment could provide higher returns to labor than wage income. Without such assets, however, self-employment could be a result of unemployment or difficulties in entering the labor market, and provide earnings lower than wage incomes. To control for regional differences in economic opportunities, regional dummy variables for the highlands and the jungle regions are included (with the coastal region being the reference region). The household variables in the vector h_j consist of three measures for productive assets: land ownership, livestock, and other owned farm equipment. Productive assets should matter little for wage labor, but for the self-employed these will constitute factors of production.

According to the theoretical model, RNAE and skills (education and work experience) are complements in generating income. This implies that the unskilled worker does not gain by switching from agriculture to RNAE, and that there are positive returns to skills in RNAE but not in agriculture. To test for interaction effects between RNAE and education, model (5) is modified as follows:

$$y_{ij} = \beta_0 + \beta_1 RNAE_{ij} + \beta_2 edu_{ij} + \beta_3\left(RNAE_{ij} \times edu_{ij}\right) + x^*_{ij}\beta^*_x + h_j\beta_4 + \varepsilon_{ij} \quad (6)$$

Education (*edu*) in equation (6) is separated from the set of individual characteristics x_{ij} (hence the asterisks on x and β_x) to make the interpretation easier. The *RNAE* indicator is interacted with the worker's years of education. There are four scenarios, in which the corresponding interaction coefficient β_3 is positive. First, the theoretical model hypothesizes that β_1 and β_2 are both zero and that β_3 is positive: education only has positive returns in RNAE, and RNAE is only beneficial given some level of education. This scenario is called S_1 and is depicted in the upper left graph of Figure 3.1. Second, there could be an unconditional earnings premium in RNAE while returns to education are positive only in RNAE. In that case, β_1 and β_3 are positive while β_2 is zero (scenario S_2). Third, it could be the case that there are positive returns to education in both sectors and that returns are higher in RNAE, but that there is no remaining (unconditional) benefit of RNAE. In this case β_1 is zero and β_2 and β_3 are positive (scenario S_3). Last, returns to education could be positive in both sectors, higher in RNAE, and there could be an unconditional earnings premium in RNAE, in which case all three coefficients are positive (scenario S_4).

Figure 3.1. *Four possible scenarios with positive interaction effects between RNAE and skills*



The test for interaction effects is then extended to allow differences in returns across sectors in *all* individual and household characteristics:

$$y_{ij} = (1 + x_{ij} + h_j)\beta^{AG} + [RNAE_{ij} \times (1 + x_{ij} + h_j)]\beta^{RNAE} + \varepsilon_{ij} \qquad (7)$$

where the β^{RNAE} coefficients constitute structural differences between agriculture and RNAE.

The analysis carried out according to models (5) – (7) assumes that sector of employment is exogenously determined. There are reasons to assume that sector of employment is a choice that is determined, at least partly, by observed and unobserved personal and household characteristics. This is shown repeatedly in empirical studies on RNAE, such as Lanjouw (1998) for Ecuador, de Janvry and Sadoulet (2001) for Mexico, and Ferreira and Lanjouw (2001) and Jonasson and Helfand (2009) on Brazil. If sector of employment, or any of the other right-hand side variables, is endogenously determined in model (5), the OLS coefficient estimates are not consistent. A two-stage least square approach is applied in order to adjust for this potential endogeneity of RNAE. Instrumental variables are needed, which are conditionally correlated with RNAE, yet uncorrelated with the error term ε, and not a direct determinant of income y. Two sets of instruments are used: two household variables and three community variables.

First, if hired labor is not a perfect substitute for household farm labor (for example, due to monitoring costs), or if there are transaction costs involved in the labor market, then the labor allocation decision may be influenced by the size of the household. Since larger households have a larger labor endowment, they might have better opportunities to let one or more household members work off farm. The number of household members, as a proxy for labor endowment, serves as the first instrument. Second, if farming involves some degree of management and monitoring that is usually under the responsibility of the household head, he or she might be less likely than other household members to take an off-farm job. An indicator for household head is used as the second instrument. Neither household size nor household position is likely to have any direct relation with unobserved characteristics that affect earnings.

Jonasson and Helfand (2009) provide some empirical evidence that the economic geography to a large extent affects the RNAE opportunities for a rural worker. They find that RNAE opportunities are higher where distance to population centers are shorter, where rural infrastructure is better, and where the local market size is larger. Their findings on the relation between geographical factors and non-agricultural income are less conclusive. Only to some lesser extent does geography seem to directly affect earnings prospects once a worker is engaged in non-agricultural work. Based on these findings, three local characteristics are used as instruments for RNAE. The first is a binary variable indicating the existence of a paved road in the community, which served as an indicator of local infrastructural development. The second variable is population size of the nearest population center, as a proxy for local market size. The third variable is a distance measure to the nearest permanent market place (which might be outside the local community).

3.5. ESTIMATION RESULTS

The estimation results of model (5) and its extensions (6) and (7) are reported in Tables 3.7–3.9 and are discussed in turn below. The primary focus is on the main variable *RNAE*, but coefficients of some of the control variables are also discussed.

Table 3.7 contains seven specifications, of which three are estimated on the full sample (columns 1–3), two on the wage labor sample (columns 4–5), and two on the self-employed sample (columns 6–7). The results suggest that there are several structural differences between laborers and self-employed, so that treating the entire rural labor force as one homogenous group could give an incomplete understanding of the relationship between income and personal characteristics.

The specification in column 1 only contains individual-specific variables. Column 2 adds household productive assets and thus corresponds to the benchmark model in equation (5). Column 3 adds interaction effects and corresponds to model (6). The reason that results are reported both with and without household assets is that these assets could be endogenous (just like employment outcome), and determined by the activity that the household is engaged in. If factor markets functioned seamlessly, this would be a major concern. Under imperfect factor markets, it is probably less endogenous to employment and income but could still be influenced by unobserved household characteristics. Adding household productive assets does not alter the coefficient estimates of the other variables in a major way, hence endogeneity of these might not be a major concern in this case.

The estimated RNAE coefficients (β_1) in columns 1and 2 are statistically significant and range between 0.41 and 0.45. This means that the estimated earnings premium of changing from the agricultural to the non-agricultural sector is approximately between 41 and 45 percent, keeping other factors constant.

In column 3, which shows the results of a regression with interaction effects, the coefficients suggest that there are positive returns to education in both sectors, higher returns to education in RNAE than in agriculture, and an unconditional earnings premium (not conditional on education) in RNAE. This is consistent with scenario S_4 discussed above. The RNAE coefficient is smaller (0.29) than in the first two columns but still significant, and the interaction coefficient is positive. It suggests that the returns to an additional year of schooling are more than twice as high in RNAE as in agriculture.

Some interesting deviations emerge when the subsample of the 552 laborers is used for the same regressions. Without interaction effects, the RNAE and education coefficients in column 4 are fairly similar to those in column 1 (with the full sample). When RNAE is interacted with education, as shown in column 5, both the RNAE and the educational coefficients become insignificant and only the interaction coefficient is significant. This is consistent with scenario S_1; that is, that there are no returns to education in agriculture, there is no unconditional RNAE premium, and the only way to gain positive returns to education is to work outside agriculture. This is what the "naive" theoretical model predicts as well. The interaction coefficient of about 0.06 suggests that the average-educated laborer will increase her income by 30 percent by moving out of agricultural wage labor into non-agricultural wage labor. For the uneducated laborer, however, the results suggest that there will be no effect on income.

The results of the regressions on the subsample of the 1 128 self-employed individuals tell a different story. The RNAE coefficient in column 6 (without interaction effects) is about 0.39, which is slightly lower than for laborers. The educational coefficient is positive and significant but about half the magnitude compared to that for laborers. In column 7, which includes the interaction effect, the RNAE coefficient is significant but both the educational and the interaction coefficients are statistically insignificant. Hence, this is not consistent with any of the four scenarios discussed in the previous sub-section.

TABLE 3.7. *Estimation results – earned income (OLS)*

	All			*Laborers*		*Self-employed*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
RNAE	0.411***	0.448***	0.288***	0.428***	0.055	0.387***	0.466***
Education	0.042***	0.039***	0.024**	0.052***	0.013	0.027**	0.025
RNAE X education			0.030**		0.063***		-0.004
Age	0.037***	0.037***	0.037***	0.064***	0.063***	0.031**	0.029**
Age squared	-0.000***	-0.000***	-0.000***	-0.001***	-0.001***	0	0
Male	0.660***	0.669***	0.677***	0.499***	0.546***	0.769***	0.786***
Non-Spanish	-0.425***	-0.420***	-0.426***	-0.116	-0.113	-0.558***	-0.552***
Work migrant	0.026	0.032	0.021	0.189**	0.137	-0.075	-0.057
Self-employed	0.064	0.051	0.073				
Sierra	-0.289***	-0.344***	-0.345***	-0.202***	-0.217***	-0.423***	-0.505***
Selva	-0.299***	-0.307***	-0.309***	-0.015	-0.002	-0.499***	-0.525***
Land ownership		0.016**	0.015**		0.012		0.019*
Livestock		0.022**	0.022**		0.008		0.031***
Equipment		-0.019**	-0.019**		-0.013		-0.023*
Constant	3.342***	3.506***	3.571***	2.837***	3.104***	3.716***	3.942***
Sample size	1 680	1 680	1 680	552	552	1 128	1 128
R-squared	0.164	0.17	0.172	0.309	0.331	0.155	0.163
F statistic	35.37	28.08	28	26.34	25.47	24.96	18.26

Note: Dependent variable is the log of earned income. Asterisks denote level of statistical significance: *** 1%, ** 5%, and * 10%.

Instead, it suggests that, for the self-employed, there is an earnings premium in the non-agricultural sector that is not conditional on education. Moreover, the results suggest that there are no returns to education in either sector for the self-employed. The insignificant coefficients for education and the interaction term, along with positive and significant coefficients for land and livestock ownership, suggest that productive assets other than human capital play the most important role for the income of the self-employed. But this conclusion should be drawn with caution. Laszlo (2005) suggests that although education may have little effect on individual earnings for the self-employed in rural Peru, there is a positive *allocative* effect of education on the income of the household. The allocative effect means that educated households are better than uneducated households at making optimal choices of income-generating activities. Once the optimal labor allocation is determined, education has little effect on labor productivity, as Laszlo argues. Yang and An (2002) find positive returns to human capital in both agricultural and non-agricultural activities for households in rural China, and they too emphasize the role of human capital in the process of allocating factors of production among activities.

Some other results are notable in Table 3.7. There is a strong gender effect in the results of the various regression specifications. For the labor force as a whole, the effect on earnings of being male is around 65 percent. The effect is more evident among the self-employed than among wage labor. Somewhat to the contrary, López and della Maggiora (2000), in their household income analysis for rural Peru, find that female-headed households are at a disadvantage among non-agricultural households but not among farming households. The results further suggest that there is a strong negative effect of having a non-Spanish mother tongue. This ethnic effect, however, differs between wage labor and the self-employed. Among the self-employed the estimated negative effect is more than 50 percent, but in the wage labor sample the ethnicity coefficient is statistically insignificant. This finding, combined with the fact that indigenous groups are under-represented in the labor market, suggests that ethnicity is more of an entry barrier to the labor market than a source of wage discrimination once in the labor market. To what extent the economic marginalization of indigenous people is due to language barriers and ethnic discrimination in the labor market (Griffin et al. 2002), or merely adverse geographic location (López and della Maggiora, 2000), is an issue for further research.

Despite structural differences in income determination between the self-employed and wage labor, there is no earnings differential observed between them in the full-sample regressions when keeping other things constant. The coefficient of the self-employment indicator is insignificant in all the full-sample specifications (see footnote 3). As for regional differences in earnings, there is a negative effect of about 30 percent of living in the mountain or jungle regions, compared to the coastal region. This effect is stronger among the self-employed but is not observed for wage workers in the lowly populated jungle region. Two of the three types of productive assets increase income for the self-employed but none of them appear to affect the income of wage laborers. The coefficient for household equipment has a negative sign, which is unexpected, but is most likely an effect of the high correlation between the household asset variables (ranging between 0.47 and 0.66). When these variables are included one at a time, the equipment coefficient is not negative.

Table 3.8 reports the estimation results of model (7), which includes RNAE interaction effects with all other variables. The lower half of the table shows the interaction coefficients. A statistically significant interaction coefficient suggests that there is a structural difference in the parameter of the respective variable between the agricultural and non-agricultural sectors. Again, the regressions are first run on the whole sample, then on laborers, and then on self-employed. Throughout all six specifications in Table 3.8, the RNAE coefficient is statistically insignificant, which partly deviates from the results reported in Table 3.7.

The differences between laborers and self-employed need emphasis. Let us therefore focus in particular on columns 4 and 6, which contain the results of the "full" model for laborers and self-employed, respectively. Column 4 suggests that, for laborers, there is no unconditional earnings premium in RNAE and there are no returns to education in agriculture. Work experience (represented by *age*) matters for agricultural wage laborers, but still has twice the returns in RNAE (0.048 + 0.049). There is no significant difference in magnitude of the gender premium for wage laborers between the two sectors. Column 6 contradicts parts of the findings about the self-employed in Table 3.7. Once all interaction effects are taken into account, there is no unconditional earnings premium for self-employed in the non-agricultural sector. Self-employed with a non-Spanish mother tongue appears to be at a much lower disadvantage in RNAE than in agriculture. There is also a strong regional effect

TABLE 3.8. *Estimation results – earned income (OLS with interaction terms)*

	All		*Laborers*		*Self-employed*	
	(1)	(2)	(3)	(4)	(5)	(6)
RNAE	-0.378	-0.63	-0.618	-0.59	-0.089	-0.369
Education	0.026**	0.023*	0.023	0.021	0.027*	0.021
Age	0.032***	0.031**	0.049***	0.048***	0.029*	0.017
Age squared	-0.000*	0	-0.000***	-0.000***	0	0
Male	0.641***	0.628***	0.556***	0.565***	0.734***	0.688***
Non-Spanish	-0.526***	-0.506***	-0.074	-0.067	-0.661***	-0.654***
Work migrant	0.021	0.031	0.169	0.168	-0.037	-0.045
Self-employed	0.143*	0.073				
Sierra	-0.339***	-0.447***	-0.393***	-0.375***	-0.389***	-0.524***
Selva	-0.443***	-0.457***	0.007	0.024	-0.671***	-0.642***
Land ownership		0.013		-0.007		0.029**
Livestock		0.047***		0.015		0.070***
Equipment		-0.019		-0.009		0.048
Interaction effects						
RNAE X education	0.029*	0.030*	0.046**	0.042**	-0.003	0.005
RNAE X age	0.03	0.033*	0.044*	0.049**	0.02	0.032
RNAE X age squared	-0.000*	-0.000**	-0.001**	-0.001**	0	0
RNAE X male	0.031	0.033	-0.009	-0.022	0.102	0.131
RNAE X non-Spanish	0.290**	0.254**	-0.107	-0.108	0.349**	0.286*
RNAE X work migrant	-0.009	-0.046	-0.122	-0.142	-0.035	-0.059
RNAE X self-empl.	-0.107	-0.063	0	0	0	0
RNAE X Sierra	0.121	0.255**	0.424***	0.395**	-0.092	0.107
RNAE X Selva	0.446***	0.512***	-0.039	0.009	0.633***	0.668***
RNAE X land own.		0.011		0.038**		-0.016
RNAE X livestock		-0.056***		-0.015		-0.089***
RNAE X equipment		0		-0.009		-0.066*
Constant	3.560***	3.847***	3.210***	3.247***	3.848***	4.083***
Sample size	1 680	1 680	552	552	1 128	1 128
R-squared	0.175	0.188	0.35	0.36	0.166	0.201
F statistic	23.57	18.77	23.31	18.07	15.11	13.11

Note: Dependent variable is the log of earned income. “RNAE X...” denotes interaction of the *RNAE* binary variable with the other respective independent variable. Asterisks denote level of statistical significance: *** 1%, ** 5%, and * 10%. Standard errors are adjusted for heteroscedasticity. A significant interaction coefficient indicates that coefficient estimates differ between agricultural and non-agricultural workers.

in the results, indicating that farm income is much lower in the jungle lowlands than in the coastal region. At the same time, non-agricultural self-employment in this region appears to be considerably more lucrative than in agriculture in either region. The results in columns 1 and 2 of Table 3.8, containing the full sample, show the aggregate outcome of these differences between laborers and self-employed.

As discussed above, a potential weakness of the income regressions outlined in models (5)–(7) is that they rely on the assumption that the sector of employment is exogenously determined. Since this is a strong assumption, a two-stage least squares approach is used to take the potential endogeneity of RNAE into consideration. Table 3.9 reports the results. The first five columns show results for the whole sample, columns 6–7 for laborers, and columns 8–9 for the self-employed. Columns 1–3 show the results of the most basic second-stage specification under three different sets of instrumental variables (the coefficients of x_{ij} and h_j in the first-stage regression are not shown). First, only the two household variables *household head* and *household size* are included. Only the latter is statistically significant. In the second case, only the set of community instruments are included – *paved road, distance to market*, and *population size* – of which two are significant. Third, all five instrumental variables are included, of which three are statistically significant. A linear combination of all the five variables is assumed to be the best instrument at hand, and is used for the other specifications in Table 3.9.

The instrumental-variable approach generates somewhat inconclusive results regarding the *RNAE* coefficient. In columns 3 the coefficient is negative and statistically significant. Adding household assets (column 4) and the interaction term (column 5) does not alter this result.[7] The interaction variable (*RNAE X education*) in column 5, which is not instrumented for, has a positive coefficient. This suggests that, while RNAE and education separately do not have any positive effect on income, education has some positive returns in RNAE.

[7] The instruments that are used for *RNAE* are less suitable instruments for the interaction term, *RNAE X education*. Since there is no obvious candidate available to serve as an instrument for this interaction term, it is treated as exogenous. This contrasts the approach used by Bertrand (2004), who faces a similar endogeneity problem of a regressor that is also part of an interaction term. She uses the same instruments for both variables.

TABLE 3.9. *Estimation results - two-stage least squares*

	All					*Laborers*		*Self-employed*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
RNAE	-1.467*	-0.866	-1.100**	-1.154**	-2.228***	0.36	-2.366	-0.810*	-1.916**
Education	0.090***	0.075***	0.081***	0.082***	-0.094**	0.054**	-0.112	0.034**	-0.071**
RNAE X education					0.319***		0.33		0.281***
Age	0.048***	0.045***	0.046***	0.047***	0.041***	0.065***	0.063***	0.025*	0.037***
Age squared	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.001***	-0.001***	0	-0.000**
Male	-0.078	0.158	0.066	0.058	0.374***	0.484***	0.586***	0.284	0.377**
Non-Spanish	-0.592***	-0.539***	-0.560***	-0.576***	-0.575***	-0.108	-0.134	-0.718***	-0.732***
Work migrant	0.077	0.06	0.067	0.064	-0.06	0.173**	0.016	-0.062	-0.065
Self-employed	-0.169	-0.095	-0.124	-0.103	0.189**	0	0	0	0
Land ownership				0.012	0.01	0.016	-0.015	0.020*	0.025**
Livestock				-0.017	-0.008	0.005	0	-0.002	-0.006
Equipment				-0.01	-0.008	-0.015	0.018	-0.041***	-0.033**
Constant	4.233***	3.948***	4.059***	4.036***	4.519***	2.903***	3.863***	4.864***	4.929***
Instruments									
Household size	0.006		0.006	0.008**	0.005*	0	0.005	0.016***	0.008**
Household head	-0.114***		-0.111***	-0.133***	-0.091***	0.072	-0.091	-0.221***	-0.123***
Paved road		0.048**	0.045**	0.006	0.008	0.013	0.008	0.007	-0.005
Distance to market		-0.006	-0.006	-0.007	-0.003	0.019*	-0.003	-0.014*	-0.003
Population size		0.04***	0.041***	0.031***	0.016***	0.046***	0.016	0.021**	0.019***
Sample size	1 680	1 680	1 680	1 680	1 680	552	552	1 128	1 128

Note: Dependent variable is the log of earned income. *RNAE* is instrumented for. Coefficients from the first-stage regression are shown for the instrumental variables only. Asterisks denote level of statistical significance: *** 1%, ** 5%, and * 10%. The regressions also included regional control variables *Selva* and *Sierra*.

The instrumental variables show less statistical significance for the subsample of laborers. Columns 6 and 7 show an insignificant RNAE coefficient, and positive or insignificant returns to education. The results for the subsample of self-employed are in line with the results for the full sample. In particular, column 9 shows negative RNAE and educational coefficients and a positive interaction coefficient.

In summary, the empirical analysis makes the following central contributions: The OLS results suggest that laborers only have positive returns to education in RNAE, and to benefit from RNAE they need some level of education. The results for the self-employed are slightly different. The results suggest that they have low or no returns to education in either sector, but that a switch of sector, out of agriculture into RNAE, is associated with increased income, given their level of productive assets. This “unconditional” premium, however, is not robust once a whole series of interaction effects are accounted for. The 2SLS results challenge the OLS results and suggest that, after adjusting for endogeneity of sector choice (RNAE), there is a negative income premium for RNAE for the uneducated but possibly a positive premium for the educated. Thus, in sum the empirical results give little support for any unconditional earnings premium in RNAE (unconditional in the sense of existing for the educated as well as the uneducated). Instead, consistent with the naive theoretical model, only the educated seem able to gain a potential RNAE income premium. Returns to education seem overall to be higher in RNAE than in agriculture, particularly for laborers.

3.6. CONCLUSION

The underlying question that motivates this study is the extent to which the rural non-agricultural sector may serve as a potential pathway out of poverty for some rural households. Over 70 percent of the rural population in Peru was estimated to live below the national poverty line in 2004. An equally large share of the rural population was engaged primarily in agriculture. This does not mean that agriculture equals poverty, but the typical household in rural Peru is nevertheless a poor farm household. Based on these characteristics, a simple farm household model is proposed to predict the earnings potential in the agricultural and non-agricultural sectors for the typical rural household. The key assumption is that there are positive returns to education in RNAE but not in agriculture. Therefore, with well-functioning markets, only skilled people will gain by working in the RNA sector and unskilled workers will have little to gain. These implications, however, are not necessarily valid as soon as market failures are allowed for.

The empirical results are somewhat mixed and might reveal some of the limitations of using the standard income regression approach on cross-sectional data to assess earnings differentials between sectors. The ideal empirical study would use panel data to trace workers over time and assess income changes for workers who change sector compared to workers who remain in the same sector. The data source is from the mid-1990s, but comparison with descriptions of more recent data on Peru reveals that there have not been any major structural changes in the rural labor force since then. Thus, the results of using the same empirical method on more recent data are unlikely to differ to any large extent from the results presented here.

While the limitations of the empirical approach should be kept in mind when interpreting them, the empirical results do not reject the basic predictions of the model. There is little support in the results for the notion that an unskilled worker would earn a higher income in RNAE than in agriculture. The results do, however, suggest that returns to education are higher in RNAE and hence that skilled people tend to do better in RNAE than in agriculture. This finding is robust across most of the regression specifications, including the instrumental-variable approach.

Before we can establish with confidence for whom RNAE is the appropriate path out of poverty and a viable livelihood strategy, we need a deeper understanding of several factors that are closely tied to RNAE. These factors include 1) the role of location for the viability of the RNA sector, 2) the importance of access to marketing channels, and 3) backward- and forward-linkages between agriculture and the RNA sector. The results in this study suggest that strategies aiming at strengthening the non-agricultural sources of income for the rural people need to contain measures for educating the rural population as well as eradicating potential entry barriers to the non-agricultural labor market.

REFERENCES

Abramo, L. & Valenzuela, M. E. (2005). Women's Labor Force Participation Rates in Latin America. *International Labor Review*, *144* (4), 369–399.

Bardhan, P. & Udry, C. (1999). *Development Microeconomics*. New York: Oxford University Press.

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

Bertrand, M. (2004). From the Invisible Handshake to the Invisible Hand? How Import Competition Changes the Employment Relationship. *Journal of Labor Economics, 22* (4), 723-765.

Dabalen, A., Paternostro, S., & Pierre, G. (2004). The Returns to Participation in the Nonfarm Sector in Rural Rwanda. World Bank Policy Research Working Paper 3462.

de Ferranti, D., Perry, G. E., Foster, W., Lederman, D., & Valdés, A. (2005). *Beyond the City: The Rural Contribution to Development*. Washington D.C.: World Bank.

de Janvry, A., Fafchamps, M., & Sadoulet, E. (1991). Peasant Household Behaviour with Missing Markets: Some Paradoxes Explained. *The Economic Journal, 101* (409), 1400–1417.

de Janvry, A. & Sadoulet, E. (2000). Rural Poverty in Latin America: Determinants and Exit Paths. *Food policy, 25* (4), 389–409.

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.

de Soto, H. (2001). *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else.* London: Black Swan.

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.* New York: Oxford University Press.

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

Escobal, J. (2004). Rural Poverty in Peru: The Role of Agriculture and Non-agriculture Sectors. Background paper to World Bank report ”Peru – Opportunities for All: Peru Poverty Assessment”. Lima: Grade.

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

Griffin, K., Khan, A. R., & Ickowitz, A. (2002). Poverty and the Distribution of Land. *Journal of Agrarian Change, 2* (3), 279–330.

INEI (2006). *Condiciones de vida en el Perú: Evolución 1997–2004*. Lima: INEI.

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.

Jonasson, E. & Helfand, S. (2009). How Important Is Economic Geography for Rural Non-agricultural Employment? Lessons from Brazil. Chapter 2 of this dissertation and Working paper 2008:3, Department of Economics, Lund University.

Key, N., Sadoulet, E., & de Janvry, A. (2000). Transactions Costs and Agricultural Household Supply Response. *American Journal of Agricultural Economics, 82* (2), 245–259.

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

Lanjouw, P. (2001). Nonfarm Employment and Poverty in Rural El Salvador. *World Development, 29* (3), 529–547.

Lanjouw, P. (2007). Does the Rural Nonfarm Economy Contribute to Poverty Reduction? In S. Haggblade, P. Hazell, & T. Reardon (Eds.), *Transforming the Rural Nonfarm Economy: Opportunities and Threats in the Developing World* (pp. 55–79). Baltimore: The John Hopkins University Press.

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

López, R. & della Maggiora, C. (2000). Rural Poverty in Peru: Stylized Facts and Analytics for Policy. In R. López & A. Valdés (Eds.), *Rural Poverty in Latin America* (pp. 281–305). New York: Saint Martin’s Press.

McCulloch, N., Weisbrod, J., & Timmer, P. (2007). Pathways out of Poverty during an Economic Crisis: An Empirical Assessment of Rural Indonesia. World Bank Policy Research Working Paper 4173.

Reardon, T., Taylor, J. E., Stamoulis, K., Lanjouw, P., & Balisacan, A. (2000). Effects of Non-Farm Employment on Rural Income Inequality in Developing Countries: An Investment Perspective. *Journal of Agricultural Economics, 51* (2), 266–288.

Sadoulet, E., de Janvry, A., & Benjamin, C. (1998). Household Behavior with Imperfect Labor Markets. *Industrial Relations*, *37* (1), 85–108.

Saith, A. (1992). *The Rural Non-farm Economy: Processes and Policies.* Geneva: International Labor Office.

Stiglitz, J. E. (1988). Economic Organization, Information, and Development. In H. Chenery & T. N. Srinivasan (Eds.), *Handbook of Development Economics, Volume 1* (pp. 93–160). Amsterdam: Elsevier Science.

Stiglitz, J. E. (1998). Markets, Market Failures, and Development. In C. K. Eicher & J. M. Staatz (Eds.), *International Agricultural Development, Third Edition* (pp. 103–112). Baltimore: The John Hopkins University Press.

Taylor, J. E. & Adelman, I. (2003). Agricultural Household Models: Genesis, Evolution, and Extensions. *Review of Economics of the Household*, *1* (1–2), 33–58.

World Bank (2003). *Reaching the Rural Poor: A Renewed Strategy for Rural Development*. Washington D.C.: World Bank.

World Bank (2005). Peru – Opportunities for All: Peru Poverty Assessment. Report No. 29825. 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: Valuable comments have been provided by Joakim Gullstrand, Göte Hansson, Pernilla Johansson, Therese Nilsson, and Sonja Opper.

APPENDIX: CORRELATION MATRIX

TABLE 3.A1. *Correlations between variables used in the empirical analysis*

	Inc.	RNAE	Edu.	Age	Male	Non-s	Work	Self.	Land	Live.	Equip	Costa	Sierra	Selva	Hh h.	Hh s.	Pop.	Pav.	Dist.
Income	1.00																		
RNAE	-0.01	1.00																	
Education	0.09	0.23	1.00																
Age	0.17	-0.22	-0.37	1.00															
Male	0.16	-0.36	0.11	0.12	1.00														
Non-Span	-0.13	-0.08	-0.15	0.05	0.00	1.00													
Work mig	0.03	-0.09	-0.09	0.10	0.13	-0.05	1.00												
Self-emp	0.13	-0.14	-0.26	0.34	-0.05	0.09	0.03	1.00											
Land	0.09	-0.12	-0.02	0.16	0.09	0.10	-0.04	0.22	1.00										
Livestock	0.06	-0.20	-0.03	0.15	0.07	0.06	-0.05	0.22	0.47	1.00									
Equip	0.08	-0.15	-0.06	0.16	0.11	0.05	0.00	0.27	0.63	0.66	1.00								
Costa	0.13	-0.02	0.09	-0.02	-0.02	-0.27	0.08	-0.17	-0.20	-0.21	-0.19	1.00							
Sierra	-0.09	0.09	-0.01	0.07	-0.03	0.36	-0.26	0.11	0.19	0.24	0.13	-0.60	1.00						
Selva	-0.02	-0.09	-0.08	-0.07	0.06	-0.15	0.22	0.04	-0.02	-0.07	0.04	-0.32	-0.57	1.00					
Hh head	0.17	-0.35	-0.16	0.48	0.54	0.09	0.20	0.27	0.08	0.05	0.06	-0.08	0.01	0.06	1.00				
Hh size	-0.08	0.07	-0.02	-0.17	-0.01	0.00	-0.07	-0.14	-0.02	0.07	0.06	-0.01	-0.01	0.03	-0.25	1.00			
Popul.	-0.04	0.13	0.09	0.01	-0.02	-0.01	-0.03	-0.02	-0.02	-0.12	-0.05	0.06	0.13	-0.21	0.02	-0.03	1.00		
Paved rd	-0.06	0.09	0.09	-0.06	-0.03	0.11	0.05	-0.09	-0.13	-0.27	-0.25	0.06	-0.06	0.01	-0.06	-0.03	0.11	1.00	
Distance	-0.05	0.00	-0.03	0.01	0.00	-0.10	0.08	0.07	0.06	0.03	0.04	-0.10	-0.01	0.11	0.03	0.04	-0.05	-0.08	1.00

Note: Sample size 1 680. See Table 3.6 for definitions.

Chapter 4

Regional Variation in Informal Employment: Skills, Norms, and Governance

4.1. INTRODUCTION

All economies, from the least developed to the most modern and wealthy, contain elements of informal, or unofficial, economic activity. It is estimated that as much as 60 percent of economic activity is unaccounted for in the official records of some countries in Africa. The size of the unofficial economy in Western Europe is believed to range from 10 percent of GDP in Switzerland to almost 30 percent in Greece. About nine percent of the economic activity in the United States is estimated to be unofficial (Schneider, 2005). These numbers are not vastly altered when defining informality as the share of the labor force that does not participate in the formal labor market. With the labor force definition, informality in Latin America ranges from 25 percent in Chile to 75 percent in Paraguay and Bolivia (Perry et al., 2007).

If a large number of market transactions in an economy take place informally, or underground, the tax base will be hollowed out, ultimately jeopardizing the standard of public goods and services (Johnson et al., 1998). For the individual firm, informality tends to imply more costly contract enforcement and limited access to financial capital, constraining its ability to grow (Loyaza, 1996). For the worker, informality to a large extent means being outside social security arrangements and being unprotected by labor regulation (Jütting et al., 2008).

Studies that seek to explain cross-country differences in the relative size of the informal economy contain mixed evidence on what causes one country to have a larger informal sector than the other. Excessive burden of taxes and social security contributions, strict regulations in the official economy, declining loyalty towards public institutions, and declining tax morale are some aspects that are frequently pointed at as driving economic activity underground (Schneider and Enste, 2000). Besides these regulatory and institutional aspects, informality is often found to decrease with the aggregate level of productivity in a country (Loayza and Rigolini, 2006).

The extent to which economic activity takes place informally varies not only across countries, but also across regions within countries. In Brazil, about 45 percent of the urban labor force works in the informal sector. Among its 5 500 municipalities, however, informality ranges between 20 and as much as 90 percent of the urban labor force. Clearly, tax burden, labor regulation, and other formal institutions that are common nationwide, are not the only factors that determine the extent to which economic activities take place informally.

This study adds to the theoretical and empirical literature on informal employment by proposing, and empirically evaluating, a model that explains regional variation in informal employment, while accommodating several micro-empirical regularities in the informal sector. The model hypothesizes that regional differences in informality are due to differences across regions in worker skill endowment and quality of local governance, including support and implementation of the formal institutional framework that governs market transactions in the formal sector. An underlying assumption is that participation in the informal sector is a choice rather than an unwanted outcome of exclusion from the formal labor market. The worker is assumed to weigh expected costs and benefits of employment in each sector and choose the sector in which the returns are expected to be highest. An extension of the model is made to allow non-pecuniary costs and benefits in the informal sector to affect the incentive structure faced by workers. One such hypothesized cost is the moral cost of evading taxes and not complying with labor market regulations.

The theoretical framework is evaluated empirically using data from the Brazilian Demographic Census for the year 2000, combined with municipal information on local governance and public sector service provision. Regional variation in informality is observed in several countries, but Brazil makes a particularly interesting study object. It has one of the strictest labor market regulations in the world (Almeida and Carneiro, 2009). At the same time, the country tends to score low in international comparisons of the perceived quality of certain institutions and in the level of public trust in politicians. In the mid-1990s the Fraser Institute gave Brazil a score of 0 (on a scale from 0 to 10) on an indicator called "Equity of Citizens under the Law and Access to a Non-Discriminatory Judiciary", compared to scores between 7 and 10 among OECD countries (Friedman et al., 2000). A more recent comparison ranks Brazil 122nd and 129th among 134 countries in terms of "public trust in politicians" and "wastefulness of government spending", respectively (World Economic Forum, 2008). This suggests that, while the formal political ambitions are high in terms of willingness to provide its citizens *Ordem e Progresso*, implementation and enforcement by local governments vary substantively across different regions. The size of Brazil in terms of area and population gives room for large differences in socioeconomic and cultural characteristics across regions, which is likely to reinforce regional differences in policy implementation.

The empirical results support the predictions of the theoretical model. In particular, human capital level, quality of local governance, and strength of social norms are all related negatively to the size of the informal sector. These results stand up to a series of robustness checks, but given the cross-sectional nature of the data, limitations of the empirical assessment still need to be acknowledged.

4.2. PREVIOUS EMPIRICAL FINDINGS

There is a fairly large empirical literature that evaluates the effects of worker and employer characteristics on the probability of participating in the informal sector. This literature generally does not test for regional or institutional effects, or for the potential role of local governance or public goods provision. Data limitations tend to prevent such analysis on a disaggregated level beyond the inclusion of regional dummy variables (Amuedo-Dorantes, 2004; El Badoui et al., 2008; Gong et al., 2004; Pisani and Pagán, 2004).[1] Since personal and firm characteristics clearly matter for sector participation, existent micro-empirical evidence constitutes important guidance in theoretical as well as empirical analysis of non-individual-specific determinants of informal employment.

---

[1] Pradhan and van Soest (1995) include the size of the local labor market and local unemployment in their empirical study on determinants of informal sector participation.

Micro-level studies generally conclude that the probability of working in the informal sector decreases with human capital endowment measured by years of education, job tenure, and experience (Funkhouser, 1996; Marcouiller et al., 1997; Amuedo-Dorantes, 2004; Gong et al., 2004; Pisani and Pagán, 2004). In terms of age, it tends to be the youngest and the oldest in the labor force who participate in the informal sector. The young mainly work as informal labor, whereas the older become self-employed (Perry et al., 2007). These findings are consistent with the "threshold view" that the worker needs a certain level of skill to benefit from participation in (or to get access to) the formal sector, as opposed to working in the informal sector. There is some evidence that position in the household, household size, and the presence of children in the household affect sector choice. This probably captures a gender effect. Women with children and married women are more likely to participate in the informal sector (Funkhouser, 1996; Amuedo-Dorantes, 2004). The reason could be that the informal sector offers a higher degree of flexibility in terms of work hours and work location (Maloney, 2004). Whether participation in the informal sector is a choice or a result of labor market segmentation and entry barriers to the formal sector is a subject of debate and empirical research. Perry et al. (2007) and Maloney (2004) provide an extensive discussion of the topic.

On the regional, within-country, level, Chaudhuri et al. (2006) analyze socioeconomic, political, and institutional factors to explain differences in the size and growth of the shadow economy across Indian states. They find some evidence that the expansion of the shadow economy is less pronounced in states where the press is free and independent (as a proxy for well-functioning democracy) and where the economy is more liberalized. Torgler and Schneider (2007) find that the shadow economy in Swiss cantons is smaller where tax morale is higher and where direct democratic participation is higher. Empirical evidence for Brazil points at the importance of labor regulation and its enforcement. Almeida and Carneiro (2009) analyze how differences in enforcement of labor regulation across regions in Brazil affect informal employment and unemployment. They find evidence that stricter law enforcement (measured as the aggregate amount of registration fines issued) in a region leads to less informal employment and to higher unemployment.

Further empirical evidence of the causes of differences in informality across space is, as far as the author is aware, exclusively on a cross-country level. Several studies find that the burden of business regulation is correlated positively with various measures of the informal sector (share of GDP – Johnson et al., 1998, and Friedman et al., 2000; share self-employed in the labor force – Loayza and Rigolini, 2006). Torgler and Schneider (2007), however, find no significant correlation between informality and the burden of labor regulation. The evidence regarding the role of tax rates is mixed. Friedman et al. (2000) find a negative correlation between taxes and informality, even when controlling for GDP per capita. They conclude that high taxes, per se, do not drive businesses underground as long as tax revenues are spent on supplying high-quality public goods and services. Instead, they argue, it is the discretion in the application of rules, and the corresponding corruption that it creates, that cause businesses to exit or avoid the formal economy. Johnson et al. (1998) reach similar conclusions. Studies that include measures for bureaucratic quality, rule of law, governance and institutional quality all find a negative relationship between these measures and the size of the informal sector (Schneider, 2005). Corruption, on the other hand, as an indicator of low institutional quality, correlates positively with the size of the informal sector in the studies above.

4.3. THEORETICAL FRAMEWORK

A static model is developed in this section to explain regional variation in informal employment. The proposed determinants of these differences are regional variations in worker skill distribution, tax rates, and government effectiveness. The latter is understood as the quality of publicly provided goods and services, including the extent to which the local government supports and implements the formal institutional framework of market transactions in the formal sector. The basic model in Section 4.3.1 is extended in Section 4.3.2 to account for non-pecuniary costs and benefits in the informal sector. The model is inspired by Loayza and Rigolini (2006), Galiani and Weinschelbaum (2007), and to a lesser extent by Rauch (1991). Sector of employment is modeled as a choice by the individual, rather than exogenously determined by segmented labor market mechanisms. There is an ongoing debate in the literature about which characterization of the informal sector is more realistic (Perry et al., 2007; Maloney, 2004). It is probably fair to claim that employment outcome is generally a choice, but within a limited opportunity set. In some cases the worker's opportunity set might be reduced to what practically becomes only one option; employment in the informal sector. The model introduced here takes this viewpoint of "choice with limited options".

4.3.1. A basic model of sector participation

The model assumes that workers differ by skill endowment s which is distributed among workers according to some density function $h(s)$ ($0 \le s \le 1$). Skill endowment determines the productivity of the worker, measured in number of homogenous labor units, $l(s)$.[2] A worker can devote labor units to work in the formal sector (FS) or the informal sector (IS), but not to both. Utility is assumed to increase monotonically in income, so that the worker maximizes utility by choosing to work in the sector that gives the highest expected income. The decision to work is pre-determined outside the model, hence the consumption/leisure decision is abstracted from. The utility maximization problem reduces to:

$$\max_j \; U(y^j), \quad j = \{FS, IS\} \qquad (1)$$

where y^j is expected labor income in sector j.

The formal and informal sectors differ from each other in several respects, which affect expected income and hence the incentives faced by the worker in the choice of sector. First, the institutional arrangements in the formal economy – whose quality and potential role depend on local government effectiveness – allow for higher labor productivity than in the informal sector. Second, the expected income in the formal sector is affected by the tax rate as well as the probability of finding a job. Third, in the informal sector there is a risk of income loss due to the risk of being charged with tax evasion. These factors are introduced below.

Labor productivity is characterized by positive but diminishing returns to skill in both sectors. For the informal sector, labor productivity, measured in units of homogenous labor, is given by:

$$l^{IS}(s) = s^{\alpha} \qquad (0 < \alpha < 1) \qquad (2)$$

[2] See Galiani and Weinschelbaum (2007) for a similar treatment of labor endowment.

For the formal sector, labor productivity for any level of skill is affected by the quality of the institutional framework that governs market transactions and business practices in the formal sector. If g denotes the effectiveness by which the government is able to maintain this institutional framework, then labor productivity in the formal sector is given by:

$$l^{FS}(s,g) = (1+g)s^{\alpha} \qquad (-1 < g < 1) \tag{3}$$

The model allows for different interpretations of the manner in which local governance affects labor productivity. A broad interpretation is that, with formal institutions that are well implemented and enforced by the government, firms face better prospects for benefiting from economies of scale through cheaper access to capital and better ability to enforce complex contracts (de Paula and Scheinkman, 2007; Straub, 2005). As a consequence, production in the formal sector will be more capital intensive on average, hence increasing the productivity of labor. A narrower interpretation is that local governments can supply public services and support institutions that directly affect labor productivity – independently of production technology – in an "efficiency-wage" manner. It could be that the worker is more productive in the formal sector due to an increased sense of transparency, security, and stability in the work environment of the formal sector, compared to the informal sector.

The formal and the informal sectors are assumed to be integrated in the sense that one unit of homogenous labor, l, is paid the same in both sectors. For simplicity, assume that the worker faces a perfectly elastic labor demand at a price equal to one.[3]

For workers in the informal sector there is a risk of being charged by the authorities for working informally (or for one's employer acting informally). In case of detection, labor income will be zero. The risk of being charged, c, is determined by the strength of the enforcement of tax and labor

[3] Self-employed workers do not receive labor income from an employer. These workers get paid for their production of goods and services. For simplicity, assume that production among self-employed is given by $q = l^{IS}(s)$ and that the price of output equals one.

regulation e ($0 < c < 1$; $e > 0$; $c'(e) > 0$)[4]. Expected income in the informal sector is given by:

$$y^{IS} = [1 - c(e)]s^{\alpha} \tag{4}$$

Expected income in the formal sector is affected by the worker's probability of finding a job in that sector and thereby gaining the corresponding productivity enhancement. The worker's perceived probability of obtaining a job is an increasing function of skills. As a simplifying assumption, let the probability be given by:

$$\pi(s) = s^{1-\alpha} \tag{5}$$

Thus, expected gross income in the formal sector is given by the product of (3) and (5). Imposing labor income tax rate t ($0 < t < 1$), the expected net income in the formal sector is:

$$y^{FS} = (1 - t)(1 + g)s \tag{6}$$

The threshold level of skill that equates the expected incomes y^{IS} and y^{FS} in the two sectors is:

$$\underline{s} = \left[\frac{(1-t)(1+g)}{(1-c)}\right]^{1/(\alpha-1)} \tag{7}$$

Thus, whenever the skill level falls short of $\underline{s}$ the individual does not expect to benefit from participating in the formal sector. Given that workers choose the sector rationally (and do not attach utility or disutility to any of the sectors beyond labor income) $y^{IS}(\underline{s}) = y^{FS}(\underline{s})$ denote the highest expected earnings in the informal sector and the lowest in the formal sector, respectively, for given levels of t, g and c.

The existence of a formal sector requires that $\underline{s}$ is less than 1. Thus, taxes, institutions, and enforcement must be such that, at least for some workers, earnings are higher in the formal sector. The condition for an interior solution of (7) is:

---

[4] The risk of being caught, c, is treated here as homogenous among workers. It could very well be that the authorities' level of enforcement is different from some workers to others, depending on ethnicity or other personal characteristics (Bigsten, et al., 2004).

$$\gamma \equiv \frac{(1-t)(1+g)}{(1-c)} > 1 \tag{8}$$

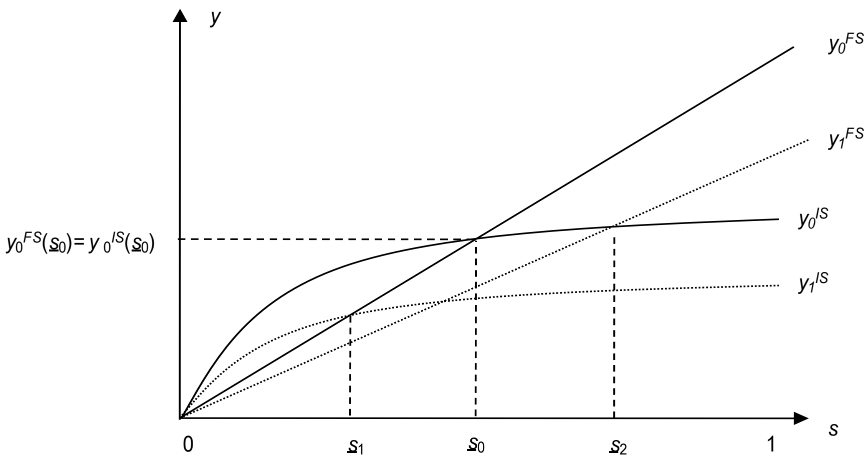
The γ denotes the ratio of formal to informal income for the highest educated (for whom $s = 1$). If this condition is not satisfied, then the tax rate is too high, governance and institutional quality too low, or the risk of income loss in the informal sector too low for anyone to benefit from working in the formal sector.

The higher the skill threshold, the higher the probability that any given worker will prefer to work in the informal sector. Differentiating (7) with respect to *t, g*, and *c* shows that the propensity to work informally increases with *t* and decreases with *g* and *c*:

$$\frac{\partial \underline{s}}{\partial t} = \frac{\gamma^{\frac{1}{\alpha-1}}}{(1-\alpha)(1-t)} > 0;\ \frac{\partial \underline{s}}{\partial g} = \frac{\gamma^{\frac{1}{\alpha-1}}}{(\alpha-1)(1+g)} < 0;\ \frac{\partial \underline{s}}{\partial c} = \frac{\gamma^{1/(\alpha-1)}}{(\alpha-1)(1-c)} < 0$$

Figure 4.1 provides an illustration of these effects. For every interior solution of (7), there is a skill threshold $\underline{s}_0$ below which workers expect to earn more by being in the informal sector than in the formal sector. This is depicted by the segment of the y_0^{IS} curve that lies above y_0^{FS}. The effect of an increase in law enforcement, which increases *c*, is shown by the lowering of the informal-sector income curve from y_0^{IS} to y_1^{IS}, which decreases the skill threshold from $\underline{s}_0$ to $\underline{s}_1$.

FIGURE 4.1. *The skill threshold that equates expected earnings in the formal and informal sector*



On the other hand, an increase in the tax rate or a decline in government effectiveness decreases the slope of the formal-sector income curve from y_0^{FS} to y_1^{FS}, which increases the skill threshold from $\underline{s}_0$ to $\underline{s}_2$.

The share of the labor force that works in the informal sector in region m is given by:

$$IS_m = H_m\left[\,\underline{s}_m(g_m, t_m, e_m)\right] = \int_0^{\underline{s}_m} h_m(s)ds \qquad (9)$$

where $H_m(s)$ is the cumulative density function of skill endowments s in regional m. Hence, differences across regions in the relative size of the informal sector are modeled as regional differences in quality institutions g, tax rate t, risk of income loss in the informal sector c, but also as differences in the skill distribution in the local labor force, $h_m(s)$. For any given levels of t, g, and c, informality will be higher (lower) in regions with lower (higher) average level of skill.

4.3.2. Model extension: non-pecuniary costs and benefits in the informal sector

The model outlined above explains why workers in the informal sector tend to have lower skills than those in the formal sector. The higher earnings in the formal sector stem from this skill difference in combination with a productivity-enhancing institutional framework in the formal sector. The model does not predict that there could be an overlap between the sectors in either skills, earnings, or both, i.e. that some workers in the informal sector are more skilled and earn more than some workers in the formal sector. Such an overlap has been found to be extensive in several empirical studies (Pradhan and van Soest, 1995; Funkhouser, 1996; Marcouiller et al., 1997; Bosch et al., 2007) and is evident also in the case of Brazil. One explanation for such a sectoral overlap is the existence of worker-specific non-pecuniary benefits and costs of participation in the informal sector. Non-pecuniary benefits of working in the informal sector could be a higher degree of flexibility in working hours or a greater sense of freedom (Maloney, 2004). A cost could be the moral cost experienced by violating tax or labor legislation (Torgler and Schneider, 2007), as well as a sense of insecurity about one's livelihood, in terms of future earnings, employment contract renewal, or enterprise survival (Jütting et al., 2008).

Let $b(x, n)$ denote the net value of the non-pecuniary benefits and costs for the worker of being in the informal sector ($-1 < b < 1$). Assume that b is distributed among workers independently of the skill distribution $h(s)$. The vector x consists of individual-specific characteristics and n represents the strength of the local social norm of obeying formal rules and regulations ($b'(n) < 0$). Assume that b is proportional to income such that the sum of benefits in the informal sector is:

$$\hat{y}^{IS} = [1 + b(x,n)]y^{IS} = [1 + b(x,n)] \times [1 - c(e)]s^{\alpha} \tag{10}$$

The skill threshold for the worker, which equates total benefits in the informal sector with income in the formal sector, $\hat{s}$, is given by:

$$\hat{s} = \left[\frac{(1-t)(1+g)}{(1-c)(1+b)}\right]^{1/(\alpha-1)} \tag{11}$$

Analogous to condition (8), an interior solution to (11) requires that at least the highest educated worker values income in the formal sector higher than the total benefits in the informal sector:

$$\delta \equiv \frac{(1-t)(1+g)}{(1-c)(1+b)} > 1 \tag{12}$$

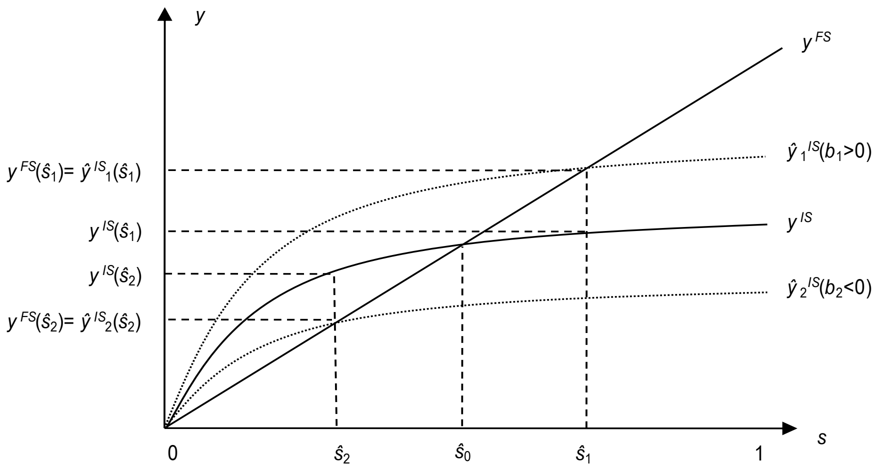
The qualitative effects of changes in t, c, and g are the same as in the case without non-pecuniary benefits. The additional effect to note is the positive effect on $\hat{s}$ of an increase in b. The derivatives of (11) are given by:

$$\frac{\partial \hat{s}}{\partial t} = \frac{\delta^{1/(\alpha-1)}}{(1-\alpha)(1-t)} > 0 \qquad \frac{\partial \hat{s}}{\partial g} = \frac{\delta^{1/(\alpha-1)}}{(\alpha-1)(1+g)} < 0$$

$$\frac{\partial \hat{s}}{\partial c} = \frac{\delta^{1/(\alpha-1)}}{(\alpha-1)(1-c)} < 0 \qquad \frac{\partial \hat{s}}{\partial b} = \frac{\delta^{1/(\alpha-1)}}{(1-\alpha)(1+b)} > 0$$

Since $\partial\hat{s}/\partial b$ is positive and $\partial b/\partial n$ is negative, then $\partial\hat{s}/\partial n$ must be negative. Figure 4.2 illustrates situations in which the worker experiences zero, positive, and negative net benefits in the informal sector. For given levels of t, c, and g, let $\hat{s}_0$ be the skill threshold in the case when the net benefit b is zero. If a worker values the net benefit by $b_1 > 0$, then the total benefit in the informal sector is given by the $\hat{y}^{IS}_1$ curve, which lies above the y^{IS} curve. The $\hat{y}^{IS}_1$ curve intersects the formal sector income curve, y^{FS}, at $\hat{s}_1 > \hat{s}_0$. The worker's monetary income in the informal sector at $\hat{s}_1$, $y^{IS}(\hat{s}_1)$, is lower than the income she expects to earn in the formal sector with the same skill level, $y^{FS}(\hat{s}_1)$. Due to the non-pecuniary benefit, however, she is indifferent between the two sectors at this skill threshold. If, instead, a worker experiences a negative net benefit $b_2 < 0$ of participating in the informal sector, then the total benefit curve falls below y^{IS} and the skill threshold decreases from $\hat{s}_0$ to $\hat{s}_2$. Despite the fact that $y^{IS}(\hat{s}_2)$ is higher than $y^{FS}(\hat{s}_2)$, the worker is indifferent between the two sectors at this point, due to the disutility attached to work in the informal sector.

FIGURE 4.2. *Skill thresholds in the presence of non-pecuniary effects in the informal sector*



Given that b is individual specific, there will be an entire distribution of thresholds $\hat{s}$ within each region. Let $s^* = s - \hat{s}$, so that whenever s^* is below 0, the individual works in the informal sector. If the distribution of s^* is given by $k(s^*)$, the size of the informal sector in region m is given by:

$$IS_m\,[k_m(s^*)] = \int_{-1}^{0} k_m(s^*)ds^* \tag{13}$$

Embedded in $k_m(s^*)$ are all the effects of taxation, governance, regulatory enforcement, and the regional distributions of skills (s) and perceived non-pecuniary benefits (b). The implication of this extension of the model is that the size of the informal sector in a region might also depend on the importance of non-pecuniary costs and benefits in the informal sectors. Assume, for example, that the two regions m_1 and m_2 have identical skill distributions and the same levels of t, g, and c, but that m_1 has a weaker norm of tax compliance and law obedience than m_2. Then the model predicts that, due to the lower moral cost of being in the informal sector, the skill thresholds in m_1 are distributed with a higher mean than in m_2, making the size of the informal sector larger in m_1 than in m_2.

4.4. EMPIRICAL APPROACH

The empirical evaluation of the model is carried out at two levels of analysis. Firstly, on worker level, a binomial probability model is estimated to assess the extent to which the individual and local factors considered in the model correlate with the probability of having informal employment. Secondly, on municipal level, a model is estimated to assess the extent to which the exogenous variables under consideration also correlate with the municipal size of the informal sector. The methods of the two approaches are discussed below.

The binomial probability model is estimated with a probit model. The difference s^* between the worker's skill level (s) and her participation threshold ($\hat{s}$) enters the probability model as an unobserved latent variable. By assumption, worker i participates in the informal sector if and only if s_i^* is below zero. Thus, if s_i^* is determined by the set of exogenous variables under consideration, then the probability that the individual participates in the informal sector is given by:

$$p_i \equiv prob(is_i = 1) = prob(s_i^* < 0 | s_i,\ x_i,\ r_m,\ z_m) \qquad (14)$$

The binary variable is_i takes the value 1 if the individual works in the informal sector and 0 if in the formal sector. s_i is a set of proxies for worker skills; x_i is a vector of other individual characteristics that are assumed to affect productivity and the individual-specific valuation of the non-pecuniary net benefit in the informal sector; r_m is a set of proxies for government effectiveness and the strength of the social norm in the municipality; and z_m consists of local productivity shifters that might affect the relative productivity in the formal and informal sectors.

The probability model is estimated under the assumption that s_i^* is a linear function of the exogenous variables:

$$s_i^* = s_i\beta_s + x_i\beta_x + r_m\beta_r + z_m\beta_z + u_i \tag{15}$$

where the β's denote vectors of parameters to be estimated and u_i denotes a residual that is assumed to be normally distributed with zero mean and variance σ^2. Let F_u denote the cumulative distribution function of u. The probit model then gives:

$$\hat{p}_i = 1 - F_u(s_i\hat{\beta}_s + x_i\hat{\beta}_x + r_m\hat{\beta}_r + z_m\hat{\beta}_z) \tag{16}$$

where $\hat{p}_i$ denotes the estimation of p_i and the $\hat{\beta}$s are the estimated coefficients ($0 \leq \hat{p} \leq 1$).

On the municipal-level, the relationship between the size of the informal sector and the exogenous variables is estimated using a fractional logit model, proposed by Papke and Wooldridge (1996). The share of the labor force in the informal sector in municipality m, IS_m (a fractional variable), is modeled as a function of aggregates of worker characteristics s_m and x_m, and the same local characteristics, r_m and z_m, as those in the probability model above. The fractional logit approach has some potential advantages over other estimation methods involving a fractional dependent variable, such as the equivalent of the linear probability model or the log-odds transformation.[5] The model has

[5] The linear probability model $P = X\beta$, estimated by OLS, relies on the strong assumption of linear relationships between the independent variables and the dependent variable. It also yields predicted probabilities that can lie outside the [0, 1] interval. A commonly applied remedy is the log-odds transformation of P, which assumes that the transformation $P^* = \ln[P/(1\text{-}P)]$ is linearly related to the independent variables (Kennedy, 2008). A disadvantage of this approach is that it does not allow P to take the extreme values 0 or 1, which is a problem if a considerable share of the observations actually takes these values. Secondly, the estimated probability $E(P|X)$ cannot be recovered without additional distributional assumptions (Papke and Wooldridge, 1996). A third alternative is to assume a specific distribution of P, such as the beta distribution, which lies between 0 and 1, and estimate the model with maximum likelihood.

similarities with the regular logit model, with the difference that the dependent binomial variable is replaced with a continuous variable that lies between 0 and 1, and that the estimation is done using a quasi-maximum likelihood procedure.[6] Define X_m as the set of explanatory variables (s_m, x_m, r_m, and z_m). The expected size of the informal sector is assumed to be related with X through a logistic function as:

$$E(IS_m|X_m) = \frac{\exp(X_m\beta)}{1+\exp(X_m\beta)} \qquad (17)$$

Model (17) is estimated with the Bernoulli log likelihood function:

$$L_m(\beta) = IS_m \times \ln[E(IS_m|X_m)] + (1 - IS_m) \times \ln(1 - E(IS_m|X_m)] \qquad (18)$$

Various robustness checks have been performed for both models. These are described separately in the results section that follows.

4.5. DATA AND DESCRIPTIVE STATISTICS

The main data source used for the empirical analysis is the Brazilian Demographic Census for the year 2000. All individual-level information is derived from these data, and several of the municipal-level variables are aggregated from them as well. The publicly available sample of the Census data includes almost 20 million individual observations, which makes it representative at the level of the municipality. In total, there were 5 507 municipalities in the year 2000, with an average population of about 30 000 people. The Census provides detailed information on employment status, earned income, and a range of socioeconomic variables. To test and control for municipal effects, two data sources are used, in addition to aggregates derived from the Census: *Base de Informações Municipais* between 2000 and 2006 and *Perfil dos Municípios Brasileiros - Gestão Pública,* 2005–2006. These databases, provided by the Brazilian Institute of Geography and Statistics (IBGE), contain

Probability estimates with this approach, however, have been found not to be robust to distributional failure (ibid.).

[6] See Papke and Wooldridge (1996) for details. Verbeek (2008) provides an in-depth discussion of quasi-maximum likelihood estimation. Gould and Villarreal (2006) provide a recent application of the fractional logit model.

detailed information on the structure of the local economy, various demographic characteristics and key public sector indicators on the municipal level.

The definition of the informal sector used in this chapter is based on the employment categories defined in the Census. Informal employment is defined here as being an unregistered employee, a self-employed person, an unpaid worker, or an employer who has fewer than five employees and does not contribute to any social security institution. Unpaid workers (who constitute about 5 percent of the informal sector) are excluded from the empirical analysis and the descriptive tables. Only the urban labor force is analyzed in this study, mainly because the majority of rural residents are engaged in agriculture and do not have access to either a formal or an informal labor market to the same extent as in urban areas.[7] Appendix 1 of this chapter discusses the definition of informal employment in greater detail and Table 4.A1 provides the percentage distribution of the urban labor force in each employment category. Using this definition, 45 percent of the urban labor force in Brazil is informal. Approximately 60 percent are employees and about 40 percent are self-employed in the informal sector.

The informal sector in Brazil has some of the characteristics that are commonly observed in studies concerned with informal employment in Latin America and elsewhere. Table 4.1 provides some key indicators. First, labor incomes are on average considerably lower in the informal than in the formal sector. Average earnings per month in the informal sector are just above 400 Reais per month (about 200 US$ in year 2000), compared to 769 Reais in the formal sector. At the 20th percentile of the earnings distribution in the informal sector, earnings are 120 R$ – well below the minimum wage (151 R$) – compared to 221 R$ at the same percentile in the formal sector. It is important to note, however, that the majority of workers in the informal sector have an income *above* the minimum wage. Moreover, the fact that the earnings at the 20th percentile in the formal sector are reached in the informal sector at approximately the 50th percentile shows that there is a considerable overlap in the earnings distributions in the two sectors. When comparing the earnings of wage labor (i.e. excluding self-employed), this intersection occurs at the 60th percentile in the earnings distribution in the informal sector.

---

[7] The urban/rural dichotomy in the Demographic Census is determined on an administrative basis and not on a certain population size or density.

TABLE 4.1. *Characteristics of the urban labor force, divided by formal and informal sector*

Age (percent in each age category)

	15-24	25-34	35-44	45-54	55-65	average age
Formal sector	21	31	27	16	5	35
Informal sector	28	27	23	15	7	34

Years of education (percent in each category)

	0-2	3-5	6-8	9-12	>12	average years
Formal sector	8	21	19	36	16	8.6
Informal sector	19	32	22	22	5	6.1

Earnings, at percentile (R$ per month)

	20%	40%	60%	80%	average
Formal sector	221	330	500	970	769
Informal sector	120	167	300	500	404

Sector of employment (percent per sector)

	Commerce	Manufacturing	Construction	Domestic services	Other services
Formal sector	21	13	12	5	46
Informal sector	18	18	5	14	30

	Gender: percent female
Formal sector	40
Informal sector	40

Note: In August 2000, the exchange rate was R$1 = US$0.56. *Source*: Brazilian Demographic Census, 2000.

Second, and most likely the main reason for the relatively low incomes, education is lower in the informal sector. On average, a worker in the informal sector has 2.5 years less education than a worker in the formal sector. Similar to the distributions of income, there is an overlap in terms of education between the two sectors. For example, 27 percent of the workers in the informal sector have nine or more years of education, while almost 50 percent in the formal sector have less than nine years of education.

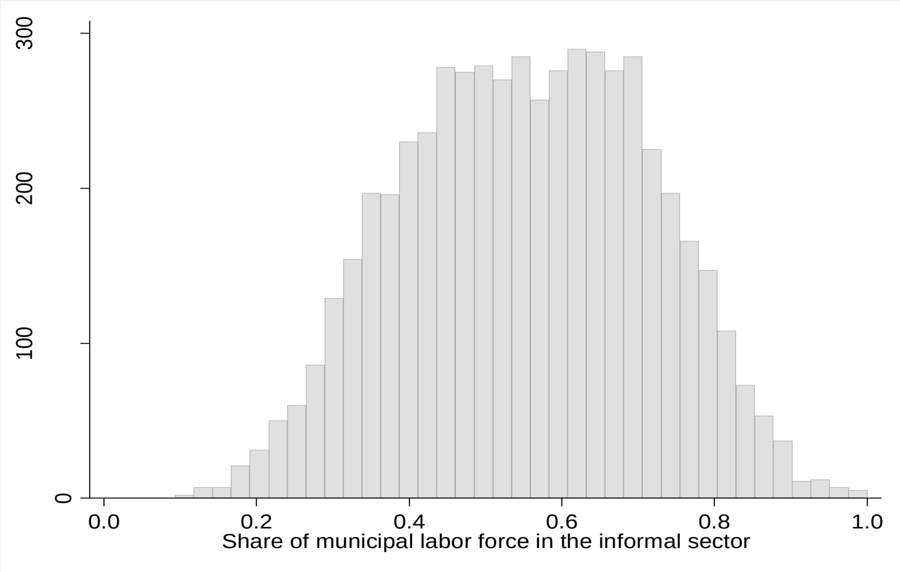
Third, the share of workers who have just entered the labor market or are close to retirement is higher in the informal sector, even though this is not as pronounced as in many other countries (Perry, et al., 2007). One aspect in which Brazil deviates from many other examples is the gender distribution across the sectors. While the informal sector in many countries tends to be over-represented by women (Kucera and Xenogiani, 2009), there is no such gender bias in Brazil. In terms of industrial composition, there is a slight bias towards manufacturing and domestic services in the informal sector compared with the composition of the formal sector.

There is a vast variation of informality in local labor markets. Figure 4.3 depicts how the relative size of the informal sectors varies across municipalities in Brazil. The unweighted average of informality on municipal level is 55 percent, but around this mean, informality varies between 20 and 90 percent. There are more than 300 municipalities with shares below 30 percent and over 1 100 with shares above 70 percent. This degree of variation in informality is not notably altered when small municipalities of 10 000 or fewer inhabitants are excluded. Even among the 200 most populated municipalities – each with 400 000 or more inhabitants – informality varies between 25 and 70 percent.

Table 4.2 gives an overview of the explanatory variables used in the empirical analysis. Individual-level variables are grouped into human capital proxies (s) and other individual characteristics (x) to be consistent with the notation in equations (14)–(17). On the municipal level, the key variables of interest (r) and are separated from the local economy control variables (z). The effect of tax rates is not assessed since the income tax is determined at federal level in Brazil and is constant across municipalities.

Human capital is represented by the individual's age and years of education. Squared age and education are included to test for non-linearity. Two additional variables are added to control for labor productivity; an

FIGURE 4.3. *Distribution of the relative size of the informal sector across municipalities in Brazil*



Note: Number of municipalities on vertical axis. *Source*: Brazilian Demographic Census 2000.

indicator for physical or mental disability, which is likely to affect work ability, and an indicator showing if the worker has recently migrated from a rural area. It could be that people who have recently entered the urban labor force from rural areas face a disadvantage in terms of knowledge of the local economy and have less access to social networks and informal institutions, which affect labor productivity (Harris and Todaro, 1970; Fields, 1975; Mazumdar, 1976).

Variables included under the assumption that they affect the utility of non-pecuniary benefits of informal-sector employment are: gender, position in the household, marital status, and indicators for the presence of formal-sector workers and young children in the household. Gender bias in household-related work and responsibilities may imply that women benefit more from the work flexibility in the informal sector than men. This effect may be strengthened if there are children in the household. The employment status by one household member may affect the work incentives by another household member. For example, Maloney (2004) notes that, in the case of Mexico, if one household member has a job in the formal sector, the entire family tends to be covered by several of the fringe benefits of the job. Thus, once one household member has a formal job, there might be little incentive for other household members to take a formal job. Instead they can work in the informal sector for cash income only. This effect is tested for, in this chapter, by variables showing the employment status and income of other household members.

Government effectiveness is tested for by using an index-variable approach. It is inspired by, yet different from, the Brazilian *IQIM* index of local institutional quality (Ministério do Planejamento, Orçamento e Gestão, 2008) as well as the *International Country Risk Guide* produced by the PRS Group. A common feature of these measures is that they are constructed as weighted sums of a range of indicators in order to obtain an index that can rank regions or countries according to quality of governance and institutions.

The composite variable *government effectiveness* (g) is the weighted sum of three indices that are constructed to capture different aspects of local governance and bureaucracy. The construction of the indices is described in detail in Appendix 2, and is summarized only briefly here. The first component is *policy formulation* (g_1), which is a proxy for the capacity of the local administration to formulate and implement policy. It is based on 20 indicators, showing to what extent the municipality has councils, development plans, ordinances, and regulations in various domains such as education, urban development, employment, and property ownership. The second component represents *bureaucratic resources* (g_2) and contains four sub-indices: i) employment form of the staff of direct administration (which share has a statutory employment contract and hence benefits from the relatively generous public-sector employment benefits), ii) competence of bureaucrats, measured as the share of the administrative staff who have secondary or higher education, iii) degree of co-ordination between units with different areas of responsibility, and iv) information technology resources. The reason for including employment form is that a bureaucrat with a generous employment contract, which offers a rich set of fringe benefits and employment security, might be more loyal and motivated to do a good job as a civil servant than if he or she has a loose and non-transparent employment contract (Evans and Rauch, 1999). The third component represents the quality of, and access to, *public goods and services* (g_3). This index is based on the teacher/pupil ratio in public primary schools, the number of health centers per municipality inhabitant, the degree of internet services offered to the public, the existence of public libraries, and the degree of support for helping people with housing. All index variables are between zero and one. The correlation coefficients of g_1, g_2, and g_3, range between 0.24 and 0.42. The effect of enforcement of tax and labor regulation (e) is not tested for explicitly since it is difficult to separate this from other qualities of the local bureaucracy (g_2).[8]

To empirically assess the possible role of a social norm influencing the moral cost of acting informally, a variable is needed that is not merely an outcome of informal employment (observed behavior) but is able to capture an attitude affecting the choice of employment. Only a survey that specifically asks for attitudes (such as the international *Latinobarómtro*) would provide a fully satisfactory indicator for this. This information is not available at such disaggregated levels as municipalities in Brazil. Instead, voting participation in municipal and presidential elections (2000 and 2002, respectively) is used as a proxy for such an attitude. The notion of rational ignorance implies that a person will only vote if the expected benefit exceeds the transaction cost of voting. But if there is a strong social norm to vote, and thereby signaling a concern for the municipality or the society as a whole, the decision not to vote could imply social sanctions and a high moral cost (Harbaugh, 1996). It is assumed here that there is some positive correlation between such a voting norm and an unobserved social norm that affects tax compliance and employment behavior.

A series of municipal control variables are included. The sectoral composition of the municipal economy is controlled for using the shares of agriculture, manufacturing, services, and public sector production in the total municipal gross product. While factors that affect informality may influence the structure of the local economy in the long run, the concern here is that the relative demand for informal labor (and products and services from the informal economy) may be higher in local economies where certain sectors dominate (such as agriculture or services). An explanation by Bosch et al. (2006) for why informality has increased in Brazil is that it is partly "due to a normal reallocation of workers to a sector that is intrinsically informal" (p. 25), by which they refer to production of non-traded goods, such as local services. Average firm size (number of employees) in the formal sector is included to

[8] For empirical analyses specifically concerned with the effects of labor regulation enforcement on employment outcomes and business performance, see Almeida and Carneiro (2005 and 2009).

TABLE 4.2. *Variables used in the empirical analysis*

Variable	Mean	S.D.	Definition
			Individual level (N = 2,222,387)
Employment and human capital			
Informal employment	0.436	0.496	Individual works in informal sector, with earned non-zero income
Age	34.5	11.6	Individual's years of age
Education	7.5	4.3	Individual's years of schooling
Other individual characteristics (x)			
Gender - female	0.397	0.489	Individual is female
Household head	0.516	0.500	Individual is household head
Married	0.454	0.498	Individual is married
Race - black	0.066	0.248	Individual is black
Disabled	0.021	0.144	Individual has reduced eyesight, hearing, paralysis, or mental problem
Rural-urban migrant	0.029	0.169	Individual has moved from a rural area 5 years or less, prior to the survey
Children in hhd	0.480	0.500	There is at least one child in the household, 10 years old or younger
Formal-sector worker in hhd	0.311	0.463	There is at least one other household member employed in the formal sector
Household income	859	2,887	Monthly total income of other household members
			Municipal level (N = 5,506)
Share informal employment	0.554	0.161	Share of the urban municipal labor force that has informal employment (see Appendix 1)
Average education in labor force	6.4	1.1	Average years of education among workers in the municipal labor force.
Municipal key variables (r)			
Government effectiveness, g	0.348	0.139	Index composed of g_1, g_2, and g_3, below, to represent government effectiveness in the municipality.
Policy formulation, g_1	0.253	0.183	Index composed of 20 variables to represent the capacity of the municipal government to formulate and implement policy. See Appendix 2 for details.
Bureaucratic resources, g_2	0.427	0.131	Index composed of g_{21}, g_{22}, g_{23}, and g_{24}, below, to represent the human, technical, and managerial resources available to the municipal bureaucracy.

Bur res, employment, g_{21}	0.620	0.290	Share of the staff of direct municipal administration with statutory employment contract, eligible for employment benefits.
Bur res, competence, g_{22}	0.204	0.124	Share of the staff of direct municipal administration with secondary or higher education.
Bur res, co-ordination, g_{23}	0.191	0.257	Index representing degree of co-ordination of activities between ten units of administration.
Bur res, IT, g_{24}	0.725	0.239	Index composed of seven variables to represent the information technology resources available to the bureaucracy.
Public goods, g_3	0.363	0.234	Index composed of 10 variables to represent the quality of, and access to, public goods in the municipality. See Appendix 2 for details.
Social norm	0.835	0.063	Index representing the strength of the social norm in the municipality that affects the moral cost of acting informally. Defined as the average share of the eligible voters who voted in the municipal and presidential elections, year 2000 and 2002, respectively.
Local economy (z)			
Share agriculture	0.280	0.183	Agriculture as share of gross municipal product.
Share manufacturing	0.190	0.162	Manufacturing as share of gross municipal product.
Share services	0.504	0.158	Services as share of gross municipal product.
Share public sector production	0.270	0.141	Public sector production share of gross municipal product.
Average firm size	5.4	5.7	Average number of employees in registered businesses in the municipality.
Local population	694	1,464	Population in surrounding municipalities, weighted by distance
Gross product per capita	4,435	5,699	Municipal gross product per capita, R$, year 2000.
Share rural immigrants	0.056	0.047	Share of the urban labor force consisting of workers who have migrated from a rural area, five years or less prior to the Census survey year.

Sources: Individual-level variables and informal employment and labor force skill on municipal level – Brazilian Demographic Census, 2000. Municipal variables for *government effectiveness, social norm, and local economy* – Gestão Pública, 2005–2006, and Base de Informações Municipais, 2000. See Appendices 1 and 2 for details.

control for technology in the formal sector, under the assumption that labor is more productive in large firms, which would increase the expected income in the formal sector for any level of worker-skill. A distance-weighted measure of local population size is included to control for market potential, assuming that the higher the market potential the higher the returns to acting formally. The share of rural-to-urban migrants in the labor force is added under the assumption that it will increase the relative supply of unskilled labor and drive down the labor income in the informal sector. Lastly, municipal product per capita is included to check to what extent the results remain robust while keeping income level constant. Since average income is most likely determined by several of the other municipal variables in the model (including government effectiveness), the results of the models that control for income should be interpreted with caution. Additional variables are used in the instrumental-variable approach and are discussed separately in section 4.6.3.

4.6. EMPIRICAL RESULTS

This section presents the results of the empirical analysis. The individual and the municipal-level models are discussed in one sub-section each. Once the empirical support has been assessed for the hypotheses of the theoretical model, some attention is given to the coefficient estimates of variables that serve primarily as control variables. The cross-sectional nature of the data prevents direct inference about causal relationships. The conditional correlations that are analyzed do not rule out the fact that causality could run in opposite directions to those being hypothesized. This limitation of the empirical analysis is discussed together with other concerns about endogeneity and omitted-variable bias in the third sub-section.

4.6.1. Individual-level probit model

Table 4.3 contains the results of the probit model specified in equations 14–16. The binary variable *is*, which indicates whether the worker is employed in the informal sector (1) or not (0), serves as the dependent variable in all specifications. To be consistent with the notation above, variables are grouped into human capital (s), other individual characteristics (x), municipal key variables (r), and municipal control variables (z). Seven specifications are reported, in which variables have been added stepwise to evaluate the validity of the hypotheses and assess the stability of the coefficient estimates as additional factors are controlled for. The coefficients show marginal effects of small changes in the independent variables or changes from zero to one for binary variables. Standard errors are adjusted to take into account clustering of the error term within municipalities. Failure to account for intra-group correlation of the error term in multi-level data analysis may lead to rejection of the null-hypothesis of a zero-value coefficient far too easily (Moulton, 1990; Primo et al., 2007).[9] Due to the large sample size (more than 2 million observations), standard errors are still small after this adjustment and most of the coefficients are statistically significant at very high levels. Specific attention should therefore be given to the economic significance when interpreting the coefficient estimates.

First, in line with the model prediction, human capital has a negative effect on the probability of working in the informal sector. Age has a negative but decreasing effect, while years of education appears to have a negative and increasing effect, as indicated by the coefficient estimates of the square terms of these variables. This relationship remains stable across the specifications, as municipal-level variables are added. Second, both of the coefficients of the key municipal variables of interest – *government effectiveness* (g) and the strength of the *social norm* (n) – are of expected sign and in most cases statistically significant.

Column 3 of Table 4.3 shows a coefficient of government effectiveness of about –0.18. When controlling for other municipal factors, the coefficient decreases to about –0.13. This means that an increase of the index value of g by one standard deviation, from the average of 0.35 to 0.49, while holding everything else constant, is associated with a two-percentage point decline in the probability of a worker being informally employed. A similar exercise for the norm coefficient suggests that a one standard deviation increase in the strength of the social norm in the municipality decreases the probability of informality by a magnitude of about six percentage points.

[9] Regression results without this standard error adjustment, not reported here, indicate significance at the 1-percent level for practically all municipal-level coefficients.

TABLE 4.3. *Estimation results – probit model (probability of informal employment)*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Human capital (s)							
Age	-0.023***	-0.027***	-0.022***	-0.022***	-0.022***	-0.022***	-0.022***
Age, squared	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
Education	-0.029***	-0.022***	-0.023***	-0.022***	-0.022***	-0.022***	-0.021***
Education, squared	-0.000***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
Other individual char's (x)							
Gender - female	0.007**	-0.001	0.008**	0.008**	0.008**	0.007**	0.007**
Household head	-0.031***	-0.057***	-0.033***	-0.033***	-0.033***	-0.033***	-0.034***
Married	-0.052***	-0.025***	-0.051***	-0.049***	-0.049***	-0.049***	-0.049***
Race - black	-0.012***	-0.006**	-0.011***	-0.010***	-0.010***	-0.009***	-0.009***
Disabled	0.052***	0.037***	0.048***	0.047***	0.047***	0.047***	0.046***
Rural-urban migrant	0.030***	0.035***	0.019***	0.011***	0.011***	0.012***	0.014***
Children in household	0.018***	0.005***	0.010***	0.009***	0.009***	0.009***	0.007***
Female X chd in hhd	0.047***	0.045***	0.048***	0.047***	0.047***	0.047***	0.048***
Household income		0.035***					
Formal-sector worker in hhd		-0.521***					
Municipal key variables (r)							
Government effectiveness, *g*			-0.178***	-0.129***			-0.014
Policy formulation, g_1					-0.030	-0.016	
Bureaucratic resources, g_2					-0.065***		
Public goods, g_3					-0.045***	-0.023*	
Bur res, employment, g_{21}						0.012	
Bur res, competence, g_{22}						-0.068**	
Bur res, co-ordination, g_{23}						-0.018*	
Bur res, IT, g_{24}						-0.115***	
Social norm, *n*			-1.247***	-0.963***	-0.952***	-0.905***	-0.668***

Local economy (z)							
Gross product per capita							-0.071***
Share agriculture				-0.112**	-0.109**	-0.129**	-0.056
Share manufacturing				-0.233***	-0.234***	-0.234***	-0.157***
Share services				-0.025	-0.024	-0.042	-0.101*
Average firm size				-0.001	-0.001	-0.001	0.001
Share rural immigrants				0.210**	0.205**	0.189**	0.213**
Local population				-0.012***	-0.012***	-0.012***	-0.008***
Sample size	2,220,387	2,220,387	2,218,167	2,213,429	2,213,429	2,213,429	2,213,429
McFadden pseudo R-squared	0.070	0.240	0.085	0.088	0.088	0.089	0.091

Note: Dependent variable is the dummy variable indicating informal employment (*is*). Coefficients show marginal effects. Asterisks denote level of significance: *** 1%, ** 5%, and * 10%. Standard errors are adjusted for error term clustering. Coefficient estimates that are reported as "0.000", and yet statistically significant, are smaller than 0.0005 in absolute magnitude.

To gain further insight into how government effectiveness might affect informality, the g index is decomposed into its three sub-components and included in the probit model. The results in column 5 suggest that resources for policy formulation (g_1) are less important than the quality of bureaucracy (g_2) and public service provision (g_3). One interpretation of these results relates to the patterns of public trust in politicians, as discussed in the beginning of the chapter; political promises delivered by means of planning, regulation, and the creation of municipal ordinances may have little effect on citizens' incentives if they doubt that these political efforts will have any real effect on them (Saavedra and Tommasi, 2007). Quality of bureaucracy and public services, on the other hand, might have a more direct effect on incentives in terms of actual enforcement of regulation and other value-added in the formal sector.

The bureaucracy index is the only index that may be further disaggregated in a meaningful manner. Column 6 shows that, among bureaucracy resources, competence of the administrative staff and technical resources appear more important than the type of employment contract of the majority of the staff or what degree of co-ordination the municipality has in some of its core undertakings. Thus there is no clear evidence that informality decreases due to a "Weberian" effect of loyal and motivated bureaucrats (Evans and Rauch, 1999).

The only time when the government effectiveness coefficient becomes statistically insignificant is when municipal gross product per capita is included as a control variable (column 7). While the social norm coefficient remains significant, this suggests that the level of economic development has a more important effect on informality than government effectiveness. However, income per capita should probably be perceived, to a larger extent than any other control variable, as endogenously determined by other municipal characteristics. This potential endogeneity calls for caution in the interpretation of the results when income is included (this is discussed further in Section 4.6.3.).

A few observations are warranted regarding the coefficients in Table 4.3 that are not part of the core hypotheses. Ten individual characteristics other than human capital proxies are included in the analysis. First, being female has a very small positive effect on the probability of being in the informal sector. This gender effect is notable only where there are young children in the household (according to the interaction variable). Disability slightly reduces the chances of participating in the formal sector, and so does being a recent migrant from a rural area. The specification in column 2 contains two additional household characteristics: earned income among other household members and an indicator showing if any other household member is employed in the formal sector. The positive effect of household income suggests that one's incentives to take a job in the formal sector are weakened if one can be supported by other household income. The negative coefficient of formal-sector employment of another household member, however, suggests the opposite: if a household member works in the formal sector, the individual is much more likely to do so too. The correlation of sector participation among household members is quite high (0.45). This high correlation, together with the large rise in pseudo-R^2 when this variable is included, raises some concerns about endogeneity. Although the inclusion of these household variables does not affect the other coefficients to a large extent, they are not included in any of the other specifications reported in Table 4.3.

Among the coefficients of the local economy control variables (z), it can be noted that the share of manufacturing in the local economy has a negative effect on informality. The share of agriculture has a negative effect as well, even though weaker. The relative size of the service sector shows no significant relationship with informality (the left-out sector is the public sector). There is a positive relationship between informality and the share of the urban labor force that consists of recent rural-to-urban migrants (even when controlling for migrant status of the individual). The presence of migrants from rural areas is likely to push down earnings in the informal sector, which would increase the incentives to look for jobs in the formal sector. It could be, though, that the probability of finding a job in the formal sector is also affected negatively by the presence of migrants, and that this effect outweighs the increased incentive due to earnings differentials. Lastly, the size of the local population has a small negative effect on informality, giving some support to the notion that there are higher returns to formality in larger markets. This could be due to greater opportunities for specialization as well as economies of scale in production.

4.6.2. Municipal-level fractional logit model

The municipal-level fractional logit model serves as a complement to the worker-level probit model. Besides constituting a robustness check of the municipal-level findings in the probit model, it allows for an alternative interpretation of the relationship between informality and the explanatory municipal variables. The dependent variable in this model is the share of the municipal urban labor force that is employed informally. The key variables of interest are the same as in the previous model, with the difference that individual human capital is aggregated here to the municipal average. Table 4.4 gives an overview of the results and includes six model specifications. Similar to the probit analysis, the composite governance index g is used in the first specifications, and is then disaggregated into its sub-indices. The coefficients show marginal effects of a small change in each explanatory variable. The results are qualitatively, and in terms of statistical significance, similar to the probit model. The coefficients are not directly comparable to the probit model, since this model contains a fractional, and not a binary, dependent variable.

The coefficient for average education is around –0.08 in the different specifications, and the interpretation is straightforward: holding everything else constant, a one-year increase in average education is associated with an eight-percentage-point decrease in the share of informal labor employment in the average municipality. The government effectiveness coefficient is about -0.16 when no municipal control variables are included (column 1). It is reduced to about half that magnitude once other municipal characteristics are accounted for. Since the index lies between zero and one, the coefficient may be interpreted as an elasticity: a ten-percentage-point increase in the g index is associated with a moderate decline of informality of about one percentage point. The same interpretation applies for the social norm, which has a coefficient between –0.4 and –0.2. A ten-percentage-point increase in the norm index is associated with a two to four-percentage-point decline in informality, when other factors are held constant. In both models, the social norm coefficient is consistently larger than the governance coefficient. This gives some support to the notion that *informal institutions* (understood here as socially sanctioned norms of behavior) play a stronger role in shaping the employment outcome and willingness to participate in the formal economy than the local government's enforcement and implementation of the *formal institutions*.

The disaggregation of the governance index g in column 4 yields a slight deviation from the results of the probit model. While all three coefficients are still negative, policy formulation (g_1) has a larger coefficient than public goods provision. Contrary to the probit results, this suggests that policy

TABLE 4.4. *Estimation results – municipal fractional logit (share informal employment).*

	(1)	(2)	(3)	(4)	(5)	(6)
Labor force skill						
Average education in labor force	-0.086***	-0.080***	-0.077***	-0.077***	-0.075***	-0.058***
Municipal key variables						
Government effectiveness, g	-0.160***	-0.077***	-0.071***			0.013
Policy formulation, g_1				-0.027***	-0.014	
Bureaucratic resources, g_2				-0.037***		
Public goods, g_3				-0.016**	-0.006	
Bur res, employment, g_{21}					0.004	
Bur res, competence, g_{22}					-0.029**	
Bur res, co-ordination, g_{23}					-0.007	
Bur res, IT, g_{24}					-0.057***	
Social norm, n	-0.425***	-0.314***	-0.212***	-0.209***	-0.220***	-0.073***
Local economy						
Share agriculture		-0.031	-0.014	-0.017	-0.018	0.032
Share manufacturing		-0.189***	-0.123***	-0.127***	-0.123***	-0.009
Share services		0.138***	0.172***	0.168***	0.152***	-0.074**
Average firm size		-0.002***	-0.001***	-0.001***	-0.001***	-0.001***
Share rural immigrants		0.058	-0.03	-0.029	-0.018	0.015
Local population			-0.024***	-0.024***	-0.023***	-0.021***
Gross product per capita						-0.107***
Sample size	5,500	5,458	5,434	5,434	5,434	5,434
Akaike Information Criterion	0.896	0.893	0.890	0.891	0.892	0.885

Note: Dependent variable is the share of the municipal labor force in the informal sector (*IS*). Coefficients show marginal effects. Asterisks denote level of significance: *** 1%, ** 5%, and * 10%. Standard errors are adjusted for heteroscedasticity.

formulation matters as well. Further disaggregation of bureaucracy in column 5 again shows that personnel and technical resources have a negative and significant effect on informality, while co-ordination and employment form of administrative staff come out as insignificant. At this level of disaggregation of bureaucracy resources, both g_1 and g_3 turn insignificant, which reveals that the relative importance of these governance sub-indices should be interpreted with some caution. The inclusion of income per capita (column 6) has the same effect as in the probit model; the coefficient of the governance index is rendered insignificant, while the social norm coefficient is smaller but still significant. The same endogeneity concerns apply here as in the probit model. Among the effects of the local economy, it can be seen that the shares of manufacturing and services in the local economy have elasticities ranging from –0.1 to –0.2 and 0.1 to 0.2, respectively. The municipal-level analysis confirms the role of the size of the local economy, represented by the population measure.

4.6.3. Robustness of the results

The robustness checks of the results are concerned primarily with the potential endogeneity of several of the explanatory variables. The endogeneity could be in the form of both reverse causality and correlation of the explanatory variables with the error term. All individual characteristics in the probit model are assumed to be exogenous (except for income and employment among other household members, which are included in only one specification). The first and main suspect for endogeneity is the government effectiveness index (g). It is plausible that the quality of governance is affected by economic and human development together with a range of unobserved factors, which also affect the level of informal employment in a region. An instrumental-variable approach is used in order to take into account this potential endogeneity. Instrumental variable candidates are needed that are correlated with g, yet uncorrelated with the error terms of the two models. Three sets of instrumental variables are used for this purpose, which are largely inspired by other studies. These are summarized as *geography*, *ethnicity*, and *political history*.

Firstly, Naritomi et al. (2007), who use governance as an indicator of institutional development in regions of Brazil, note that "distance to the equator, rainfall, sunshine, average temperatures, and types of soil are all significantly related to both economic and institutional development.

Geography alone explains 65, 30, and 20 percent of the variation in, respectively, ln income per capita, governance, and land Gini" (p. 22). Almeida and Carneiro (2009) analyze a narrower piece of governance: the effect of labor regulation enforcement on informal employment. They suspect that enforcement may be endogenous and use distance to the nearest enforcement office interacted with the local intensity of labor inspectors as an instrumental variable. Nee and Opper (2009) suggest that the quality of bureaucracy may be affected either positively or negatively by the size of a country, in terms of area; while smaller countries tend to be more homogeneous and may respond more easily to citizens' preferences, large countries might benefit from economies of scale in bureaucracy. Inspired by these studies, four variables are included as geographical instruments: latitude, longitude, area of the municipality, and transportation cost to the state capital.

The second set of instruments is inspired by La Porta et al. (1999). They find some empirical evidence that governance performance is worse in countries with higher ethno-linguistic fractionalization and where Catholicism or Islam is the dominant religion. Ethno-linguistic fractionalization is not a pronounced problem in Brazil. To capture another aspect of possible ethnic fractionalization, *racial* fractionalization is used as an instrument. Moreover, the share of Roman Catholics in the municipal population is used to capture the possible effect of religion on governance.

Third, the age of the municipality is used as an instrument. More than 1 400 of Brazil's over 5 500 municipalities were created after the constitutional reform in 1988. There is some anecdotal evidence that some of these municipalities were partly created out of rent-seeking motives and that governance performance has developed quite poorly in some of these new municipalities (IBGE, 2001; The Economist, 2008).

Table 4.5 describes the instrumental variables. The third and fourth columns show the correlations between the instrumental variables and the g index and municipal share of informal employment, respectively. The right-most column shows coefficient estimates from OLS regressions, in which g is regressed on the instruments and the other exogenous municipal variables. All the instruments, except racial fractionalization, show a statistically significant conditional correlation with g.

The instrumental-variable approach is applied both to the probit model and to the municipal-level model. For the probit model, specification (4) from Table 4.3 is used as a benchmark, which includes the composite g index.

TABLE 4.5. *Instrumental variables*

Variable	Mean	Std. Dev.	Correlation with *g* index	Corr with informal empl.	OLS coefficients
Geography					
Transportation cost to state capital	469	409	-0.20	0.25	-0.022***
Municipal area	1,549	5,739	0.01	0.11	0.022***
Latitude	-16	8	-0.41	0.66	-0.005***
Longitude	-46	6	-0.23	0.30	-0.001*
Ethnicity					
Racial fractionalization	0.44	0.14	-0.20	0.43	-0.018
Share Roman Catholic	0.82	0.12	-0.22	0.16	-0.048***
Political history					
Age of municipality	55	57	0.28	-0.12	0.000***

Note: The OLS coefficients are from a regression that includes other municipal control variables. Asterisks show statistical significance level with robust standard errors (*** 1%, ** 5 %, * 10%, respectively). Coefficient estimates that are reported as "0.000", and yet statistically significant, are smaller than 0.0005 in absolute magnitude. N = 5,500.

Attempts have been made to instrument the disaggregated indices g_1, g_2, and g_3 in the probit model, but without success of finding concavity of the likelihood function (IV results in which g_1, g_2, and g_3 are separated are only available for the municipal-level model and are discussed below).

The instrumental variables are added, set by set. Table 4.6 shows the results of the analysis. The author has been unable to find a reliable method to recover the marginal effects from an instrumental probit model; therefore the raw coefficients are shown in column 1. Converting the coefficients in column 1 to marginal effects reproduces column 4 in Table 4.3. Column 2 of Table 4.6 reveals that all geographical instrumental variables are statistically significant.

Column 3 suggests that the ethnic instruments are weak, as their coefficients in the first-stage regression are statistically insignificant; hence the results in columns 2 and 3 differ little from each other. The political history of the municipality is added as an instrument in column 4. The coefficients of the worker characteristics are just minimally affected by the IV approach compared to the regular probit model. Considering the municipal key variables, the g index gets a larger negative coefficient as a result of using instrumental variables, whereas the coefficient on the social norm gets a smaller negative

TABLE 4.6. *Probit model with instrumental variables (probability of informal employment)*

	(1) PROBIT	(2) IV 1	(3) IV 1+2	(4) IV 1+2+3
Human capital				
Age	-0.056***	-0.054***	-0.054***	-0.056***
Age, squared	0.001***	0.001***	0.001***	0.001***
Education	-0.057***	-0.050***	-0.050***	-0.055***
Education, squared	-0.001***	-0.001***	-0.001***	-0.001***
Other individual char's				
Gender - female	0.019**	0.023***	0.023***	0.021**
Household head	-0.086***	-0.083***	-0.083***	-0.085***
Married	-0.127***	-0.128***	-0.128***	-0.128***
Race - black	-0.025***	-0.027***	-0.026***	-0.025***
Disabled	0.119***	0.103***	0.104***	0.114***
Rural-urban migrant	0.029***	0.049***	0.048***	0.037***
Children in household	0.024***	0.010**	0.011***	0.019***
Female X chd in hhd	0.120***	0.118***	0.118***	0.120***
Municipal key variables				
Government effectiveness, *g*	-0.331***	-1.929***	-1.873***	-0.928**
Social norm, *n*	-2.469***	-1.802***	-1.827***	-2.233***
Local economy				
Share agriculture	-0.287**	-1.578***	-1.534***	-0.769**
Share manufacturing	-0.597***	-1.213***	-1.193***	-0.830***
Share services	-0.064	-0.642**	-0.622*	-0.28
Average firm size	-0.002	0.012**	0.011**	0.004
Share rural immigrants	0.538**	-1.103	-1.045	-0.076
Local population	-0.031***	-0.022	-0.022	-0.028**
Instrumental variables				
Transp. cost to state capital		-0.000***	-0.000***	-0.000**
Municipal Area		0.000**	0.000***	0.000***
Latitude		-0.006***	-0.007***	-0.007***
Longitude		-0.004**	-0.005**	-0.005***
Racial fractionalization			0.07	0.11
Share Roman-Catholic			0.085	0.018
Age of municipality				0.000***
Sample size	2,213,429	2,213,429	2,213,429	2,213,429

Note: Dependent variable is the dummy variable indicating informal employment (*is*). Asterisks denote level of significance: *** 1%, ** 5%, and * 10%. Standard errors are adjusted for error term clustering. Coefficient estimates that are reported as “0.000”, and yet statistically significant, are smaller than 0.0005 in absolute magnitude.

coefficient. The coefficients of the local economy control variables are altered in magnitude but not in their signs.

For the municipal-level model, the method of two-stage least squares is used, which is compared to ordinary least-squares results. Table 4.7 presents the results of the instrumental approach as follows: Two model specifications are used as benchmarks. The first uses the aggregated g index and corresponds to model (3) in Table 4.4. Columns 1–4 in Table 4.7 have this specification. The second specification uses the disaggregated g_1, g_2, and g_3 indices and corresponds to model (4) in Table 4.4. Columns 5–8 in Table 4.7 have this specification. A comparison of columns 1 and 2 shows that the results of the fractional logit ("flogit") are in fact very similar to the results of the linear (OLS) probability model. The two instrumental specifications reported in columns 3 and 4 show results that are qualitatively similar to the OLS specification, but the magnitudes of the coefficients change. In absolute magnitudes the coefficient of labor force skill declines, and the coefficients for social norm and the instrumented governance variable increase. The IV results of the specification that uses the disaggregated g indices are somewhat inconclusive; in columns 7 and 8, informality is positively related to public goods, while the negative relationships with the two other governance indicators remain. In addition, the coefficient of the social norm becomes statistically insignificant.

In sum, the instrumental-variable approach for both the individual-level model and the municipal-level model confirms the results in Tables 4.3 and 4.4, but when the governance index is disaggregated so that all three sub-indices are instrumented for, some noticeable deviations are observed. Clearly, the choice of instruments plays a major role when all three indices are instrumented for.

Some further endogeneity concerns and limitations of the empirical results are elaborated on briefly in the remainder of this sub-section. The regression output of the six additional robustness checks, which are discussed here but not included in the chapter, are available from the author upon request.

First, per-capita income, as a general indicator of local economic development, is likely to affect – and be affected by – several observable and unobservable variables (including quality of governance). Due to this endogeneity, it is included only in one specification of each model (column 7 of Table 4.3 and column 6 of Table 4.4). This resulted in a statistically insignificant coefficient estimate of the governance index. To assess the importance of the

TABLE 4.7. *Municipal model with OLS and instrumental variables (share informal employment)*

	(1) FLOGIT	(2) OLS	(3) IV 1+2	(4) IV 1+2+3	(5) FLOGIT	(6) OLS	(7) IV 1+2	(8) IV 1+2+3
Labor force skill								
Average education in labor force	-0.077***	-0.073***	-0.020***	-0.041***	-0.077***	-0.073***	-0.028***	-0.035***
Municipal key variables								
Government effectiveness, g	-0.071***	-0.071***	-1.212***	-0.756***				
Policy formulation, g_1					-0.027***	-0.028***	-0.912***	-1.124***
Bureaucratic resources, g_2					-0.037***	-0.034***	-0.925***	-0.998***
Public goods, g_3					-0.016**	-0.017**	0.530**	0.916***
Social norm	-0.212***	-0.209***	-0.377***	-0.310***	-0.209***	-0.206***	-0.121	-0.031
Local economy								
Share agriculture	-0.014	-0.014	-0.528***	-0.323***	-0.017	-0.016	-0.600***	-0.593***
Share manufacturing	-0.123***	-0.122***	-0.439***	-0.312***	-0.127***	-0.125***	-0.668***	-0.728***
Share services	0.172***	0.164***	-0.402***	-0.176***	0.168***	0.160***	-0.552***	-0.563***
Average firm size	-0.001***	-0.001***	-0.002***	-0.002***	-0.001***	-0.001***	-0.003***	-0.003***
Share rural immigrants	-0.03	-0.031	-0.042	-0.037	-0.029	-0.029	0.133	0.202**
Local population	-0.024***	-0.023***	-0.019***	-0.020***	-0.024***	-0.023***	-0.024***	-0.026***
Constant		1.300***	1.971***	1.703***		1.304***	1.985***	1.925***
Sample size	5,434	5,434	5,434	5,434	5,434	5,434	5,434	5,434
R^2		0.556				0.556		

Note: Dependent variable is the share of the municipal labor force in the informal sector (*IS*). FLOGIT refers to fractional logit, OLS to ordinary least squares, and IV to 2-stage least squares. Asterisks denote level of significance: *** 1%, ** 5%, and * 10%. FLOGIT columns are identical to columns 3 and 4 in Table 4.4. Instrumental variables are the same as in Table 4.6.

municipal key variables while still holding local per-capita income “constant”, the models have been evaluated on a sub-sample of municipalities with relatively homogenous income. A 40-percent sub-sample of municipalities, consisting of the 2 113 “middle-income“ municipalities with per-capita income of between 2 000 and 5 000 R$, was selected for this purpose. While coefficient estimates change slightly in magnitude, no qualitative changes occur with this sub-sample. A tentative conclusion is that the results are not driven to any large extent by differences in productivity or per-capita income.

Second, the structure of the local economy could be endogenously determined in the model, just as governance might be. The structure of the local economy could be affected by human capital intensity, institutional and economic development, geography, or by unobserved characteristics. The models have therefore been evaluated on sub-samples with relatively homogenous structures of the local economy. While municipalities dominated by services and manufacturing do not deviate from the previously obtained results, agriculture-dominated municipalities do not show the same strong relationships between informality and governance.

Third, the empirical literature on spatial human capital externalities is usually concerned about endogenous sorting of skilled people to certain regions (Moretti, 2004). Thus the average level of human capital in a city or region might not be exogenously determined. While it is outside the scope of this study to fully satisfactorily adjust for this possible endogeneity (by the means of additional instrumental variables), the method of evaluating the models on sub-samples has been applied here as well. By using the sub-sample of individuals who have never moved from one municipality to another, some of the endogenous-sorting problem is taken care of.[10] Even if the resulting sample size decreases by half, the key coefficient estimates remain robust, with only minor changes in magnitude. Another sub-sample includes only municipalities with relatively homogenous education (those less than ½ standard deviation away from the average level of education). Parameter estimates from regressions on this sample are similar to those on the full sample, with the exception for two of the disaggregated governance indices in some of the specifications.

[10] It should be recognized, however, that staying in the municipality in which one was born is to some extent an endogenous choice too.

Fourth, in the results discussed above there is no distinction made between the self-employed and employees in the formal sector. While these two groups are treated as homogenous in this study, they might face different income prospects and hence different incentives regarding sector choice. Some empirical studies focus solely on employees (Pratap and Quintin, 2006) or self-employed (Blau, 1985) in the informal sector, due to their potentially fundamental differences. No major deviations in the results are observed when the probit model is estimated on sub-samples with a) all informal employees excluded and b) with all informal self-employed excluded.

Fifth, there is some regional variation in the results. When the models are estimated on each of the five macro regions in Brazil, statistically significant relationships cannot be established between informal employment and the disaggregated governance indices and the social norm indicator for the North or the Center-West. These two regions are sparsely populated and together only account for 14 percent of the sample.

Last, the models have been tested for sensitivity to certain outliers. The models have been estimated with the “tail” municipalities excluded, defined here as those that are at least two standard deviations away from the average level of informality and education. The results do not change noticeably. In sum, few of the results of the sub-sample evaluations of the models deviate from the findings in Tables 4.3 and 4.4. The cases in which key coefficient estimates do come out insignificant, or even of the opposite sign, are those in which small sub-samples are used, such as the Northern and Center-West macro regions, and agriculture-dominated municipalities.

The conclusion from these additional robustness checks is the same as above; the coefficient estimates of main interest generally remain stable and significant with the expected sign. In cases where the sample size is shrunk to a small subset, some deviation in the results is observed.

4.7. CONCLUSION

This chapter aims at shedding some new light on the question of what causes informal employment to vary across regions. It complements the existing empirical literature by seeking to fill the gap between micro-level studies – which explain informal employment as something determined by individual characteristics – and cross-country studies – which explain it in terms of differences in tax systems, labor regulation, and quality of institutions and governance. The challenge in this case is to explain the within-country differences in informality observed in Brazil, ranging from 20 percent of the urban labor force in some municipalities (which is comparable to the informality in Chile) to 80 percent in others (which is comparable to Paraguay).

A theoretical model is proposed to explain these differences in terms of worker skills, local quality of governance, tax rates, enforcement of tax and labor regulations, and non-pecuniary costs and benefits in the informal sector. The empirical assessment of the model supports the main hypotheses: informality is higher where education is lower, where governance is less effective, and where social norms on tax compliance are weaker. There is also some evidence that social norms have a stronger effect on employment outcome than the authorities' enforcement and implementation of formal institutions. These results complement previous studies by showing that regional factors, which are exogenous to the individual, affect individual employment outcomes. Moreover, they complement the cross-country studies by showing that regional differences within a country may cause informality to vary just as much as between countries, despite the fact that the legal system, labor and tax regulation, and other formal institutions all are held constant.

The empirical strategy relies on several indices to represent government effectiveness. Although results are presented for disaggregated indices, to separate the effects of bureaucracy resources, public service provision etc., the question that ultimately remains is exactly what kind of government action is most effective in terms of including its citizens in the formal economy. The analysis in this chapter is not specific enough to provide the answer. To prevent "exit and exclusion" from the formal economy (to use the title of a recent World Bank report on the topic; Perry et al., 2007), the results suggest that the worker's incentive structure needs to be taken into consideration. Without strong incentives to engage in the formal economy, the worker will either "opt out" from it or avoid trying to engage in it in the first place. While education will increase the worker's chances to overcome the skill threshold to the formal sector, flexible labor legislation could probably improve the prospects for some workers to find a job in the formal sector.

The incentive structure need not only contain economic aspects. For the local government the most challenging task – besides providing education, efficient bureaucracy, and other public services to its citizens – might be to improve the quality of the "social contract" between the authorities and the citizen (Saavedra and Tommasi, 2007). This implies making participation in the formal sector the norm rather than an exception. While such norms are likely to change only slowly over time, a government can seek ways to improve the sense of political participation and inclusion among its citizens. This includes transparency in the political decision process and in the spending of public resources, as well as recognizing the needs of the people outside the formal sector just as much as the needs of those who are already in it.

REFERENCES

Almeida, R. & Carneiro, F. (2005). Inequality and Employment in a Dual Economy: Enforcement of Labor Regulation in Brazil. IZA Discussion Paper 3094.

Almeida, R. & Carneiro, F. (2009). Enforcement of Labor Regulation and Firm Size. *Journal of Comparative Economics, 37* (1), 28–46.

Amuedo-Dorantes, C. (2004). Determinants and Poverty Implications of Informal Sector Work in Chile. *Economic Development and Cultural Change*, *52* (2), 347–368.

Bigsten, A., Kimuyu, P., & Lundvall, K. (2004). What to Do with the Informal Sector? *Development Policy Review, 22* (6), 701–715.

Blau, D. (1985). Self-Employment and Self-Selection in Developing Country Labor Markets. *Southern Economic Journal*, *52* (2), 351–363.

Bosch, M., Goni, E., & Maloney, W. (2007). The Determinants of Rising Informality in Brazil: Evidence from Gross Worker Flows. World Bank Policy Research Working Paper 4375.

Chaudhuri, K., Schneider, F., & Chattopadhyay, S. (2006). The Size and Development of the Shadow Economy: An empirical Investigation from States of India. *Journal of Development Economics, 80* (2), 428–443.

de Paula, A. & Scheinkman, J. (2007). The Informal Sector. NBER Working paper 13486.

de Soysa, I. & Jütting, J. (2007). Informal Institutions and Development: How They Matter and What Makes Them Change. In J. Jütting, D. Drechsler, S. Bartsch, & I. de Soysa (Eds.), *Informal Institutions: How Social Norms Help or Hinder Development*. Paris: OECD Development Centre Studies.

Economist (2008). Why Elections, Even to Lowly Office, Are So Expensive. September 25.
[http://www.economist.com/world/americas/displaystory.cfm?story_id=12305155]

El Badaoui, E., Strobl, E., & Walsh, F. (2008). Is There an Informal Employment Wage Penalty? Evidence from South Africa. *Economic Development and Cultural Change*, *56* (3), 683–710.

Evans, P. & Rauch, J. E. (1999). Bureaucracy and Growth: A Cross-National Analysis of the Effects of "Weberian" State Structures on Economic Growth. *American Sociological Review*, *64* (5), 748–765.

Fields, G. S. (1975). Rural-Urban Migration, Urban Unemployment and Underemployment, and Job-Search Activity in LDCs. *Journal of Development Economics, 2* (2), 165–187.

Friedman, E., Johnson, S., Kaufmann, D., & Zoido-Lobatón, P. (2000). Dodging the Grabbing Hand: the Determinants of Unofficial Activity in 69 Countries. *Journal of Public Economics*, *76* (3), 459–493.

Funkhouser, E. (1996). The Urban Informal Sector in Central America: Household Survey Evidence. *World Development*, *24* (11), 1737–1751.

Galiani, S. & Weinschelbaum, F. (2007). Modeling Informality Formally: Households and Firms. CEDLAS working paper 47.

Gasparini, L. & Tornarolli, L. (2007). Labor Informality in Latin America and the Caribbean: Patterns and Trends from Household Survey Microdata. CEDLAS working paper 46.

Gong, X., van Soest, A., & Villagomez, E. (2004). Mobility in the Urban Labor Market: a Panel Data Analysis for Mexico. *Economic Development and Cultural Change 53* (1)*,* 1–36.

Gould, B. W. & Villarreal, H. J. (2006). An Assessment of the Current Structure of Food Demand in Urban China. *Agricultural Economics, 34* (1), 1–16.

Harbaugh, W. T. (1996). If People Vote Because They Like to, Then Why Do So Many of Them Lie? *Public Choice*, *89* (1–2), 63–76.

Harris, J. R. & Todaro, M. P. (1970). Migration, Unemployment and Development: A Two-Sector Analysis. *American Economic Review, 60* (1), 126–142.

Henley, A., Arabsheibani, G. R., & Carneiro, F. G. (2009). On Defining and Measuring the Informal Sector: Evidence from Brazil. *World Development, 37* (5), 992–1003.

IBGE (2001). IBGE traça o perfil administrativo de todos os municípios brasileiros. Press release, April 17, 2001. [http://www1.ibge.gov.br/home/presidencia/noticias/1704munic.shtm]

Johnson, S., Kaufmann, D., & Zoido-Lobatón, P. (1998). Regulatory Discretion and the Unofficial Economy. *American Economic Review, 88* (2), 387–392.

Jütting, J., Parlevliet, J., & Xenogiani, T. (2008). Informal Employment Re-loaded. Working Paper 266, OECD Development Centre.

Kennedy, P. (2008). *A Guide to Econometrics.* Malden, MA: Wiley-Blackwell.

Kucera, D. & Xenogiani, T. (2009). Women in Informal Employment: What Do We Know and What Can We Do? In J. Jütting & J. R. de Laiglesia (Eds.), *Is*

Informal Normal?: Towards More and Better Jobs in Developing Countries (pp. 89–113). Paris: OECD Development Centre.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. (1999). The Quality of Government. *Journal of Law, Economics and Organisations*, *15* (1), 222–279.

Loayza, N. (1996). The Economics of the Informal Sector: A Simple Model and Some Empirical Evidence from Latin America. *Carnegie-Rochester Conference Series on Public Policy, 45*, 129-162.

Loayza, N. & Rigolini, J. (2006). Informality Trends and Cycles. World Bank Policy Research Working Paper 4078.

Maloney, W. (2004). Informality Revisited. *World Development, 32* (7), 1159–1178.

Marcouiller, D., Ruiz de Castilla, V., & Woodruff, C. (1997). Formal Measures of the Informal-Sector Wage Gap in Mexico, El Salvador, and Peru. *Economic Development and Cultural Change, 45* (2), 367–392.

Mazumdar, D. (1976). The Rural-Urban Wage Gap, Migration, and the Shadow Wage. *Oxford Economic Papers*, 28 (3), 406–425.

Ministério do Planejamento, Orçamento e Gestão (2008). Estudo de Atualização do Portfólio dos Eixos Nacionais de Integração e Desenvolvimento, de 2000-2007 para 2004-2011. Mimeo. [http://www.planejamento.gov.br/secretarias/upload/Arquivos/spi/downloads/081014_DOWN_EX_PNC_ArDep_area.pdf]

Moretti, E. (2004). Human Capital Externalities in Cities. In J. V. Henderson & J. F. Thisse (Eds.), *Handbook of Regional and Urban Economics, Volume 4* (pp. 2243–2291). Amsterdam: Elsevier Science.

Moulton, B. R. (1990). An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Unit. *Review of Economics and Statistics*, *72* (2), 334–38.

Naritomi, J., Soares, R. R., & Assunção, J. J. (2007). Rent Seeking and the Unveiling of 'de facto' Institutions: Development and Colonial Heritage within Brazil. NBER Working Paper 13545.

Nee, V. & Opper, S. (2009). Bureaucracy and Financial Markets. *Kyklos, 62* (2), 293–315.

Papke, L. & Wooldridge, J. (1996). Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates. *Journal of Applied Econometrics*, *11* (6), 619–632.

Perry, G., Maloney, W., Arias, O., Fajnzylber, P., & Mason, A. (2007). *Informality: Exit and Exclusion.* Washington, D.C.: World Bank.

Pisani, M. & Pagán, J. (2004). Sectoral Selection and Informality: a Nicaraguan Case Study. *Review of Development Economics*, *8* (4), 541–556.

Pradhan, M. & van Soest, A. (1995). Formal and Informal Sector Employment in Urban Areas of Bolivia. *Labour Economics, 2* (3), 275–297.

Pratap, S. & Quintin, E. (2006). Are Labor Markets Segmented in Developing Countries? A Semiparametric Approach. *European Economic Review*, *50* (7), 1817–1841.

Primo, D. M., Jacobmeier, M. L., & Milyo, J. (2007). Estimating the Impact of State Policies and Institutions with Mixed-Level Data. *State Politics and Policy Quarterly*, *7* (4), 446–459.

Rauch, J. E. (1991). Modelling the Informal Sector Formally. *Journal of Development Economics*, *35* (1), 33–47.

Saavedra, J. & Tommasi, M. (2007). Informality, the State and the Social Contract in Latin America: A Preliminary Exploration. *International Labour Review*, *146* (3–4), 279–309.

Schneider, F. (2005). Shadow Economies around the World: What Do We Really Know? *European Journal of Political Economy*, *21* (3), 598–642.

Schneider, F. & Enste, D. (2000). Shadow Economies: Size, Causes, and Consequences. *Journal of Economic Literature, 38* (1), 77–114.

Straub, S. (2005). Informal Sector: The Credit Market Channel. *Journal of Development Economics*, *78* (2), 299–321.

Torgler, B. & Schneider, F. (2007). Shadow Economy, Tax Morale, Governance and Institutional Quality: A Panel Analysis. CESifo Working Paper 1923.

Verbeek, M. (2008). *A Guide to Modern Econometrics.* West Sussex: Wiley.

World Bank (2008). *World Development Indicators.* Washington D.C.: World Bank.

World Economic Forum (2008). *The Global Competitiveness Report 2008–2009*. Geneva.

Acknowledgments: David Fairris, Roger Martini, and Sonja Opper have provided valuable feedback on this chapter. I am also grateful for comments from participants in seminars at the Trade & Agriculture Directorate and the Development Centre at OECD. Francisco Figueiredo greatly facilitated the access to several of the data sources used in the empirical analysis.

APPENDIX 1: DEFINING THE INFORMAL SECTOR

In a detailed discussion on the definition and measurement of the informal sector, Perry et al. (2007) compare several measures of the informal sector. Four of these are i) percent informal labor force under the legalistic definition (the worker is informal if she does not have the right to a pension linked to employment when retired), ii) percent informal labor force under the productive definition (the worker is considered informal if she is unskilled self-employed, a salaried worker in a small private firm, or an unpaid worker), iii) percent self-employed, and (iv) percent of labor force not covered by a pension scheme (close to the legalistic definition, following the definition used by World Bank, 2008). The correlations of these measures across countries are high, especially between the legalistic and productive definitions (0.90), which suggests that the employment status largely determines what rights and conditions a worker may expect in the labor market. Common for these measures of the informal sector is that they are aimed at identifying those in the labor force who are unable to benefit from mandated benefits (such as paid sick leave or pensions) and who work under conditions that do not comply with labor market regulations (maximum hours of work per week, minimum pay, safety standards, etc.).

When the informal sector is defined as consisting of the self-employed only, a large group of wage laborers is excluded. Still the correlation of the share of self-employed and the informal sector as defined by the other measures is fairly high (around 0.7). The correlations between the various informality measures that define the informal sector as a share of the labor force and the size of the shadow economy (as share of GDP, defined by Schneider, 2005) are between 0.4 and 0.6 (Perry et al., 2007), suggesting some, but far from complete, coincidence between the two measures.

The definition used in this study lies conceptually closest to the legalistic definition mentioned above. The nine employment categories defined in the Brazilian Demographic Census are provided in Table 4.A1, which shows the percentage of the urban labor force in each category. Informal employment is defined here as being an unregistered employee (categories 2 and 4), an employer with less than five employees who do not contribute to any social security institution (parts of category 5), a self-employed person who does not contribute to any social security institution (parts of category 6), or an unpaid worker (categories 7–9). Henley et al. (2009) provide an analysis of three alternative measures of informal employment in Brazil, using the PNAD survey. The definition used here corresponds to a large extent to their measure of informality defined as "no signed labor card".

TABLE 4.A1. *The urban labor force by employment status*

Employment status	Percent of the labor force	Formal or informal sector (FS/IS)	Percent of informal sector	Percent of informal sector, paid
1 – Registered domestic employee	2.5	FS	..	..
2 – Unregistered domestic employee	5.9	IS	13.1	13.9
3 – Registered employee	38.8	FS	..	..
4 – Unregistered employee	25.9	IS	42.5	45.2
5 – Employer	2.3	FS/IS [a]	1.6	1.7
6 – Self-employed	21.9	FS/IS [a]	36.8	39.2
7 – Trainee (unpaid)	0.4	IS	1.0	..
8 – Helping household member (unpaid)	1.4	IS	3.0	..
9 – Working for self-consumption (unpaid)	0.9	IS	2.0	..
Total	100		100	100

Note: The labor force includes everyone who is between 15 and 65 years old and reports an occupation. [a] All employers with five or less employees, and self-employed, who do not pay social security contributions are considered informal. *Source*: Brazilian Demographic Census 2000.

Table 4.A2 shows how the informality measure used in this study compares with the measures discussed by Perry et al. (2007). The reason that the measure used here is ten percentage points lower than the productive definition used by Gasparini and Tornarolli (2007) is that their measure includes both the urban and the rural labor force. When they exclude the rural labor force, their measure for Brazil is 45.6 percent for 2003.

TABLE 4.A2. *Size of the informal sector in Brazil and Latin America by different measures (percent)*

Measure:	This study [a]	Legalistic [b]	Productive [b]	Self-employed [c]	Lack of pensions [d]	Shadow economy [e]
Brazil	45	35	55	31	63	40
Latin America	..	53	60	34	68	42

Sources: [a] Author's calculation based on the Brazilian Demographic Census 2000; [b] Gasparini and Tornarolli (2007); [c] Loayza and Rigolini (2006); [d] World Bank (2008); [e] Schneider (2005).

Schneider's (2005) measure of the shadow economy (as percent of GDP) is close to the measure of the informal sector used in this study, but, given that his estimate includes the entire economy, it should be compared with informality measures that consider the entire (urban and rural) labor force. Following the productive definition, it appears as if the share of labor under informal employment arrangements is considerably higher than the share of GDP that is classified as unofficial.

APPENDIX 2: CONSTRUCTION OF THE GOVERNMENT EFFECTIVENESS INDICES

This appendix describes the construction of the governance index variables that are used in the empirical analysis. Table 4.2 (pages 110–111) gives some basic information about these variables. The composite index for *government effectiveness, g*, is constructed as the average of the three disaggregated indices *policy formulation* (g_1), *resources for the bureaucracy* (g_2), and quality of *public services provision* (g_3):

$$g = (g_1 + g_2 + g_3)/3 \tag{A1}$$

These three sub-indices are defined below.

Policy formulation (g_1): The policy formulation index is constructed by means of factor analysis of 20 binary variables. The index consists of the first principal factor normalized to lie between 0 and 1. The complete list of variables is provided in Table 4.A3.

Resources for the municipal bureaucracy (g_2): The index is constructed as a weighted sum of the four sub-indices *employment contract* (g_{21}), *competence of administrative staff* (g_{22}), *degree of coordination* (g_{23}), and *IT resources* (g_{24}).

Index g_{21} is based on the share of the administrative staff in the municipality with a statutory employment contract and eligible for employment benefits. Index g_{22} is based on the share of the administrative staff with secondary or higher education. The g_{23} index is constructed with factor analysis, based on 12 binary variables to capture the degree of co-ordination of key municipal activities; these are listed in Table 4.A3. The g_{24} index is also constructed with factor analysis and is based on seven variables to capture the information technology resources available to the municipal bureaucracy; these are listed in Table 4.A3.

In the g_2 index, the first two variables are each given a weight of 1/3, and the last two variables are each given 1/6:

$$g_2 = (g_{21} + g_{22})/3 + (g_{23} + g_{24})/6 \quad \text{(A2)}$$

Public service provision (g_3): The index is constructed with factor analysis and is based on ten variables that measure the quantity and quality of municipal public service provision. The variables contained in the index are listed in Table 4.A3.

TABLE 4.A3. *Variables used in government effectiveness indices*

Variables for g_1 (all binary)	***variable code***
1. A prefeitura desenvolve política ou plano de inclusão digital	A100
2. Lei orgânica municipal	A138
3. Lei do orçamento anual (LOA)	A139
4. Lei de diretrizes orçamentárias (LDO)	A140
5. Plano Plurianual de investimentos (PPA)	A141
6. Portaria do órgão gestor da educação	A142
7. Plano Municipal de educação	A166
8. Conselho Municipal de Educação	A186
9. Conselho Municipal de Segurança Pública	A219
10. Conselho Municipal de Política urbana, Desenvolvimento Urbano, da Cidade ou similar	B64
11. Lei de parcelamento do solo	B68
12. Lei de zoneamento ou equivalente	B69
13. Código de obras	B70
14. Existência de lei específica de Solo criado	B71
15. Existência de lei específica de Contribuição de melhoria	B72
16. Existência de lei específica de Operação urbana consorciada	B73
17. Existência de lei específica de Estudo de impacto de vizinhança	B74
18. Legislação específica que dispões sobre regularização fundiária	B201
19. Programa específico de regularização fundiária	B202
20. Conselho municipal de Cultura	B263

TABLE 4.A3. *(Continued)*

Variables for g_{23} (all binary)	
Articulações interinstitucionais:	
1. Educação	B116X
2. Saúde	B121X
3. Assistência e desenvolvimento social	B126X
4. Direito da Criança e Adolescente	B131X
5. Emprego e/ou Trabalho	B136X
6. Turismo	B141X
7. Cultura	B146X
8. Habitação	B151X
9. Meio ambiente	B156X
10. Transporte	B161X
11. Desenvolvimento urbano	B166X
12. Saneamento e/ou Manejo de Resíduos Sólidos	B171X
Variables for g_{24} (all binary)	
1. Cadastro imobiliário informatizado – existência	A3
2. Planta Genérica de Valores informatizada – existência	A8
3. Cadastro ISS informatizado – existência	A10
4. Computadores ligados em rede na prefeitura – existência	A56
5. Intranet - existência	A69
6. Computadores com acesso a internet – existência	A70
7. Cadastro informatizado ou levantamento de famílias interessadas em programas habitacionais	B181
Variables for g_3 (1–8 are binary)	
1. Página na internet da prefeitura: Serviços informativos do município e notícias	A83
2. Página na internet da prefeitura: Acesso a documentos e formulários	A84
3. Programas ou ações na área de habitação: Construção de unidades	B182
4. Programas ou ações na área de habitação: Oferta de material de construção	B187
5. Programas ou ações na área de habitação: Oferta de lotes	B190
6. Programas ou ações na área de habitação: Outras ações	B193
7. Bibliotecas públicas: Alguma é mantida pelo poder público municipal	B269
8. A prefeitura garante o acesso ao público aos derviços disponibilizados na sua página na internet	A96X
9. teachers per 100 pupils in public schools, *Ensino Fundamental*	BIM3E07X
10. health centers per 1,000 municipal citizens	BIM3S04X

Note: The variables were obtained from the following sources: variables for g_1, g_{23}, and g_{24} – Gestão Pública, 2005-2006; variables for g_3 – Gestão Pública, 2005–2006 and Base de Informações Municipais, 2000.

APPENDIX 3: CORRELATION MATRICES

TABLE 4.A4. *Correlations between individual-level variables used in the empirical analysis*

	Informal	Age	Education	Female	HH head	Married	Black	Disabled	Migrant	HH child	HH formal	HH income
Informal	1.00											
Age	-0.04	1.00										
Education	-0.28	-0.15	1.00									
Female	0.00	-0.03	0.14	1.00								
HH head	-0.04	0.37	-0.14	-0.42	1.00							
Married	-0.09	0.39	0.00	-0.08	0.25	1.00						
Black	0.02	0.01	-0.09	-0.01	0.01	-0.04	1.00					
Disabled	0.03	0.08	-0.07	0.00	0.02	0.01	0.01	1.00				
Migrant	0.05	-0.03	-0.11	-0.02	0.01	-0.01	0.00	0.01	1.00			
HH child	0.04	-0.13	-0.10	-0.03	0.14	0.13	0.02	-0.01	0.02	1.00		
HH formal	0.00	-0.09	0.08	0.16	-0.36	0.02	-0.01	-0.01	0.00	-0.07	1.00	
HH income	-0.05	0.00	0.18	0.08	-0.14	-0.02	-0.03	-0.01	-0.02	-0.07	0.14	1.00

Note: Sample size: 2,222,387. See Table 4.2 for details.

TABLE 4.A5. *Correlations between municipal-level variables used in the empirical analysis (part 1 of 3)*

	Inf.	g	g_1	g_2	g_{21}	g_{22}	g_{23}	g_{24}	g_3	N
Informality	1.00									
g	-0.41	1.00								
g_1	-0.36	0.79	1.00							
g_2	-0.28	0.61	0.36	1.00						
g_{21}	-0.02	0.27	0.07	0.78	1.00					
g_{22}	-0.29	0.34	0.24	0.45	0.06	1.00				
g_{23}	-0.15	0.40	0.31	0.42	0.01	0.09	1.00			
g_{24}	-0.41	0.55	0.44	0.47	0.05	0.20	0.19	1.00		
g_3	-0.29	0.82	0.42	0.24	0.00	0.17	0.24	0.37	1.00	
Norm	-0.34	0.12	0.10	0.17	0.06	0.32	-0.01	0.10	0.05	1.00
Agriculture	0.10	-0.17	-0.16	0.02	0.08	-0.02	-0.07	-0.02	-0.19	0.00
Manuf.	-0.42	0.36	0.32	0.15	-0.04	0.17	0.14	0.29	0.30	0.18
Services	0.38	-0.25	-0.23	-0.24	-0.06	-0.19	-0.09	-0.34	-0.14	-0.22
Public sec	0.54	-0.48	-0.43	-0.33	-0.04	-0.26	-0.19	-0.51	-0.33	-0.25
Firm size	-0.04	-0.02	-0.02	-0.06	-0.04	-0.03	0.00	-0.07	0.02	-0.01
Local pop.	-0.25	0.20	0.18	0.04	-0.09	0.09	0.10	0.14	0.19	0.11
Local inc	-0.40	0.27	0.24	0.17	-0.01	0.18	0.12	0.26	0.21	0.21
Rural imm	0.10	-0.07	-0.04	0.01	0.04	0.00	-0.06	0.00	-0.10	-0.02
Trp cost	0.25	-0.20	-0.16	-0.13	-0.01	-0.16	-0.03	-0.20	-0.16	-0.24
Mun. area	0.11	0.02	0.04	-0.02	-0.02	-0.10	0.10	-0.02	0.02	-0.20
Latitude	0.66	-0.40	-0.36	-0.36	-0.10	-0.29	-0.13	-0.48	-0.24	-0.35
Longitude	0.32	-0.24	-0.26	-0.22	-0.07	-0.16	-0.14	-0.25	-0.10	-0.14
Fractional.	0.43	-0.20	-0.19	-0.18	-0.05	-0.26	0.00	-0.20	-0.11	-0.35
Roman	0.17	-0.22	-0.22	-0.06	0.07	0.04	-0.20	-0.20	-0.18	0.09
Age of m.	-0.11	0.28	0.23	0.08	-0.05	0.02	0.18	0.17	0.27	-0.07

TABLE 4.A5. *(Part 2 of 3)*

	Agriculture	Manuf.	Services	Public sector	Firm size	Local pop.	Local income	Rural imm.	Trp cost	Mun. area
Agriculture	1.00									
Manuf.	-0.63	1.00								
Services	-0.47	-0.35	1.00							
Public sector	-0.25	-0.42	0.82	1.00						
Firm size	-0.15	0.12	0.03	0.07	1.00					
Local pop.	-0.28	0.29	-0.02	-0.15	0.11	1.00				
Local income	-0.04	0.47	-0.50	-0.51	0.05	0.15	1.00			
Rural imm.	0.34	-0.21	-0.21	-0.06	-0.07	-0.13	0.00	1.00		
Trp cost	0.23	-0.28	0.04	0.14	-0.11	-0.27	-0.12	0.22	1.00	
Mun. area	0.03	-0.07	0.03	0.04	-0.02	-0.10	-0.03	0.05	0.34	1.00
Latitude	-0.21	-0.21	0.50	0.65	0.18	-0.16	-0.36	0.00	0.22	0.19
Longitude	-0.34	0.08	0.39	0.43	0.18	0.09	-0.21	-0.28	-0.32	-0.26
Fractional.	-0.19	-0.08	0.31	0.35	0.08	-0.01	-0.21	-0.04	0.10	0.11
Roman	0.10	-0.19	0.13	0.26	-0.03	-0.22	-0.16	-0.09	0.00	-0.12
Age of mun.	-0.30	0.23	0.10	-0.10	0.04	0.14	0.06	-0.33	-0.22	0.05

TABLE 4.A5. *(Part 3 of 3)*

	Latitude	Longitude	Fractional.	Roman	Age of mun.
Latitude	1.00				
Longitude	0.47	1.00			
Fractional.	0.55	0.36	1.00		
Roman	0.13	0.25	-0.08	1.00	
Age of mun.	0.02	0.19	0.08	-0.07	1.00

Note: Sample size: 5,506. See Table 4.2 for details.

Lund Economic Studies

1. Guy Arvidsson	Bidrag till teorin för verkningarna av räntevariationer, 1962
2. Björn Thalberg	A Trade Cycle Analysis. Extensions of the Goodwin Model, 1966
3. Bengt Höglund	Modell och observationer. En studie av empirisk anknytning och aggregation för en linjär produktionsmodell, 1968
4. Alf Carling	Industrins struktur och konkurrensförhållanden, 1968
5. Tony Hagström	Kreditmarknadens struktur och funktionssätt, 1968
6. Göran Skogh	Straffrätt och samhällsekonomi, 1973
7. Ulf Jakobsson och Göran Norman	Inkomstbeskattningen i den ekonomiska politiken. En kvantitativ analys av systemet för personlig inkomstbeskattning 1952-71, 1974
8. Eskil Wadensjö	Immigration och samhällsekonomi. Immigrationens ekonomiska orsaker och effekter, 1973
9. Rögnvaldur Hannesson	Economics of Fisheries. Some Problems of Efficiency, 1974
10. Charles Stuart	Search and the Organization of Marketplaces, 1975
11. S Enone Metuge	An Input-Output Study of the Structure and Resource Use in the Cameroon Economy, 1976
12. Bengt Jönsson	Cost-Benefit Analysis in Public Health and Medical Care, 1976
13. Agneta Kruse och Ann-Charlotte Ståhlberg	Effekter av ATP - en samhällsekonomisk studie, 1977
14. Krister Hjalte	Sjörestaureringens ekonomi, 1977
15. Lars-Gunnar Svensson	Social Justice and Fair Distributions, 1977
16. Curt Wells	Optimal Fiscal and Monetary Policy - Experiments with an Econometric Model of Sweden, 1978
17. Karl Lidgren	Dryckesförpackningar och miljöpolitik - En studie av styrmedel, 1978
18. Mats Lundahl	Peasants and Poverty. A Study of Haiti, London, 1979
19. Inga Persson-Tanimura	Studier kring arbetsmarknad och information, 1980
20. Bengt Turner	Hyressättning på bostadsmarknaden - Från hyresreglering till bruksvärdesprövning, Stockholm 1979
21. Ingemar Hansson	Market Adjustment and Investment Determination. A Theoretical Analysis of the Firm and the Industry, Stockholm 1981
22. Daniel Boda Ndlela	Dualism in the Rhodesian Colonial Economy, 1981
23. Tom Alberts	Agrarian Reform and Rural Poverty: A Case Study of Peru,

	1981
24. Björn Lindgren	Costs of Illness in Sweden 1964-75, 1981
25. Göte Hansson	Social Clauses and International Trade. An Economic Analysis of Labour Standards in Trade Policy, 1981
26. Noman Kanafani	Oil and Development. A Case Study of Iraq, 1982
27. Jan Ekberg	Inkomsteffekter av invandring, 1983
28. Stefan Hedlund	Crisis in Soviet Agriculture?, 1983
29. Ann-Marie Pålsson	Hushållen och kreditpolitiken. En studie av kreditrestrik-tioners effekt på hushållens konsumtion, sparande och konsumtionsmönster, 1983
30. Lennart Petersson	Svensk utrikeshandel, 1871-1980. En studie i den intraindustriella handelns framväxt, 1984
31. Bengt Assarsson	Inflation and Relative Prices in an Open Economy, 1984
32. Claudio Vedovato	Politics, Foreign Trade and Economic Development in the Dominican Republic, 1985
33. Knut Ödegaard	Cash Crop versus Food Crop Production in Tanzania: An Assessment of the Major Post-Colonial Trends, 1985
34. Vassilios Vlachos	Temporära lönesubventioner. En studie av ett arbetsmarknadspolitiskt medel, 1985
35. Stig Tegle	Part-Time Employment. An Economic Analysis of Weekly Working Hours in Sweden 1963-1982, 1985
36. Peter Stenkula	Tre studier över resursanvändningen i högskolan, 1985
37. Carl Hampus Lyttkens	Swedish Work Environment Policy. An Economic Analysis, 1985
38. Per-Olof Bjuggren	A Transaction Cost Approach to Vertical Integration: The Case of Swedish Pulp and Paper Industry, 1985
39. Jan Petersson	Erik Lindahl och Stockholmsskolans dynamiska metod, 1987
40. Yves Bourdet	International Integration, Market Structure and Prices. A Case Study of the West-European Passenger Car Industry, 1987
41. Krister Andersson and Erik Norrman	Capital Taxation and Neutrality. A study of tax wedges with special reference to Sweden, 1987
42. Tohmas Karlsson	A Macroeconomic Disequilibrium Model. An Econometric Study of the Swedish Business Sector 1970-84, 1987
43. Rosemary Vargas-Lundius	Peasants in Distress. Poverty and Unemployment in the Dominican Republic, 1989
44. Lena Ekelund Axelson	Structural Changes in the Swedish Marketing of Vegetables,

	1991
45. Elias Kazarian	Finance and Economic Development: Islamic Banking in Egypt, 1991
46. Anders Danielson	Public Sector Expansion and Economic Development. The Sources and Consequences of Development Finance in Jamaica 1962-84, 1991
47. Johan Torstensson	Factor Endowments, Product Differentiation, and International Trade, 1992
48. Tarmo Haavisto	Money and Economic Activity in Finland, 1866-1985, 1992
49. Ulf Grönkvist	Economic Methodology. Patterns of Reasoning and the Structure of Theories, 1992
50. Evelyne Hangali Maje	Monetization, Financial Development and the Demand for Money, 1992
51. Michael Bergman	Essays on Economic Fluctuations, 1992
52. Flora Mndeme Musonda	Development Strategy and Manufactured Exports in Tanzania, 1992
53. Håkan J. Holm	Complexity in Economic Theory. An Automata Theoretical Approach, 1993
54. Klas Fregert	Wage Contracts, Policy Regimes and Business Cycles. A Contractual History of Sweden 1908-90, 1994
55. Per Frennberg	Essays on Stock Price Behaviour in Sweden, 1994
56. Lisbeth Hellvin	Trade and Specialization in Asia, 1994
57. Sören Höjgård	Long-term Unemployment in a Full Employment Economy, 1994
58. Karolina Ekholm	Multinational Production and Trade in Technological Knowledge, 1995
59. Fredrik Andersson	Essays in the Economics of Asymmetric Information, 1995
60. Rikard Althin	Essays on the Measurement of Producer Performance, 1995
61. Lars Nordén	Empirical Studies of the Market Microstructure on the Swedish Stock Exchange, 1996
62. Kristian Bolin	An Economic Analysis of Marriage and Divorce, 1996
63. Fredrik Sjöholm	R&D, International Spillovers and Productivity Growth, 1997
64. Hossein Asgharian	Essays on Capital Structure, 1997
65. Hans Falck	Aid and Economic Performance - The Case of Tanzania, 1997
66. Bengt Liljas	The Demand for Health and the Contingent Valuation

		Method, 1997
67.	Lars Pålsson Syll	Utility Theory and Structural Analysis, 1997
68.	Richard Henricsson	Time Varying Parameters in Exchange Rate Models, 1997
69.	Peter Hördahl	Financial Volatility and Time-Varying Risk Premia, 1997
70.	Lars Nilsson	Essays on North-South Trade, 1997
71.	Fredrik Berggren	Essays on the Demand for Alcohol in Sweden - Review and Applied Demand Studies, 1998
72.	Henrik Braconier	Essays on R&D, Technology and Growth, 1998
73.	Jerker Lundbäck	Essays on Trade, Growth and Exchange Rates, 1998
74.	Dan Anderberg	Essays on Pensions and Information, 1998
75.	P. Göran T. Hägg	An Institutional Analysis of Insurance Regulation – The Case of Sweden, 1998
76.	Hans-Peter Bermin	Essays on Lookback and Barrier Options - A Malliavin Calculus Approach, 1998
77.	Kristian Nilsson	Essays on Exchange Rates, Exports and Growth in Developing Countries, 1998
78.	Peter Jochumzen	Essays on Econometric Theory, 1998
79.	Lars Behrenz	Essays on the Employment Service and Employers' Recruitment Behaviour, 1998
80.	Paul Nystedt	Economic Aspects of Ageing, 1998
81.	Rasha M. Torstensson	Empirical Studies in Trade, Integration and Growth, 1999
82.	Mattias Ganslandt	Games and Markets – Essays on Communication, Coordination and Multi-Market Competition, 1999
83.	Carl-Johan Belfrage	Essays on Interest Groups and Trade Policy, 1999
84.	Dan-Olof Rooth	Refugee Immigrants in Sweden - Educational Investments and Labour Market Integration, 1999
85.	Karin Olofsdotter	Market Structure and Integration: Essays on Trade, Specialisation and Foreign Direct Investment, 1999
86.	Katarina Steen Carlsson	Equality of Access in Health Care, 1999
87.	Peter Martinsson	Stated preference methods and empirical analyses of equity in health, 2000
88.	Klas Bergenheim	Essays on Pharmaceutical R&D, 2000
89.	Hanna Norberg	Empirical Essays on Regional Specialization and Trade in Sweden, 2000
90.	Åsa Hansson	Limits of Tax Policy, 2000
91.	Hans Byström	Essays on Financial Markets, 2000

92. Henrik Amilon	Essays on Financial Models, 2000
93. Mattias Lundbäck	Asymmetric Information and The Production of Health, 2000
94. Jesper Hansson	Macroeconometric Studies of Private Consumption, Government Debt and Real Exchange Rates, 2001
95. Jonas Månsson	Essays on: Application of Cross Sectional Efficiency Analysis, 2001
96. Mattias Persson	Portfolio Selection and the Analysis of Risk and Time Diversification, 2001
97. Pontus Hansson	Economic Growth and Fiscal Policy, 2002
98. Joakim Gullstrand	Splitting and Measuring Intra-Industry Trade, 2002
99. Birger Nilsson	International Asset Pricing, Diversification and Links between National Stock Markets, 2002
100. Andreas Graflund	Financial Applications of Markov Chain Monte Carlo Methods, 2002
101. Therése Hindman Persson	Economic Analyses of Drinking Water and Sanitation in Developing Countries, 2002
102. Göran Hjelm	Macroeconomic Studies on Fiscal Policy and Real Exchange Rates, 2002
103. Klas Rikner	Sickness Insurance: Design and Behavior, 2002
104. Thomas Ericson	Essays on the Acquisition of Skills in Teams, 2002
105. Thomas Elger	Empirical Studies on the Demand for Monetary Services in the UK, 2002
106. Helena Johansson	International Competition, Productivity and Regional Spillovers, 2003
107. Fredrik Gallo	Explorations in the New Economic Geography, 2003
108. Susanna Thede	Essays on Endogenous Trade Policies, 2003
109. Fredrik CA Andersson	Interest Groups and Government Policy, A Political Economy Analysis, 2003
110. Petter Lundborg	Risky Health Behaviour among Adolescents, 2003
111. Martin W Johansson	Essays on Empirical Macroeconomics, 2003
112. Joakim Ekstrand	Currency Markets - Equilibrium and Expectations, 2003
113. Ingemar Bengtsson	Central bank power: a matter of coordination rather than money supply, 2003
114. Lars Pira	Staples, Institutions and Growth: Competitiveness of Guatemalan Exports 1524-1945, 2003
115. Andreas Bergh	Distributive Justice and the Welfare State, 2003

116. Staffan Waldo	Efficiency in Education - A Multilevel Analysis, 2003
117. Mikael Stenkula	Essays on Network Effects and Money, 2004
118. Catharina Hjortsberg	Health care utilisation in a developing country -The case of Zambia, 2004
119. Henrik Degrér	Empirical Essays on Financial Economics, 2004
120. Mårten Wallette	Temporary Jobs in Sweden: Incidence, Exit, and On-the-Job Training, 2004
121. Tommy Andersson	Essays on Nonlinear Pricing and Welfare, 2004
122. Kristian Sundström	Moral Hazard and Insurance: Optimality, Risk and Preferences, 2004
123. Pär Torstensson	Essays on Bargaining and Social Choice, 2004
124. Frederik Lundtofte	Essays on Incomplete Information in Financial Markets, 2005
125. Kristian Jönsson	Essays on Fiscal Policy, Private Consumption and Non-Stationary Panel Data, 2005
126. Henrik Andersson	Willingness to Pay for a Reduction in Road Mortality Risk: Evidence from Sweden, 2005
127. Björn Ekman	Essays on International Health Economics: The Role of Health Insurance in Health Care Financing in Low- and Middle-Income Countries, 2005
128. Ulf G Erlandsson	Markov Regime Switching in Economic Time Series, 2005
129. Joakim Westerlund	Essays on Panel Cointegration, 2005
130. Lena Hiselius	External costs of transports imposed on neighbours and fellow road users, 2005
131. Ludvig Söderling	Essays on African Growth, Productivity, and Trade, 2005
132. Åsa Eriksson	Testing and Applying Cointegration Analysis in Macroeconomics, 2005
133. Fredrik Hansen	Explorations in Behavioral Economics: Realism, Ontology and Experiments, 2006
134. Fadi Zaher	Evaluating Asset-Pricing Models in International Financial Markets, 2006
135. Christoffer Bengtsson	Applications of Bayesian Econometrics to Financial Economics, 2006
136. Alfredo Schclarek Curutchet	Essays on Fiscal Policy, Public Debt and Financial Development, 2006
137. Fredrik Wilhelmsson	Trade, Competition and Productivity, 2006
138. Ola Jönsson	Option Pricing and Bayesian Learning, 2007

139. Ola Larsson	Essays on Risk in International Financial Markets, 2007
140. Anna Meyer	Studies on the Swedish Parental Insurance, 2007
141. Martin Nordin	Studies in Human Capital, Ability and Migration, 2007
142. Bolor Naranhuu	Studies on Poverty in Mongolia, 2007
143. Margareta Ekbladh	Essays on Sickness Insurance, Absence Certification and Social Norms, 2007
144. Erik Wengström	Communication in Games and Decision Making under Risk, 2007
145. Robin Rander	Essays on Auctions, 2008
146. Ola Andersson	Bargaining and Communication in Games, 2008
147. Marcus Larson	Essays on Realized Volatility and Jumps, 2008
148. Per Hjertstrand	Testing for Rationality, Separability and Efficiency, 2008
149. Fredrik NG Andersson	Wavelet Analysis of Economic Time Series, 2008
150. Sonnie Karlsson	Empirical studies of financial asset returns, 2009
151. Maria Persson	From Trade Preferences to Trade Facilitation, 2009
152. Eric Rehn	Social Insurance, Organization and Hospital Care, 2009
153. Peter Karpestam	Economics of Migration, 2009
154. Marcus Nossman	Essays on Stochastic Volatility, 2009
155. Erik Jonasson	Labor Markets in Transformation: Case Studies of Latin America, 2009