Fall 2007

Spatial aspects of infant mortality and informal workers: The case of Ceara State -Brazil

Ricardo Soares
University of New Hampshire, Durham

SPATIAL ASPECTS OF INFANT MORTALITY AND INFORMAL WORKERS: THE CASE OF CEARA STATE - BRAZIL

BY

RICARDO SOARES

B.A., Economics, Federal University of Ceara in Brazil, 1996

M.A., Economics, University of New Hampshire, 2001

DISSERTATION

Submitted to the University of New Hampshire

in Partial Fulfillment of

the Requirements for the Degree of

Doctor of Philosophy

in

Economics

September, 2007

This dissertation has been examined and approved.

Dissertation Director, Karen S. Conway,
Professor of Economics

Ju-Chin Huang, Associate Professor
of Economics

Bruce T. Elmslie, Professor of Economics

Jonathan C. Rork, Assistant Professor of
Economics, Georgia State University

Robert S. Woodward, Forrest McKerley
Professor of Health Economics

5/18/07
Date

ACKNOWLEDGEMENTS

I am for life indebted to my advisor Karen Conway for her outstanding support and guidance on the development of this dissertation. I am also very grateful to Robert Woodward for his encouragement and editing lessons. My gratitude also goes to the committee members of this dissertation – Bruce Elmslie, Ju-Chin Huang, Jonathan Rork, and Robert Woodward - whose insights and suggestions were primordial to improve the quality of this work.

I would like to thank my wife Wladia and my family with all my love for their incredible support and understanding before and during the elaboration of this thesis. I am also thankful to all my friends from UNH, and especially Edinaldo, Fabricio, Minghua, Liping, Andrea, Nevin, and Dora. My thanks also go to Sinthy for her support during all these years at UNH.

I would like to thank all my friends from the UNH community for sharing all the ups and downs of my stay at UNH (Analia, Heloisa, Heloisa, Carlos, Neide, Daniel, Isabel, Leninha, Inês, Humberto, Beta, Evilene, Gonzalo, Isabel, Cassiano, Leticia, Hudson, Manoel, Marcos, Lina, Davi, Amanda, Flavio e Jocka).

Finally, I would like to thanks the office of the National System of Employment (SINE/IDT – CE) located in Fortaleza for generously providing the data set for this thesis, and the same thanks go to the Health Secretariat of Ceara State.

TABLE OF CONTENTS

LIST OF TABLES

PART 1

PART 2

LIST OF FIGURES

PART 1

PART 2

LIST OF GRAPHS

PART 1

PART 2

ABSTRACT

SPATIAL ASPECTS OF INFANT MORTALITY AND INFORMAL WORKERS: THE CASE OF CEARA STATE – BRAZIL

by

Ricardo Soares

University of New Hampshire, September, 2007

High levels of infant mortality and high participation of informal workers in the labor market are living conditions faced by many developing countries. In Ceara State - Brazil, their trends during the last three decades have followed opposite directions. Whereas infant mortality has decreased substantially since the 1980s, suggesting that the country is on the right path to development, the labor market has presented increasing levels of informality, which challenges traditional theories of development. The essays of this thesis aim to investigate some aspects of these two phenomena. In particular, the first essay offers an approach to analyze if the health care program known as the Community Health Worker Program has been an important factor explaining the downward trend in infant mortality. Moreover, it provides a framework to identify if the effectiveness

of this program is under-estimated when traditional methods of evaluation do not consider its spatial spillover effect. This effect occurs when improvements in this health care program in one municipality affect not only infant mortality in that municipality but also the infant mortality of neighboring municipalities. The second essay studies heterogeneity among informal workers and investigates possibilities of social interactions and spatial segmentation among different types of workers. Besides sharing the same study site (Ceara Sate – Brazil), both essays incorporate and empirically verify the hypotheses that neighborhood conditions (at city or urban sector levels) play an important role in: i) amplifying the benefits of health care programs, and ii) affecting workers' behavior in the labor market.

INTRODUCTION

High levels of infant mortality and high participation of informal workers in the labor market are living conditions faced by many developing countries. In Ceara State - Brazil, their trends during the last three decades have followed opposite directions. Whereas infant mortality has decreased substantially since the 1980s, suggesting that the country is on the right path to development, the labor market has presented increasing levels of informality, which challenges traditional theories of development. The essays of this thesis aim to investigate some aspects of these two phenomena. The common hypothesis included in both essays is that externalities in health policies or labor market outcomes can be transmitted through neighboring areas (municipalities or neighborhoods). These externalities take the form of spillover effects in health policies (Essay I) or neighborhood effects in the informal labor market (Essay II).

The first essay of this thesis, therefore, investigates the determinants of infant mortality rate (IMR) with a suggestive modeling perspective, which is able to capture and empirically verify the significance of spillover effects for an important healthcare program (Community Health Worker Program). Spillover effects exist when improvements in the program affect not only the IMR of the benefited municipality but also the IMR of neighboring municipalities. These spillovers may occur because the populations of neighboring municipalities share not only (crowded) public health services but also behavioral attitudes related to the risk of infant death. A health care program providing basic information (water filtering, baby formulas, adequate sanitation, breastfeeding incentives, etc) and regular follow up for families can improve the health care of these families, and can be spontaneously replicated across the border of municipalities by other families, also improving the quality of the shared public health services[1].

Essay I investigates if the use of traditional reduced form models to investigate variations in infant mortality rate among contiguous municipalities can underestimate the effectiveness of health policy programs when spillover possibilities are not taken into consideration. In this case, the under-valuation of the health care program may have a tremendous social cost when under investments in the program are measured in fewer infants saved.

The second essay of this thesis aims to provide evidence that: i) formal versus informal dichotomies may not be adequate to classify workers in the labor market, and ii) workers' position in the labor market may also be influenced by social interactions in urban neighborhoods. These two possibilities have usually been neglected in labor market segmentation studies, which have enjoyed renewed interest in the literature with the persistence and growth of the informal economy worldwide during the 1990s. In the labor market of Fortaleza City – Brazil, for example, more than half of the workers are informal as they are either self-employed or work in small-firms usually evading labor legislation (absence of social security contributions, for example).

[1] This improvement in the quality can occur due to reductions in the demand for particular types of shared health services or with a better selectivity from illness conditions avoidable by primary health care (diarrhea, for example).

Labor market segmentation occurs when there are institutional (minimum wage, labor legislation, etc) or social (stigma) barriers to mobility between sectors. That is, individuals with the same endowments may find different difficulties to acquire formal and informal jobs. Limited social mobility in the labor market is a sign of inefficiency in this market which may aggravate the traditional problems associated with a large informal economy (the weakening of tax and social security bases and the camouflage of usual labor market indicators (unemployment rate, for example)[2].

Empirically, more recent segmentation studies have tested the possibility of segmentation in urban labor markets by analyzing asymmetries in the likelihood of transitions between sectors in a dual labor market (formal versus informal sectors) and/or without considering the possibility of social interactions between workers. This essay argues that neither the duality in the differentiation of workers nor the absence of social interaction should be the norm in segmentation studies. Informal workers as have been empirically defined by lack of social security contributions or by small-size firms may encompass a set of differentiated workers (and activities) for which aggregation is inadequate. Some informal positions, for example, may not correspond to the segmented view of under-employment. The aggregation in this case hides important dynamic features of informal workers which should be considered for labor market policies.

---

[2] A low level of unemployment rate can associated with a high level of informality rate which not necessarily indicates good conditions in the labor market.

It is also interesting to note that the informal literature has devoted little or no attention to the influence of social interactions on worker's sector decision. Networking and other forms of spontaneous informational channels (observing or hearing about) are responsible for a considerable portion of new jobs acquired (Topa, 2001). Neglecting the possibility of neighborhood interactions in the (informal) labor market can also compromise the reliability of segmentation studies (models).

Essay II, therefore, suggests a sector allocation model which not only allows differentiation among informal workers, but also includes the possibility of neighborhood interactions among the different types of workers. These two modifications may also be included in (future) segmentation studies, which require the use of panel data not yet available for the labor market of Fortaleza City, Brazil.

PART 1: SPATIAL EFFECTS IN INFANT MORTALITY MODELS: THE CASE OF CEARA STATE – BRAZIL

CHAPTER 1

INTRODUCTION

In 1999 approximately 5 percent of infants died before reaching one year of age in Ceara State Brazil. Although this average is considered high by the standards of the World Health Organization (WHO), it was much worse at the end of the 1980's, when the level of infant mortality was twice that level. Efforts to reduce infant mortality in Ceara State have been recognized not only by other States in Brazil but also by international organizations, who have suggested the adoption of similar health policies in other developing countries[1]. The basic prescription of the local government involves the combination of investments improving socio-economic conditions (mainly education and sanitation) with investments in a health care program where the population assisted receive primary medical attention at home from trained community health workers[2] (Community Health Work Program - CHWP). Although these are the general recommendations for the state as a whole, each municipality[3] within the state has adopted more or less such prescriptions. As infant mortality rates among these municipalities also vary considerably, one should ask if they are

[1] The World Bank (WDR, 1998/99 p. 122) and the World Health Organization have cited the Ceara State as an example of a poor region that has decreased considerably the levels of infant mortality.
[2] Community health workers are civilians who are trained by nurses and physicians to provide basic health care information to families.
[3] Municipalities are the smallest geographical areas with their own local government. It is equivalent to counties in the US.

correlated[4]. The basic objective of this research, therefore, is to provide an answer.

Home visiting programs based on community participation are still one of the most important flags of the primary health care model of the World Health Organization (WHO, (1978) and WHO, (2005)) in combating infant mortality in less developed areas. Barnes-Boyd et al (2001) counted at least 14 published experiences worldwide with mixed results in significantly improving infant health. Differences in methods of analysis were indicated by the authors as the main cause of this mixed result in the evaluation of the programs worldwide.

In Ceara State, Brazil, the CHWP has been referred to as a well established community program in descriptive analyses such as in Tendler (1997) and Svitone (2000)[5]. Regression analyses based on reduced form models derived from health economics frameworks (Corman and Grossman, 1985), however, are missing. These empirical strategies are appropriate to identify the causal link between health policies and health outcomes (infant mortality rate) at the municipal data level. On the other hand, the general use of these frameworks so far has neglected the possibility that improvements in health policies in one municipality may also affect the infant mortality risk of the families living in neighboring municipalities. This spillover effect can be induced by improvements in the quality of the public health services shared by the population of

[4] The infant mortality rate among the municipalities ranges from 10 deaths for each 1000 live births to 120 deaths per 1000 births.

[5] These descriptive analyses refer to the operational description of the program and suggestive (rather than inferential) associations between reductions in infant mortality rate and the adoption of the program.

neighboring municipalities[6], or even by the sharing of information between families in neighbor municipalities (well) covered or otherwise by the program.

This essay proposes modeling perspectives to reduced form models that allow spillover effects in health policies at the municipal data level. It uses Ceara State's experience with the CHWP to show how a traditional reduced form model can underestimate the efficacy of these programs when the model neglects extensive effects in neighboring municipalities. In order to accommodate these spatial spillover effects this essay suggests a methodology that combines the theoretical scope of reduced form models (Corman and Grossman (1985), Rosezweigh and Schultz (1982), Corman et al (1986)) with the statistical treatment of spatial analysis (Anselin (1986), Anselin and Hudak (1992)).

The following seven sections of this essay discuss in greater details how spatial considerations may improve reduced form models for variations in infant mortality rates among municipalities within a state. They provide information respectively describing: i) the study site and the health policy under consideration; ii) literature review on the evaluation of the health care program, reduced form models of infant mortality and spatial analysis; iii) theoretical modeling perspectives; iv) empirical models; v) empirical methodology; vi) empirical results; and vii) further discussions about the findings of this paper and complementary analyses.

---

[6] The preventive health care behavior advised by the programs can reduce the usual overcrowded demand for health services in public hospitals, for example, which benefit all the populations sharing these services.

It is shown here that the exclusion of health spillover effects in traditional regional models of infant mortality can lead to incorrect guidance for policy design. In the case of Ceara State, it is estimated that under certain conditions the productivity of the health care policy program installed can be up to 39% greater than the productivity predicted in traditional models when spillover effects are ignored. The empirical analysis also confirms the hypothesis that a multidisciplinary approach is needed to reduce significantly the level of infant mortality rate in Ceara State. Improvements in the health structure have to be followed by better socio-economic conditions (and vice-versa) in order to decrease considerably the risk of infant death among the municipalities of Ceara State.

CHAPTER 2

STUDY SITE AND HEALTH POLICIES

With an area[7] of 146,348 Km^2 and a population of 7,430,661[8] individuals in 2000, Ceara State (Figure 1) is located in the poorest region of Brazil (Northeast). The state illiteracy rate is 40% and the income inequality is one of the worst in the country. These signs of underdevelopment across individuals are also verified across municipalities. Differences in socio-economic, geographic and health structure conditions contribute directly to the great variance in infant mortality rates across municipalities.

The concentration of state production in a few municipalities also reflects disparities in the provision of adequate health services. In 2000, 6 out of 184 municipalities produced more than 60% of the gross state product (GSP). Four of these municipalities are located in the metropolitan area of Fortaleza (the State's capital) which also contains 80% of the State's beds for neonatal intensive care units. Populations in small underprivileged municipalities usually have to commute to receive medical services in the next better-endowed municipality. In 1998, for example, approximately 26% of the mothers commuted to other municipalities for delivery service.

---

[7] Ceara state has approximately the same land area as New York State.
[8] Population Census of 2000.

Cyclical outbreaks of diseases such as cholera, dengue and less intensively malaria still characterize the epidemiological pattern of the state with 23% of the causes of death for babies less than a year old in 1998 being related to infectious or bacterium diseases. Therefore, the high percentage of infant deaths that can be avoided by adequate primary care, the intensive practice of commuting for health services, and the persistence of epidemic diseases make Ceara state a good study case in analyzing spatial effects in infant mortality models. In particular, spillover effects in health policy are of great interest in this essay[9]. These effects occur when the benefits of health policies expand to beyond the borders of the jurisdiction (municipality) where they were implemented. That is, the health policy affects the risk of infant death not only in the municipality where the program was implemented (or improved) but also in the neighboring municipalities.

Health spillover effects are usually conditioned by the inter-municipal demand for health services or simply by behavioral influences between the populations of two regions. Improvements in the health structure[10] in one municipality can affect the population of all municipalities involved in this inter-municipal commuting for health services. In addition, health care programs that promote habit changes toward a more preventive healthy pattern can also be spontaneously replicated through informal channels in neighbor municipalities. Therefore, the transmission mechanism of spillover in health policies can be

[9] Additional spatial effects are modeled as spatial error or spatial lag effects that are further analyzed ahead in this essay.
[10] The health structure of the city includes the human and capital resources available for prevention and treatment of medical needs.

structural or behavioral. The former reflects the sharing of health infrastructures by populations of different municipalities, and the latter is related to the sharing of better information on preventive measures against illness conditions.

Home visiting programs are mostly designed to encourage preventive care by providing information on how behavioral attitudes can affect the risk of survival of infants. Improvements in these programs can produce spillover effects as they may improve the quality of the health services shared by populations of different municipalities. This may occur if the preventive behavior advised to and adopted by the population of one municipality diminishes demand for crowded and (municipal) shared health services. That is, if the population of one municipality demands less public health services, for example, the population of neighboring municipalities has to wait less to get these services or they can receive them with better quality. The spillover of home visiting programs can also occur simply because the populations of neighboring municipalities share information as well as they share behavioral attitudes. Therefore, even if the population of one municipality has not received the proper attention of a health care program, they may adopt some primary health care measures (water filtering, for example) because others (relatives or friends) are adopting them in neighboring municipalities.

Since 1989, besides additional investments in sanitation and health structure, the government of Ceara State in Brazil has implemented a health care program in each municipality known as the Community Health Worker Program (CHWP - *Programa Agentes Comunitários de Saúde*). These investments and program are the result of the combined efforts of governments at the local (municipal), state and federal level. Each municipality uses resources from its own budget, and from state and federal transfers, to provide public and universal health services to their citizens and municipal neighbors.

The CHWP employs community health workers (CHWs) whose main activities consist of visiting families at their homes to provide health and nutritional education. They weigh children less than two years old, provide baby formulas for oral rehydration, collect information on health indicators, and refer care in health units for pregnant women or any other family member in need of health treatment. The data collected by the CHWs are pooled at the municipal level by their supervisors.

The program initially targeted the 45 poorest municipalities in 1989, when CHWs were recruited from individuals with at least 5 years of local residence with basic education and proven commitment to social service. The program was extended to other municipalities at the demand of unassisted populations (Svitone, 2000) and in 1998 all 184 municipalities were receiving the program. The CHWP is designed to select one CHW for every 75 homes in rural areas and every 225 homes in urban areas. However, this basic prescription is not always followed. The number of community health workers per capita in each municipality varies because while the state sponsors the salaries of the CHWs the municipalities have to hire nurses and physicians as supervisors and instructors. Therefore, the distribution of CHW among the municipalities depends on the shared responsibility of the municipality, represented by its mayor, and the

State Health Secretariat that directs the entire program. Each municipality, however, is allowed to increase the number of CHWs using their own resources, which makes the policy "semi-independent" with respect to the each municipal jurisdiction of the state.

The program targets not only infant health but also the health of the family as a whole, and thus, by extension, the health of the community (as it allows the tracking and controlling of contagious diseases). In addition, the policy creates new job opportunities for the population, which can alleviate the adverse economic conditions of the region. Therefore, the use of the policy in each municipality is not restricted to the target of reducing infant mortality. Its cross-municipal variability with respect to the general prescription is determined by the preferences of each mayor regarding priority policies. These conditions suggest that reverse causality problems in empirical models are of less concern than the immediate perception suggests[11]. In addition, this essay follows Corman and Grossman (1985) in using time lag values of infant mortality rates as an additional regressor to control for potential problems of reverse causality[12].

The benefits of the CHWP in reducing infant mortality in Ceara State have been partially promulgated by expositional (Svitone et al (2000), Tendler (1997), World Bank (1998/1999)) and empirical studies (Souza et al (1999), Sourza et al (2000), Souza et al (2001), Lindsay et al (2002)). These studies, however, have neither considered empirical strategies based on reduced form models to

[11] Reverse causality occurs when the intensity of the policy is induced by the level of infant mortality. This makes the policy endogenous generating biased estimates for the model.
[12] The pros and cons of this strategy are further discussed in section 5.1.3.

analyze policy efficiency, nor included the possibility of spatial effects of any type. The main contribution of this essay is to offer an empirical strategy of estimation of infant mortality models that combines the theoretical framework of reduced form equations with spatial statistical analysis. Within this framework it is possible to fully capture the efficiency of the health care program in reducing infant mortality rate locally and statewide.

CHAPTER 3

LITERATURE REVIEW

Infant mortality is a prime example of a challenging issue that welcomes multiple methodological approaches to understand its causes. Models explaining variations in infant mortality rates have been estimated under distinct perspectives motivated by epidemiological concerns, population trends, family studies, or economic growth and policy. As the main contribution of this essay is to investigate municipal externalities in the CHWP in Ceara State by using spatial analysis in reduced form models of infant mortality, this review focuses on three types of literature: i) studies examining municipal variations in infant mortality rate in Ceara State, ii) references describing the framework for estimating reduced form models of infant mortality, and iii) studies exemplifying the growing use of spatial analysis.

3.1. Ecological Models of Infant Mortality in Ceara State

Ceara State in the northeast part of Brazil was a pioneer in implementing the CHWP in the country. As the program developed through the 1990's, covering more municipalities and generating a reliable data source, some scholars began to use its informational data to investigate variations in infant mortality rate. Souza et al (1999) and Lindsay (2002), for example, examine variations in the infant mortality rate and the diarrhea specific infant mortality rate among the municipalities of Ceara state during the 1994-96 period[13].

Both studies used multivariate ecological models, where predictors of infant mortality are differentiated as proximate determinants (prevalence of low birthweight, breastfeeding and infant weight gain), health-related behavioral variables (participation in growth monitoring programs, immunizations, and prenatal care) and socio-economic conditions (sanitation, illiteracy rate and income). The community health worker program is the source of the data for the calculations of the first two types of predictors (proximate and health behavioral variables), but it is not included as an additional predictor in the empirical models[14]. The ecological models also fail to identify and treat the proximate determinants and the health behavioral variables as endogenous predictors in the regression models. In addition, they disregard any possibility of spatial effect such as the spillover effect of any health care program. These ecological models, therefore, present empirical limitations as they (i) lack theoretical identification of exogenous and endogenous determinants of infant mortality rates and (ii) neglect to account any form of spatial effect.

In general, empirical models based on the theory of demand for health (Grossman, 1972) respond precisely to the first critique but fail to adjust to the second one; in contrast, the "social-medical" spatial literature produces models

[13] In both studies 140 out of 184 municipalities were used in the sample. The selection of the municipalities was based on availability of monthly information of the set of information provided by the CHWP.
[14] The only variable possibly directly linked to the CHWP is the percentage of infants participating in growth monitoring program. The authors however do not specify if this monitoring program is in fact the CHWP. Even in this case this variable should not be included in reduced form models as the behavioral demand for health services is endogenous.

that are just the opposite[15]. This essay merges these approaches and shows in the next sections the missing and the favorable parts of some representative studies.

3.2. Reduced Form Models of Infant Mortality in Health Economics

Grossman's 1972 paper on demand for health provide the reference-framework in Health Economics to understanding variations in health conditions and health behaviors. One important feature of this framework is the distinction between intermediate inputs in production of health conditions and exogenously given inputs that shape and move demand functions for such intermediate inputs. The modeling consequences of this distinction are the differentiations between reduced form and structural models of infant mortality rate. The first include only exogenous variables and are attractive as they provide straightforward estimates for the total impact of health policies on the risk of infant death. The second usually divide the total impact of health policies in direct and indirect effects as they affect infant mortality rate directly or through endogenous intermediate inputs[16].

Proximate determinants of infant mortality like low birth weight or inadequate weight gain are actually outcome variables and therefore should not be included in reduced form equations. Behavioral variables indicating demand

---

[15] For expositional purposes the social-medical spatial literature in this essay refers to those studies using spatial analysis in infant mortality models without necessarily identify the distinction between reduced form and structural form models.

[16] For a more complete discussion on the differentiation between reduced form and structural models in health economics see Rosenzweig and Schultz (1982).

for health services like prenatal or neonatal care are also considered endogenous intermediate inputs, as they are also induced by exogenously given health policies or socio-economic conditions. Regarding health service policy variables, Corman and Grossman (1985) emphasize the differences between variables expressing their availability and their use. The first are exogenous because they express the instruments of health policies, while the second are endogenous because they are the results of the policies. In general, this seems to be the basic difference between social-medical studies and health economic studies. Whereas the distinction between outcome and policy variables or exogenous and endogenous variables is a major concern for the latter, the former makes less effort to draw this distinction.

Corman and Grossman (1985) and Corman et al (1988) estimate reduced form models for infant mortality rates among counties of the US, where policy availability of health services such as abortion or family planning programs were some of the policy variables of interest.

When reduced form models are estimated with macro level data including bordering municipalities, the relationship between availability and use of health services and their impact on infant mortality requires additional spatial considerations. In many circumstances, the population affected by the health policy is not restricted to the population of the municipality sponsoring the policy. In addition to this spatial spillover in health policies, other spatial effects such as spatially defined ecological problems (river contamination, for example) or even the mismatch between the municipal geographical space and the risk of infant death may generate spatial correlation problems that bias the model's estimates. The modeling and identification of these spatial effects though, have been increasingly considered by the literature on spatial analysis.

3.3. Spatial Effects in Economics and Socio-Medical Studies

In their 1985 empirical model of infant mortality Corman and Grossman used non-bordering US counties with population size greater than 50,000 as the unit of analysis. The sample selection was justified by two reasons. The first, more statistically based, was that infant mortality rate estimates can be miscalculated for small regions as the ratio of infant deaths per live births are subjected to small number irregularities. That is, one additional infant death for a small population, for example, may produce a variation in the measurement of infant mortality rate that is much higher than the true variation in the probability of infant death. The second, more theoretically based, was that small regions are not self-sufficient in producing and consuming their own level of health services, and therefore they lack, to a certain extent, control over their health production possibilities. If the first issue is attenuated in small regions of less developed countries by the unfortunate reality that the number of infant deaths is much higher than in developed countries, the second issue can find a solution in the use of spatial analysis.

Since the work of Moran (1948) and more recently Cliff and Ord (1981) and Anselin et al (1986), spatial analysis has been used in many different studies to capture the importance of the geographical position of the unit of analysis in the hypothesis being investigated[17]. Externalities in economic growth (Fingleton, 2001; Ford, 2002), in tax policies (Case et al, 1993; Conway and Rork, 2004), in regional prices (Case, 1991) or in disease incidence (Cliff and Ord, 1981) are examples of situations where spatial effects are important.

With respect to health indicators, epidemiological studies have made great progress in using spatial analysis to estimate distributional maps for infant mortality rates (Assuncao et al (1998), Shimakura et al (2001) and Clayton et al (1993)). Few exceptions however have made use of spatial econometric methodology to estimate mortality rate models[18]. Andrade and Szwarcwald (2001) use Moran's I index to observe if there is spatial correlation in neonatal infant mortality rates among the barrios of the municipality of Rio de Janeiro, Brazil. They observe how the spatial correlation disappears as explanatory variables are added to the regression model. Lorant et al (2001) compared a traditional weighted least square (WLS) regression model with a spatial autoregressive model (SAR) for standard mortality rate among the municipalities of Belgium. Also testing for spatial correlation with a modified Moran's I index, they find that the estimates of the SAR model are more reliable than those of the WLS.

Both previous studies (Andrade and Szwarcwald (2001) and Lorant et al (2001)) follow the spatial econometric methodology of testing and adjusting

---

[17] This geographical position can be understood as the physical location of the region, or even its economic position (weight) with respect to different partners in trade.
[18] The spatial econometric methodology is referenced here by Anselin (1992) where diagnostic tests for spatial correlation are conducted in traditional regression models. In cases where spatial correlation is significant spatial models are suggested.

regressive models when spatial correlation is present. These studies, however, do not present any specific health care policy as explanatory variables, concentrating efforts in measuring associations between demographic characteristics and mortality rates. Consequently, they also do not consider the possibility that spatial spillovers in health policy variables may be driving the significant presence of spatial correlation.

This essay proposes a modeling perspective that not only adds spatial analysis to reduced form models but also allows spillover possibilities in health policies. This combination of a theoretical framework for health policy analysis and a statistical spatial methodology is an important contribution to the literature. The application for a specific case study (CHWP) reveals how traditional models without spatial effects may undervalue the effectiveness of policy programs when compared to the spillover models proposed here. A better exposition of how this modeling perspective is accomplished is found in the next section.

CHAPTER 4

SPATIAL MODELING AND EXPLANATORY VARIABLES

This section provides the modeling perspectives of reduced form models with spatial effects together with some contextualization for the case of infant mortality rate in Ceara State. In addition, it also identifies the main explanatory variables used in this essay.

4.1. Reduced Form Models with Spatial Effects

In general, reduced form equations for infant mortality are expressed as:

(1) $$IMR = X\beta + H\gamma + \varepsilon$$

where IMR is the vector of infant mortality rates (or its algebraic transformation such as square root or log) for n different regions, X is an n x k matrix of covariates indicating socio-economic conditions, H is an n x s matrix expressing the availability of health services (health structure and health policies), and ε is a vector of random shocks assumed to be independent of each other. In this model, spatial effects caused by contagion possibilities, regional spillover effects of health policies, or even by the practice of commuting for health services are ignored. Any of these effects can violate the assumption of independence of the residuals. The presence of spatial correlation implies that the lack of independence of the residuals follows a geographical pattern given by some weighting matrix (W) [19]. The result is that the $E(\varepsilon_i, \varepsilon_j) \neq 0$, which compromises the estimates of equation 1.

The presence of spatial correlation implies biased and/or inefficient estimates depending on the form of the spatial effect. This correlation can capture uncontrolled possibilities of spatial effects (in isolation or jointly) such as: i) spatial health spillover; ii) unobserved shocks or variables that affect jointly neighbors' infant mortality level; and/or iii) spatial mismatch between the phenomenon being investigated and the unit of analysis. In these cases, modifications in the reduced form model are necessary and the following section specifies how each of these spatial effects may be modeled.

4.1.1. Policy spillover models

When spillovers in health services are present, the correct specification for the infant mortality model is given by equation 2.

$$IMR = X\beta + H\gamma + W^h H\delta + \varepsilon \tag{2}$$

where W^h is an n x n weighting matrix averaging the availability of health services in the neighboring municipalities, and δ is a vector of parameters which captures the general spillover effect of health services[20] in neighboring municipalities. The presence of spatial correlation when equation 1 is estimated could be significant

[19] The covariance matrix is such that the Cov $(\varepsilon_i, \Sigma_j w_{ij}\varepsilon_j) \neq 0$. $\Sigma_j w_{ij}\varepsilon_j$ is the weighted average of the residuals of municipality i's neighbors.

[20] The estimation of this equation can be carried out by OLS (or WLS) given that only additional explanatory variables are added to the model.

due to the omission of the health spillover variables (W^hH). In this case, equation 2 can be carried out by traditional OLS (or WLS) given only additional explanatory variables are added to the model[21].

The survival probability of infants in one particular region is not exclusively affected by the availability of health services only in that particular region. Commuting for health services is a common practice for a population living in small and/or underprivileged municipalities. In this case, it is very likely that any improvement in these services also affects the infant mortality rates of neighboring regions. Simply, a public hospital serves not only the population of the municipality where it is located, but also others. Even if a health policy is exclusive for the population in one particular region (as in the case of the CHWP), information spillover effects of the policy can also contribute to reduce infant mortality in the surrounding areas.

In neighboring regions equally characterized by high susceptibility of the population to acquiring infectious disease due to the lack of adequate information on preventive behavior, the effects of informational health policies will probably be extended to those regions where the policy was not implemented. Failing to account for this possibility may make the health policy look less effective. In a

[21] Two additional assumptions are however necessary to use these traditional methods. The first refers to the exogenous character of the policy and the second to the absence of interaction in the distribution of the health services. The interaction occurs when the health policy decisions of one municipality affect and is affected by the health policy decisions of the neighbors. Whereas it is assumed that this second assumption holds in our case, additional empirical adjustments can be made when the first assumption does not hold. These adjustments are explained in the next section.

cost/benefit analysis this distortion can result in a less than efficient distribution of health care programs.

Model 2 assumes that only the neighbors' availability of health services has a direct impact on the levels of infant mortality rates. That is, the socio-economic conditions of neighboring municipalities are not included directly as an additional variable in the model. These conditions may also affect the risk of infant death. This possibility however, requires additional assumptions regarding the transmission mechanism that goes beyond the main objective of this essay. This essay, however, argues that if neighbors' socio-economic conditions are important, model 2 would still present spatial correlation, and their effect can also be captured with the spatial lag or the spatial error models suggested ahead.

4.1.2. Spatial error models

The level of infant mortality in neighboring regions can be correlated because these regions are susceptible to the same risk of temporary and unpredicted epidemic diseases, ecological disturbance, or any other unobserved factor that crosses municipal borders. In this case, the spatial effect operates via the error term. This last term in equation 1 or 2 is given by:

$$\varepsilon = \lambda W^h \varepsilon + \mu \quad (3)$$

where λ is the parameter detecting the unobservable spatial effect, and μ is a well-behaved disturbance with mean 0 and variance matrix $\sigma^2 I$. The correct specification for the model in this case should be:

$$(4) \qquad IMR = X\beta + H\gamma + W^hH\delta + (I - \lambda W^h)^{-1}\mu$$

Equation 4 suggests that additional unobserved factors are important to explain the distribution of infant mortality[22]. That is, it is possible that not all important socio-economic and health conditions have been included as predictors for infant mortality in the model, or that unmeasured environmental or epidemiological factors are also playing a role in explaining infant mortality variation. These unobserved spatial externalities in infant mortality models can be interpreted as any risk factor of infant death that would be given by unmeasured regional behavioral variables (cultural aspects of the population for example), by disease incidence, and/or by ecological problems (river contamination or other pollution problems).

The distinction between purely random spatial shocks and unobserved variables with spatial effects is difficult to address. However, in the spatial context it is possible at least to identify if potential "unobservable" variables are a spatial combination of the observable (measurable) phenomena at hand. That is, what is really missing in equation 2 is not an additional (possibly non-measurable) spatially determined explanatory variable (a lake contamination affecting the adequate supply of water to neighboring municipalities, for example) but the spatial externality of the available explanatory variables. This possibility is explored in the next model specification.

[22] Equation 4 can be estimated by maximum likelihood as in Anselin (1992).

4.1.3. Spatial lag models

In infant mortality models, spatial lag specifications work as spatial filters necessary for model specification. The spatial lag model is defined as:

(5) $$IMR = X\beta + H\gamma + W^h H\delta + \rho W^h IMR + \varepsilon$$

where the parameter ρ is the spatial autoregressive coefficient. Spatial lag models have been more recently used in geography economics to express the presence of externalities in economic conditions (Fingleton, 2001) and to underlie regional competition in prices and taxes (Case et al, 1993; Conway and Rork, 2004). Following Anselin (1992), spatial lag models can be interpreted in two ways. In the first, the spatial lag variable ($W^h IMR$) works as an additional (endogenous) explanatory variable. That is, the model expresses spatial dependence where the value of the dependent variable in one region is a function of their values in the surrounding areas. The spatial lag factor captures the interactivity of the phenomenon itself, which is usually modeled theoretically (Brueckner, 1998). In this case, the focus of analysis is the possibility of significant interaction while controlling for the effect of other variables. The main interest is the endogenous spatial effect itself and its meaning. This interpretation however is not taken in this essay which primarily investigates efficiencies in health policies.

This essay gives another interpretation to the spatial lag term in equation 5. It is only a statistical modeling requirement to control the misspecification of the model. It is the filtering factor that corrects the difference between the area of jurisdiction (municipality) and the dimensionality of the phenomenon (risk of infant mortality). In this case, the main interest of the analysis is the unbiased marginal effect of the explanatory variables (health policy indicators) once the spatial correlation problem is controlled for. Although contagion possibilities would be a theoretical explanation for spatial lag models of infant mortality rates, this essay only assumes a controlling role for the spatial lag variable.

Equation 5 can be estimated with the use of an instrumental variable for the spatial lag[23]. In a first step, the predicted value for *WIMR* is estimated from the spatial lag structure of the exogenous explanatory variables (*WX*). This predicted value is included afterwards as an explanatory variable in equation 5 to provide consistent estimators for the parameters of the model (Kelejian and Prucha (1998)). The use of instrumental variable (IV) for this model specification has some advantages. First, it does not rely on parametric assumptions (normal distribution of residuals, for example) to provide consistent estimators as the maximum likelihood estimation does (Anselin, 1992). Especially for small samples, this propriety can be translated into efficiency advantages over parametric estimation methods.

Second, the spatial lag model with IV estimation is less sensitive to the presence of spatial correlation in the error term. That is, this estimation method is still adequate for the spatial lag structure when the spatial error specification is also important. The instrumental variable reduces the potential correlation problem between unobservables and the spatial lag variable (*WIMR*) determined

[23] Equation 5 can also be estimated by maximum likelihood as in Anselin (1992).

by other reasons (outbreak of a contagious diseases like dengue, for example) not captured by the spatial lag.

Finally, the IV approach not only corrects the endogeneity problem of the spatial lag but also gives it a controlling interpretation which accounts for the possible spatial effects of observed socio-economic conditions. This way, the health spillover effect, if significant, will capture only the direct effect that neighbor's policy has on infant mortality rates. It is a net effect for the policy that goes beyond improvements in the socio-economic conditions in neighboring municipalities, for which potential effect is captured by the spatial filter (spatial lag)[24].

It is important to mention that the spatial effects suggested in this paper and their modeling structure are complementary explanations for the variation of infant mortality. The main predictors are socio-economic conditions and health services, the definitions of which are the theme of the next section.

4.2. Determinants of Infant Mortality Rate

In reduced form models of infant mortality the explanatory variables usually refer to socio-economic and geographic conditions, health structure, and health policies. Socio-economic conditions are related to the characteristics of the population and to the physical structure of each municipality that can affect

[24] Another advantage of the IV approach over the maximum likelihood (ML) estimation refers to the computational procedure. The ML estimation is interaction based and for large sample size it becomes very time consuming.

the risk of infant mortality. Income, education, sanitation and urbanization may affect the probability of infant survival in different ways.

Higher income is usually associated with better living conditions, general hygiene, and nutrition intake. Children in low income families are more likely to receive less than the necessary amount of nutrients to gain adequate weight during the first year of life. At the municipality level however, it is the increase in the proportion of children at risk of being underfed that can make the index of infant mortality increase. Therefore, the relationship between income and infant mortality can be misrepresented by the average income of the region when income distribution is very unequal. This research uses the level of unemployment rate in each municipality as the variable that represents income disparities and (health) productive capacity of the families. This variable is usually included in social deprivation indexes that are normally used as the main determinant of infant mortality[25].

Education is considered one of the most important socioeconomic determinants of infant mortality because it represents the technology of the health production function in families[26]. It expresses the reaction capacity of the families with respect to their socio-economic constraints. In addition, it can also be a proxy for cultural beliefs or preferences which stimulate demographic

---

[25] The Towsand deprivation index, for example, is composed of the standardized sum of unemployment rate, percentage of households owning no car, percentage of households not owner occupied, and the percentage of households with more than one person per room (Bithell et al, 1995).

[26] See Grossman, 1972.

transition of regions[27]. The educational variable in this research is represented by the illiteracy rate for woman between 15 and 49 years of age.

In less developed regions sanitation conditions are epidemiologically associated with infant mortality. Outbreak of diseases, water contamination, and pollution are problems induced by low quality sanitation. However, at the aggregate level of municipality the relationship between sanitation conditions and infant mortality has been dubious. Souza et al (1999), for example, find low explanatory power in sanitation conditions[28] when analyzing variations in infant mortality among the municipalities of Ceara State. On the other hand, Alves and Belluzo (2004) also working with municipality level data for the whole country and Simoes (1996), working with a broader geographic data for Brazil, show that regional differences in infant mortality can be explained mostly by differences in water and sanitation quality. This is currently the general view, although differences in measuring sanitation quality may affect this result.

In this research, sanitation conditions are represented by the percentage of households with adequate waste removal. In Ceara State, waste removal services are provided by each municipality. This variable, therefore, represents the efforts of each municipality in improving sanitation conditions. On average, less than 40% of the population of the municipalities has adequate garbage

[27] The demographic transition begins with the reduction in the level of population growth motivated by the realization that the capacity of survival of infants in a family is more correlated to the nourishment condition of the infants than to the number of infants in the family. That is, the chances of infant survival come with quality investments rather than quantity investments in the children.

[28] Souza et al (1999) used percentage of households without sewage connection or septic tanks as a measure for sanitation conditions. In a univariate ecological model better sanitation is a significant inhibitor of infant mortality. But, in multivariate models, the sanitation indicator becomes statistically insignificant.

removal services. The accumulation of garbage affects environmental quality and induces the proliferation of infectious diseases.

Demographic conditions for each municipality can also affect the risk of infant death. Living conditions in urban and rural areas are usually different. Infant mortality in rural areas is usually higher following harsher conditions related to deprivation of health facilities and economic development. The empirical evidence for Ceara State in Souza et al (1999), however, has shown that after controlling for health structure and socio-economic conditions, urban residence has no significant impact on infant mortality rate, with the sign of the coefficients shifting between positive and negative. A positive correlation between infant mortality and urbanization can be hypothesized by epidemiological concerns. In urban areas, the transmission of contagious diseases may spread faster, exposing proportionally more infants to the risk of contagion. In this case, it is possible that the degree of urbanization in the municipalities can elevate the level of infant mortality.

The health structure conditions are represented in this paper by the number of physicians and hospital beds per ten thousand individuals in each municipality. These variables capture variation in human and physical capital which together can provide adequate health services for the population of the municipalities. The health structure is complemented with the health policy variable that accounts for the number of community health workers per each group of ten thousand individuals. This variable represents therefore the quality and not the presence of the community health worker program in each municipality given that all municipalities were engaged in the program by 1994. It is therefore a measure of intensity rather than simple availability of the health policy in each municipality. However, as community health workers are responsible for the coverage of different and pre-determined areas within the municipalities, more CHWs also represent expansive coverage for families not covered before. Therefore, they also represent availability for the additional (marginal) covered families/areas.

Besides socio-economic conditions and the availability of health services, geographical location is also a potential predictor of the risk of infant death. Currie and Gruber (1997) for example, verified that distance is an important variable affecting the utilization of health services. In addition, distance is a key variable defining the probability of survival of the individuals in situations of emergency. The distance from the municipality of reference to the capital of Ceara State (the municipality of Fortaleza) was added to the models because Fortaleza contains most of the neonatal emergency services. It can also capture climate variation since more arid conditions are found in the countryside[29]. Another geographical variable is a dummy variable for municipalities that are bordered by municipalities in another state. The inclusion of this variable aims to adjust the spatial mismatch of the sample selection. That is, the borders of the State are not closed with respect to the distribution of infant mortality, and

[29] Fortaleza is located on the coast of the Atlantic Ocean with more amenable conditions than the countryside of the State which is mostly characterized by semi-arid conditions with shortage of water for some regions. Figure 2 shows the location of Fortaleza in Ceara State. Whereas Fortaleza presents an average rain precipitation level of 1,378mm/year, the average for all municipalities is 927mm/year.

bordering municipalities can also be affected by spatial effects coming from municipalities in different states.

The descriptive statistics of all variables included in the empirical analysis together with their data sources and definitions are shown in Table 1. The socio-economic characteristics show the usual regularities associated with less developed regions such as precarious sanitation (only 38% of families with adequate waste removal), high unemployment rate (15% average), and high illiteracy rate among women (28%). The CHWP presents a municipal average of 1 community health worker for each group of 500 individuals which is below the recommendation of 1 for each 225 individuals.

CHAPTER 5

EMPIRICAL MODELS AND SPATIAL MATRIX

This section describes the procedures used to estimate empirical ecological models. Two analyses are important at this point: the choice of model specification form and choice of the spatial arrangements (spatial matrices) used to define neighborhoods.

5.1. Model Specification

Ecological models of infant mortality vary considerably, not only with respect to their functional form but also to their time frame. In this regard, it is not unprecedented that data limitations for small regions may be an important factor in choosing the best strategy of analysis. The following discussion clarifies how data constraints played a role in this process and what strategies are used in this essay to adequately estimate marginal effects for health policy variables.

5.1.1. Functional form

The source of data to calculate infant mortality rate is the community health worker program. Compared to the official statistics given by death and birth certificates, the data collected by the program is less subject to undercounting problems because it is directly collected by community health workers (Svitone, 2000)[30]. This personal identification of cases also gives more credibility to the data as historical information is attached to each case.

At the municipal level, the index of infant mortality is calculated using a three year accumulation of live births (denominator) and infant deaths under 1 year of age (numerator). This procedure has been used in the literature because one additional death or live birth in small communities may cause the index to vary considerably, given that the ratio of deaths per live births is multiplied by 1000 to produce the index[31]. This type of data generates heteroskedastic problems where small municipalities tend to present higher variance than more populated municipalities.

Using three consecutive years, only a few municipalities (26 out of 184) presented a number of live births less than 500, which is a bottom limit suggested by local government[32] to calculate the index of infant mortality for small regions. In this paper, however, these municipalities are not excluded given that their exclusion would compromise the overall spillover effect that involves all contiguous locations (municipalities) of a determined region (state).

After accumulating three consecutive years of live births and infant deaths, the municipal index of infant mortality rate is then converted to its square root

[30] In very poor regions the undercounting of deaths and live births can distort the real risk of infant death when proportionally more deaths than live births are counted in the infant mortality index. If these uncounted cases were added, the index of infant mortality would change considerably. Given the community health worker program also counts non-hospital registered live births and deaths, the miscounting cases are only those not covered by the program and not registered. The use of the CHWP data information to calculate infant mortality rate is also present in Souza et al (1999).

[31] See Corman and Grossman (1985), for example.

[32] The Health Secretariat of Ceara cites the World Health Organization in prescribing carefulness when calculating infant mortality rate for small municipalities.

value. This conversion serves mostly two purposes: it suggests a non-linear functional form where the marginal effect of the covariates increases with the level of infant mortality, and it usually produces a distribution for equation residuals that is closer to the Gaussian distribution than when the level or other transformations are used (log values, for example)[33]. Given that the spatial correlation diagnostic tests in this essay rely on the assumption of normal distribution of residuals, square root transformation provides more robustness to these tests.

The nonlinear functional form for infant mortality rate suggests that municipalities with high levels of infant mortality have more room for improvement and consequently have comparatively higher marginal effects[34]. Although this functional form may alter productivity levels per municipality it does not compromise the hypothesis raised in this essay that health policy productivity can be undervalued in traditional reduced form models if policy spillover effects are neglected. This is because the contrasting models (with and without spillover effects) have to be compared under the same model specification (at level or square root transformation). This essay follows Leal and Szcwarcwald (1997) who also tests spatial correlation for infant mortality rate with the square root transformation. Therefore, although it is not common in the literature, this

[33] Level and log models will also be estimated to check the robustness of the results.
[34] With square roots conversion, the marginal effect for the index of infant mortality is given by: $\partial IMR / \partial X = 2\beta \times IMR^{1/2}$, where β is the coefficient of the covariate X, and IMR is the index of infant mortality in one particular municipality. It is assumed, therefore that the higher the infant mortality rate higher will be the marginal effect. This assumption says that municipalities in worse situations have more room for improvement and consequently can reduce faster their levels of infant mortality.

transformation brings spatial modeling benefits (more robust diagnostic tests for spatial correlation) and does not conflict with any reasonable assumption for infant mortality models[35].

5.1.2. Patterns of time modeling

The choice of the period of analysis was driven by data availability and quality. The cross section analysis regresses infant mortality rate for the 1998/2000 period against health related variables of the same period and socio-economic variables collected from the Brazilian Census of 2000. This contemporaneous regression is the only available alternative to substitute models with lagged values for the explanatory variables. In the latter, the marginal effect of policy variables in health outcomes is felt with a time lag in which the population adapts to and acknowledges the new policy. However, lagged values for the endogenous (IMR) and policy variables are not available as the CHWP only became more effective in data management after 1994, and as 5 new municipalities had no prior information because they were created (became independent jurisdictions) during the 1990s. Although lagged values for the explanatory variables could be extracted from the Brazilian Census of 1991 and extrapolated for the newly created municipalities, the dependent variable (IMR), the policy variable of interest (CHW) and the other health related variables (Physicians and Hospital Beds) could not be calculated consistently for earlier years[36]. This data limitation therefore inhibited the use of difference or fixed effect

[35] Models in levels and logs are also estimated for robustness check.

[36] In estimating the IMR for the municipalities of Ceara state for the 1994-1996 period, Souza et al (1999) included in the sample only those municipalities (140) which provided complete

model specifications which could also provide consistent marginal effects for the policy variables.

Despite this modeling constraint imposed by data limitation, the contemporaneous regression proposed here serves well the purpose of analyzing how ignoring spatial externalities in traditional reduced form models may undervalue the efficacy of health policies. This efficiency effect, however, may also be biased in contemporaneous equations if reverse causality problems exist.

5.1.3. Policy Endogeneity

Policy endogeneity occurs when the intensity of the health policy is induced by poor health outcomes causing a downward or upward bias on the estimates of policy variables. If there is a positive correlation between contemporary policy variables and the lagged value of infant mortality rate there is a downward bias on the estimates of marginal effects. This is because more community health workers would be hired in municipalities with high levels of infant mortality, and one would therefore find more CHWs where the situation is worse. On the other hand, when there is a significant negative correlation between the lagged infant mortality rate and the actual policy intensity (maybe due to municipal administration negligence or over-concern with respect to health policy) there should be an upward bias in the estimation of policy effects. The

---

information at that time. The inconsistency in the provision of the data for this earlier period occurs when some municipalities do not report any information regarding live births or infant deaths, or report information inconsistent with their history for one particular year. These problems were not observed for the 1998-2000 period.

traditional solution in this case is to include the lagged value of infant mortality rates as an additional explanatory variable.

On the other hand, this solution may also generate a serial correlation bias when the lagged variable is correlated to the (actual) error term. The sources of this correlation can be unmeasured health endowments at the municipal level not fully captured by explanatory variables. As these unmeasured resources also affect infant mortality in the early years it will also affect indirectly the actual level of infant mortality and could bias other coefficients.

Instrument variables for the time lagged infant mortality rate, however, are difficult to be estimated. This is especially due to the unavailability of suitable lagged values for explanatory variables. As a solution to the policy endogeneity problem this essay follows Corman and Grossman (1985) who report both models with and without time lagged variables for the infant mortality rate and interpret their marginal effects as bounded values. That is, these models present upper and lower values for marginal effects, and any policy simulation perspective can be analyzed at the average of these two. Although not very precise, this solution does not compromise the investigation of how traditional reduced form models may understate health policies and spillovers. These spillovers, however, depend on the spatial distribution of the policy.

5.2. Spatial Matrices

The distribution of infant mortality rates among the municipalities of Ceara State can be observed in Figure 3. It can be seen at first sight that neighbor municipalities are usually in the same distribution interval of infant mortality levels, which suggests that spatial correlation may be an issue for empirical models. The issue at hand, therefore, is to identify what type of spatial patterns for the definition of neighbors better fits the actual distribution.

In this paper four types of spatial patterns are used not only to identify spatial correlation possibilities (lag or error) but also to test the hypothesis of spatial spillover in health related variables: the simple contiguity weighting matrix (W1), the population-contiguity matrix (W2), the inverse distance matrix (W3), and population-inverse distance matrix (W4). Whereas matrices W1 and W3 are purely based on traditional geographic definitions of neighborhood, the matrices W2 and W4 add an economic content to these weighting matrices as more populated municipalities have a higher weight in calculating neighbor's average infant mortality rate and policy.

The first matrix (W1) identifies related municipalities only by the sharing of common borders. Before row standardization, $W1_{ij} = 1$ if municipalities j and i share a common border, and 0 otherwise. Therefore, spatial correlation occurs between infant mortality in one municipality and the weighted sum of infant mortality in its contiguous (immediate) neighbors. Among the neighboring municipalities, their weights are equal in calculating the neighbor's infant mortality rate or availability of health services.

In the second matrix (W2), neighbors are also defined by common borders, but the weights for each neighbor municipality are given by their proportional size (population). Before row standardization, $W2_{ij}$ = population of j if municipalities j and i share a common border, and 0 otherwise. The assumption here is that more populous (neighboring) municipalities have stronger impact on neighboring conditions, and should therefore receive a higher weight when counted as a neighbor. As spillovers in health policies can be driven by behavioral influence between municipalities, it is reasonable to suggest that the direction of this influence is more prominent from large to small municipalities.

The third weighting matrix (W3) assumes that all municipalities of the state are directly correlated. But the extension of this correlation depends on the proximity of the municipalities. Before row standardization $W3_{ij}$ = 1/distance between municipalities j and i. The decay distance matrix is used in spatial econometrics under the assumption that the shorter the distance the higher the influence of one municipality on another.

The fourth matrix (W4) is the combination of the distance decay effect and the population size effect. That is, before row standardization, $W4_{ij}$ = (population size of municipality j/distance between municipalities j and i)[37]. In this case more importance is given to large municipalities which receive a regular inflow of patients from other municipalities located all around the state. Municipalities such as Fortaleza or Juazeiro do Norte provide specialized services (neonatal units,

[37] Two other spatial matrices were also tested in earlier stages of this essay, with their weights based on the number of delivery services provided for non-resident mothers. The undesirable endogeneity of the weights in this case, however, were essential for their displacement in this essay.

for example) not found in other municipalities. Therefore, they should have a higher weight even for municipalities located relatively distant from them.

The evidence that policy spillover effects depend on spatial patterns defined by the spatial weight matrices is by itself a source of information that can be used to analyze state distributional policies among the municipalities. If the health spillover variable (W^hH) is significant it is possible to infer a per-municipality marginal effect that will depend directly on the participation (the weights) that the availability of health services in one municipality will have in reducing infant mortality rate on all its neighbors ($\Sigma_i \delta W_{ij}$).

The hypothesis of simple spatial correlation in infant mortality can be tested using the Moran's I index which is a correlation coefficient between infant mortality in the municipalities (IMR_i) and in their neighbors ($W_{ij}IMR_j$). For all weighting matrices specified above, infant mortality presents significant spatial correlation as shown in Table 2. However, this significant correlation (or spatial auto-correlation) tends to disappear when we add the predictors of infant mortality to the model. This is especially true if the socio-economic and health related variables are also spatially correlated as shown in Table 2. That is, the simple spatial correlation is detected because the neighboring municipalities present similar socio-economic, health or geographic conditions. Once we control for these factors, the spatial correlation tends to be insignificant. When the spatial correlation is still present in infant mortality models, some (spatial) adjustments must be made to the model in order to avoid imprecision and/or the bias of the estimated coefficients.

CHAPTER 6

SPATIAL METHODOLOGY

This essay follows traditional spatial econometrics methodology, where first reduced form models without spatial effects are estimated by ordinary least square (OLS) in order to test for significant presence of spatial correlation (Florax et al, 2003). That is, theoretical model 1 is estimated and diagnostic tests for spatial correlation are used to verify the presence and type of spatial correlation[38]. There are many diagnostic tests used in the literature for different types of spatial models.

This essay uses four diagnostic tests based on the Lagrange Multiplier principle, which, combined, provide a decision rule for alternative spatial models. This guidance to alternative modeling is an advantage of the Lagrange Multiplier tests over the Moran's I index, for example, which is very powerful in capturing spatial error models but is comparatively less powerful at detecting spatial lag specifications (Anselin and Florax, (1995)). Two of the tests are designed to capture spatial error models as alternatives to the null-hypothesis of no-spatial correlation (LM-ERR and LM-EL), and two are directed to spatial lag models (LM-

[38] The models are estimated with and without the time lagged value (1994-1996 period) for infant mortality rates as an additional explanatory variable. These are the traditional reduced form models for which comparative analysis will be made with respect to models including spatial effects.

LAG and LM-LE)[39]. These statistics have good asymptotic properties and have also proven to perform well in small samples. Based on a robustness check from monte-carlo simulations Anselin and Florax (1995) observed that:

> "*The robust LM-EL and LM-LE tests performed remarkably well against one-directional alternative, which suggests that they may be usefully combined with the LM-ERR and LM-LAG tests to indicate which of the two forms of dependence (lag or error) is the proper alternative. In other words, this may result in an augmented decision rule,(...): when LM-LAG is more significant than LM-ERR and LM-LE is significant while LM-EL is not, a lag dependence is the likely alternative; and when LM-ERR is more significant than LM-LAG and LM-EL is significant while LM-LE is not, an error dependence is the likely alternative*" (Anselin and Florax 1995, p. 48).

In addition to this rule (which is not always the case for all data sets), two additional criteria may be used. The first is simply to verify the significance of the spatial correlation coefficients for lag or error (ρ – in spatial lag model or λ – in spatial error model) in their respective models, and the second is to also test them but using the other spatial model specification (ρ – in spatial error model or λ – in spatial lag model). That is, the second test analyzes if there are any additional form of spatial correlation (lag or error) when spatial lag or spatial error models are estimated (Bruckner, 1998). The better specified spatial model is that with insignificant spatial correlation.

[39] See Appendix B for more information on the statistics and their distributions and sources.

Different from the traditional spatial methodology, however, the first spatial adjustment to be made in this essay in case of significant spatial correlation is the estimation of health policy spillover models (model 2). These models are estimated by traditional methods (OLS) as only exogenous variables are present on the right hand side. The objective is to verify if the spillover variables can explain and control most (if not all) of the spatial correlation found. If spatial correlation still persists, however, spatial lag or spatial error models are estimated following the suggestion provided by the Lagrange Multiplier diagnostic tests and the other two complementary rules.

The estimation of spatial models, however, requires different methods of estimation. Instrumental variable (IV) estimations are used as they provide more robust estimation especially for spatial lag models (Kelejian and Prucha, 1997). As was said before, when compared to maximum likelihood estimation, the IV procedure has comparative advantages with respect to non-parametric assumptions and robustness to model specification with spatial correlated errors.

All types of spatial patterns are used in order to diminish the possibility of spatial model misspecification while searching for the models with the best adjustments. The Akaike Information Criterion[40] (AIC) indicator will be the reference guide for cross model comparison in terms of adjustment[41]. Additional criteria for model selection are the absence of spatial correlation in health spillover models, and in case of spatial lag models (model 5), their spatial

[40] AIC = -2L + 2K, where L is the maximized log likelihood function and K is the number of variables. See Anselin (1992).
[41] Different from traditional measures of fit such as R^2 (or adjusted R^2) the AIC measure allows comparing models with different number of explanatory variables and spatial effects.

correlation coefficients (ρ) needs to be less than one. When spatial lag models are estimated through 2sls it is possible to find ρ greater than one, which is a clear indicator of model misspecification. This is because the spatial effect captured by this coefficient does not diminish when transmitted successively from municipality to municipality and the model becomes unstable[42]. That is, any marginal effect of policies tends to infinity. A ρ less then one is a required assumption to produce consistent estimative based on 2sls procedure (Kelejian and Prucha, 1997).

Once the models fit the criteria suggested above their estimates can be used for inferences and policy simulations. In order to make measurement comparisons, however, this essay follows Corman and Grossman (1985) in estimating average marginal effects for models with and without initial conditions (time lagged infant mortality rates). This is because they have higher and lower bounds estimates for marginal effects as they represent a trade-off between policy endogeneity problems (models without time lag) and serial correlation problems (models with time lag).

[42] It is the spatial version of an explosive time series model.

CHAPTER 7

EMPIRICAL RESULTS

This section follows the methodology suggested above, ordering the empirical analysis sequentially by: i) searching for model specifications with and without spatial effects (model 1 versus model 2 versus model 4 or 5), ii) interpretation of marginal effects found for the models of reference, iii) sensitivity analysis, and iv) policy simulations.

7.1. Model Specification

Traditional reduced form models of infant mortality rate (without spatial effects) are the starting point to search for models of reference, which should present the best fit (lowest AIC) and no spatial correlation. Table 3 shows traditional models without spatial effects but with (3.1) and without (3.2) time lag for infant mortality rates. In both models the presence of spatial correlation is significant for all types of spatial matrices, especially when the alternative model specifications are spatial lag models. In this case any inference from socio-economic or health related variables from these traditional models would be biased.

The next step is to identify if health spillover possibilities are significant and what general type of spatial pattern would describe these cross-bordering benefits from health services. Table 4 presents spillover models for all types of spatial matrices. The inclusion of spillover possibilities from health related variables seems to be adequate considering the spatial patterns defined by W1 (model 4.1) and W2 (model 4.2). This is because not only the models are comparatively more adjusted (lower AIC value) but also a jointly significant test for the coefficients of the spillover variables rejects the null hypothesis that they are all equal to zero. More interestingly, after controlling for spillover possibilities spatial correlation becomes insignificant for these two types of spatial patterns. In addition, a jointly significant test for the variables related to the policy program (CHW and W*CHW) and to the availability of physicians (Physicians and W*Physicians) attested the significance of the spillover effect for these variables. This is to say that improvements in the policy in one municipality will affect infant mortality rate not only in this municipality but also in neighboring municipalities.

In order to test alternative model specifications spatial lag models were also estimated with (Table 5) and without (Table 6) policy spillover variables[43]. The results form Table 5 suggest robustness for the health spillover models when spatial patterns defined by W1 (model 5.1) and W2 (model 5.2) are the reference. Neither the spatial lag variables for these models were significant nor were the adjustment (AIC) of the models improved[44]. Models 5.3 and 5.4 show significant spatial lag variables, but their correlation coefficients (ρ) were too high,

[43] The spatial lag specification was used as their diagnostic spatial correlation tests were more significant.
[44] The same also occurred for spatial error models estimated through maximum likelihood procedure.

indicating misspecification problems for these types of spatial patterns (W3 and W4).

The robustness of the spillover models for W1 and W2 are evident when other forms of spatial models are also estimated and when the time lag is included in the model. Table 6 brings spatial lag models without possibilities of spillovers. Again, the spatial correlation coefficients for spatial lag variables based on distance (W3 and W4) are inconsistent with modeling selection requirements, and the absence of spillover effects reduced expressively the adjustment of the models with spatial patterns based on simple contiguity (model 6.1 and model 6.2). When time lag variables are included in the analysis (Tables 7 and 8) spillover models (models 7.1 and 7.2) based on spatial patterns W1 and W2 are still the models attending the selection criteria. Following Corman and Grossman (1985), therefore, this essay uses averages of marginal effects for model specifications with and without time lag to generate quantitative predictions. There are, however, two spatial patterns of references in spillover models presenting similar levels of adjustments (AIC). In this case, marginal effects are computed not only as average effects for models with and without time lag for the same spatial pattern, but also as a range of values between each different type of spatial matrix.

Table 9, therefore, presents the models for which the average marginal effects are calculated for each spatial pattern of reference: models 9.1 and 9.2 (W1), and models 9.3 and 9.4 (W2). These average marginal effects form the range of possible values to infer predictions for changes in policy variables.

7.2. Marginal effects

This section identifies the main predictors of infant mortality by quantifying their marginal effects as shown in Table 9. It is evident that socio-economic variables are significant predictors of infant mortality rates. With exception of illiteracy rate, this qualitative evidence seems to be robust with respect to model specification regarding spatial effects or initial conditions (lagged infant mortality rates). Significant marginal effects remain across the models of Table 9 for unemployment, sanitation, and urbanization variables.

As expected, unemployment rate has a positive correlation with infant mortality rate. For an average municipality initially experiencing an infant mortality rate equal to 36 deaths per 1000 births[45], the effect of a 10% point reduction in unemployment rate would produce a predicted change ranging from 4.22 (average of models 9.3 and 9.4) to 4.47 (average of models 9.1 and 9.2) of avoided deaths per 1000 live births. It is interesting to note that in traditional models without spatial spillover (models 3.1 and 3.2), the average marginal effect is comparatively lower (3.75 infant deaths avoided per group of 1000 live births).

Improvements in sanitation conditions can also be an effective policy to reduce infant mortality. For the same average municipality suggested above, a 10 percent increase in proportion of homes with adequate waste removal would reduce the mortality probability from 2.24 to 2.36 infants per groups of 1000 live

[45] This is the approximate municipal average value for infant mortality rate. As stated before, marginal effects in square root models are positively related the actual level of infant mortality rates in each municipality.

births. As sanitation conditions are still very poor among the municipalities, there is much room for improvement. With respect to unemployment rate, policies targeting full employment would be comparatively more limited at reducing infant mortality rates. This is because average levels of unemployment rates for the municipalities are not very far away from zero. This is consistent with the hypothesis that there is a limit for reduction in the level of infant mortality rates when only economic, social or health policies are triggered.

Although the illiteracy rate had a significant impact on the infant mortality rate in traditional models (3.1 and 3.2) it was not significant for the spillover models of reference (Table 9). Misspecifications in the educational variable could be playing a role in explaining the result above. Formal education may not be the best reference for analyzing the capacity of families to understand the behavioral risks that lead to changes in the survival probability of their infants. That is, formal education is not necessarily the mechanism that induces the acknowledgement that water filtering and breast-feeding are important behavioral conditions to reduce the chance of infant death. In this case, the health policy spillover of the CHWP can already capture a great part of this learning behavior (health productive booster) given it is primarily an educational health care program[46].

The positive relationship between urbanization and infant mortality is to some extent puzzling given that urbanization represents for most cases a sign of development. However, once we control for other socio-economic conditions that

[46] The Pearson correlation coefficient between neighbor's illiteracy rate and neighbor's community health workers availability is 0.59.

also characterize development, the urban effect can be ambiguous. It may represent stress conditions or even potential facilities for in-municipal diffusion of contagious diseases as well as better informational access for health programs (immunization programs, for example). Whereas the first two effects cited above assume a positive correlation with infant mortality, the latter suggests a negative one. According to the estimations provided in Table 9, the rationale of the first effects seems to be dominant. In addition, the beneficial side of urban effects may also be controlled by the municipal location variable.

The empirical results show that location is important to define the survival probability of infants. The state capital (Fortaleza) provides most of the specialized health services for infants, including exclusive emergency services. Travel time in these cases can directly affect the probability of survival of infants. Following the estimates (Table 9) the risk of infant death is expected to increase from 3.36 to 3.42 deaths per thousand live births for each 100 km (62.5 miles) from the capital[47]. This result suggests policies improving accessibility to the capital of the state (transportation facilities) and the decentralization of specialized health services as additional measures to reduce the risk of infant death.

If socio-economic and geographic conditions prove to be robust enough to model specifications, as they allow a straightforward cross-model comparison of their marginal effects, the same can not be said with respect to variables related to health structure and police. This is however expected as the availability of

[47] This simulation is also made for an average municipality previously experiencing a level of infant mortality rate equal to 36.

physicians and community health workers within the municipality and in neighboring regions are significantly correlated, as the Moran's I index in Table 2 indicates. Thus, bearing in mind that there is spatial correlation among health related variables, some interesting results appear when comparing models with and without spatial effects.

It seems that in traditional models (3.1 and 3.2) the health structure variables (hospital beds and physicians per 10,000 individuals) have no effect on infant mortality rate. Once the models allow for health spillover (Table 9), neighbors' availability of physicians (W*Physicians) becomes very significant, although the in-municipal availability (Physicians) continues to be insignificant. A joint significance test for these variables (Table 9) suggests that it is regional rather than local availability of physicians that can affect the risk of infant mortality. That is, it takes a group of physicians providing services in the surrounding municipalities to affect significantly the level of infant mortality rates. In this case, there is a policy mismatch between regional availability measures and local (municipal) risk of infant death. When this mismatch is accounted for with the introduction of spillover variables, the relationship between physician services availability and infant mortality is properly modeled. Therefore, the modeling of spillover effects in this case makes a significant difference in evaluating the effectiveness of municipal health structures for reducing infant mortality.

The spillover effect of physicians is consistent with both types of spatial patterns (W1 and W2) and time modeling. In order to analyze the total effectiveness of increasing the availability of physicians for one average municipality, however, it is necessary to consider the average weights (spatial weights) for neighboring municipalities and the average number of neighbors benefiting from more physicians around. As these two are respectively 0.195 and 5, an additional physician per group of ten thousand individuals in one municipality is expected to reduce infant mortality rate from 0.54 to 0.62 infant deaths per thousand live births. This effect is very substantial, especially when compared to traditional models without spatial effects, which indicated ineffectiveness for the availability of physicians. For policy purposes, therefore, this empirical result suggests that physicians are more productive where regional (pool of municipalities) rather than municipal level of infant mortality is higher.

Regarding the effectiveness of the CHWP, spillover effects are also significant, suggesting again a mismatch between where the policy makes itself locally available and the geographical dimension of the phenomenon being explained (municipal infant mortality rate). The joint significance test for related variables (CHW and W*CHW) in Table 9 shows that the benefits of improving the program in one municipality spread to neighboring municipalities especially when W2 is the spatial pattern defining neighbor's weights. In this case, for each CHW added per group of 10,000 individuals, there is an expected reduction ranging from 0.38 to 0.65 in the number of infant deaths per thousand live births. Without spillover possibilities (Table 3) the average marginal effect of the policy program reduces to 0.27 infant deaths per thousand live births for an average municipality.

Another advantage of models with spillover effects is the possibility of a case by case estimation of marginal effects for each municipality[48]. These municipal marginal effects depend on the spatial matrix of the spillover policy variables, which assigns the weights that each municipality exerts on others. It is evident from Table 9 that W2 is the spatial pattern that describes more precisely the general conditions for the spillover of the CHWP. In this case, the spillover of the policy occurs more effectively from larger to smaller neighbor municipalities. This is consistent with the idea that informational spillovers inducing behavioral changes are more likely to flow from large to small municipalities as the former are more populated and usually represent role models to be followed. It is also possible that the preventive character of the health care program (CHWP) reduces more effectively the overcrowded demand for health services in public hospitals of larger municipalities. This improves the quality of these services and benefits all the populations from the neighboring municipalities. This form of spatial pattern for policy spillover therefore brings additional conjectures for distributional policies.

As larger municipalities (in size and population) have higher weights in defining neighborhood average availability of health services, they tend to present higher productivity levels for the policy program. This productivity, however, also depends directly on other factors such as the prevailing level of infant mortality rate for the municipality of reference and for its neighbors. In general, therefore, allocation policies would be more effective, in the case of

[48] This case by case estimation of marginal effects is independent of the functional form assumed for the infant mortality rate model. Even in linear models municipal specific marginal effects can be differentiated given municipalities have different weights on regional averages.

Ceara state, if more CHWs were added to clusters of relatively larger (size and population) municipalities presenting high levels of infant mortality. Without entering into the fairness issues, as health policies are not always guided by efficiency, this essay shows that traditional reduced form models can undervalue the effectiveness of health policies when they ignore significant policy spillover effects. This evidence is also illustrated by the policy simulations in the section ahead. However, as the simulations depend on model specifications, the next section investigates if the results found stand for different functional forms.

7.3. Sensitivity Analysis

As the square root transformation of infant mortality is not in common use, the spatial effects found could be unique to that model specification. The spatial methodology described above was also followed for level models of infant mortality (Tables 10 and 11). As in the case for square root transformation, spatial spillover models with spatial patterns defined by W1 and W2 are the models of reference fitting all the criteria of selection for inferences. Table 11 is the level version of the reference models of Table 9. With respect to the policy variable (CHW), spillover effects are significant when the spatial pattern W2 is used (models 11.3 and 11.4). This finding reinforces the idea that the effectiveness of the program is comparatively higher in regions more populous than others. This conclusion therefore applies independently of the functional forms of the infant mortality models[49].

[49] Log models were also estimated with similar results.

The level version models, however, present residuals for which the Jarque-Bera statistical test clearly rejects the assumption of a normal distribution (Table 10). As this assumption is important for spatial correlation diagnostic tests[50] and for efficiency in inferences with finite samples, the square root model of infant mortality is preferred therefore for policy simulation exercises. This is because the assumption of normal distribution for the residuals was not rejected for this type of functional form (Table 3).

7.4. Policy Simulations with Spatial Effects

This essay proposes a spatial modeling approach to estimate reduced form models for small contiguous bordering areas. When spillover effects are significant, policy prescriptions may change when compared to what would be prescribed if traditional models without spatial effects were the reference. Health care programs may look less effective and allocation policies could be distorted if the policy guide were from models without spillover possibilities. This section provides a policy simulation to illustrate these points.

The simulation compares effectiveness in predicted reductions of infant deaths or risk of infant deaths (IMR) when each municipality receives one additional community health worker for each group of 10,000 individuals. This hypothesized policy follows the general prescription of the program where each community health worker covers a certain number of families living next to him/her. Although each municipality receives a different number of additional

[50] Anselin (1992).

community health workers - as they have different population sizes[51] - the municipal productivities can be compared as they count child deaths avoided (or IMR reduced) per community health worker added. These policies are evaluated in models with (Table 9) and without (Table 3) the spillover of the health care program[52]. The objective is to estimate how average state and municipal policy effectiveness changes once spillovers are considered. It is important to emphasize that the measure of policy effectiveness suggested here (infant life saved per CHW added) values infant life saved equally across municipalities. That is, all municipalities have the same weight when the average level of policy effectiveness for the whole state is calculated[53].

Table 12 provides the predicted effectiveness of the program[54]. For the state as a whole, the simulated policy involves distributing approximately 743 more community health workers across all the municipalities, and this will result in 1858 predicted child deaths avoided in models without spillover (effectiveness rate of 2.502). In models with spillover the range of predicted child deaths avoided goes from 1790 (W1) to 2594 (W2) which generates effectiveness rates ranging from 2.410 to 3.485. This result shows that neglecting possibilities of

[51] The number of additional community health workers is equal to the population size divided by 10,000. A municipality such as Fortaleza with 2,140 thousand individuals, for example, will receive approximately 214 community health workers.

[52] The predicted productivity without spillover is based on the average of the estimated effectiveness of models 3.1 and 3.2. In models with spillover the averages of models 9.1 and 9.2 and models 9.3 and 9.4 are calculated. This averaging strategy follows Corman and Grossman (1985) where marginal effects from models with and without initial conditions are averaged for prediction purposes. The marginal effects for the spillover models with different spatial patterns are presented as a range of values.

[53] This means that no (political) relative preference is given to life saved in any particular municipalities.

[54] The process to calculate the predicted productivity is addressed in the Appendix A of this essay.

spillovers in traditional models may compromise the effectiveness of the program by up to 39%. This difference in the valuation of the program can be considered crucial for the whole society as this extra benefit of the program is measured in fewer infant deaths. There is therefore an immense social cost to be paid when policy makers are misguided by traditional reduced form models neglecting the possibility of policy spillover.

The municipal comparison in models with (9.3 and 9.4) and without (3.1 and 3.2) spillovers is illustrated in Figures 4 and 5. In Figure 4 the municipal productivity intervals are based on the quintiles of the productivity distribution when spillover effects are ignored. Figure 5 uses the same distribution intervals as in Figure 4 but they are applied for the municipal productivity in models with significant policy spillover. It is evident that the municipal productivity increased considerably for most of the municipalities as many of them are now in the highest quintile of the earlier (no spillover) distribution (the darkest areas)[55].

When productivity is measured in unit reductions in the risk of infant deaths (IMR) per community health worker the comparative analysis represents even a higher difference in models with and without policy spillover[56]. Two interesting study cases refer to the municipalities of Fortaleza and Salitre. Without spillover these two would represent the municipalities with the lowest and highest productivity respectively, in terms of IMR reduction per community

[55] Similar effectiveness measures are found for models in levels of infant mortality. Without spillover the state average effectiveness of the policy is 2.77 infant deaths avoided per community health worker added. In models with spillover the range goes from 2.45 to 3.78 infant deaths avoided per community health worker added.
[56] Differences in productivity results are due to the weight that the number of live births has on each of the productivity measures. More details are shown in the Appendix A.

health worker added (Panel B of Table 12). This is because they would have the highest and the lowest predicted level of infant mortality before the policy is introduced. In models without spillover the relationship between productivity and IMR is directly linked to the functional form of the model and is illustrated in Graph 1. The higher the infant mortality, the higher the effectiveness of the program in reducing the risk of infant death.

Once the spillover effects are considered, effectiveness in the municipality of Fortaleza increases by up to 8.8 times (becoming the 4th highest) whereas the productivity for the municipality of Salitre increases comparatively much less (79%), becoming only the 73rd highest among all municipalities. This is because effectiveness when spillovers are significant depends not only on the previous infant mortality rate but also on other factors such as infant mortality in neighboring municipalities and the weights the municipalities have on their neighbors. Graph 2 illustrates how the relationship between policy effectiveness and IMR changes once the spillovers are significant.

The huge increase in effectiveness of the policy in Fortaleza is basically due to the high spillover effect, which is a function of the population size modeled through the spatial spillover matrix (W2). As Fortaleza is the most populated municipality, its total spillover effect[57] is expected to be high, which more than compensates for its low in-municipality productivity. Its mechanism of transmission can be related to cross-border reinforcement in the health educational purpose of the program, or to improvements in the quality of the

[57] Total spillover effect means the sum of the reductions in IMR in the neighbors of Fortaleza once the program increases by 1 additional agent per group of 10000 people.

shared public health services. The latter may be induced by the preventive character of the program, which may reduce the crowding of the intra and inter municipal demand for those services. Given that the municipality of Salitre has a relatively small population size, its spillover effect is lower and consequently, its statewide effectiveness will not be boosted considerably.

Another important factor contributing to increased productivity once spillovers are allowed is the average level of infant mortality in neighboring regions. Municipalities with the same population share in their neighborhood regions present different policy productivities as their neighboring municipalities present higher levels of infant mortality. In this case, productivity is higher where neighboring municipalities present the highest level of infant mortality. This result however is conditioned by the model specification used here. In models at the level value for infant mortality rate the marginal effect is independent of the previous municipal values of infant mortality rate. In this case (level model), it is only the spatial spillover pattern (W2) that defines comparatively higher effectiveness measures for the policy. As this essay focuses more on comparisons between models with and without spillover effects rather than comparisons for different functional forms, it remains cautious with respect to possible policy prescriptions conditioned on the functional form of the IMR model.

Therefore, the relative size of the population affects the policy productivity of the community health worker program (CHWP) because the spillover effects are significant and determined by W2. This interesting finding suggests a distributional policy for Ceara State where more populated municipalities would be relatively more efficient (statewide) than their neighbors in reducing the regional risk of infant mortality. That is, among neighboring municipalities, improvements in the CHW program are expected to reduce relatively more the (statewide) level of infant mortality rate if implemented in large rather than small municipalities. This is because these larger municipalities have greater spillover effects which can be induced by preventive behavioral change or improvements in the quality of shared health services.

Another important conclusion from the simulation above is that if the policy program is more efficient than suggested by traditional reduced form models of infant mortality, then a simple benefit/cost analysis may lead to an inefficient (and reduced) distribution of available resources toward this policy program. In (level) models without spillover effects (10.1 and 10.2) the effect of increasing the program by one CHW per 10,000 individuals has the same effect in reducing the risk of infant mortality as an increase of 0.89 percent point in the number of families with adequate waste removal[58]. In models with spillover effects (11.3 and 11.4) this equivalence increases to 2.71 percent point, which could represent a very significant difference when associated costs are attached to each policy alternative[59]. On the other hand, when compared to physicians, the relative effectiveness of the CHW goes from 2.06 to 0.978 in models without and with

[58] In models with spillover the marginal effect for the spillover variable (W2*CHW) is computed for a municipality with 5 neighbors and spatial weights equal to 0.195. These are respectively the average number of neighbors per municipality and the average weight per neighbor.

[59] This huge difference occurs due to the joint effect of gains in the effectiveness of the policy and reduction in the marginal effect of the sanitation indicator. The assumption of 5 neighbors with equal weights of 0.195 also contributes to this difference. Assuming only 2 neighbors (with the same weights), for example, the equivalence in the reduction of IMR from sanitation would be equal to 1.15 point reduction in models with spillover.

spillovers respectively. This is because the physician indicator also has a very strong spillover effect[60]. Therefore, under-investments (or over-investments) in this program can be induced by erroneous traditional models without spillover possibilities. The direct consequences of insufficient (or inefficient) investments may represent differences in infant lives saved which are invaluable for the suffering families.

In addition, the program may also have similar effects on other health indicators as it targets the family as a whole. In this case, a number of health outcomes and especially those considered avoidable with primary care (incidence of infant diarrhea, for example), may also be misrepresented by models ignoring spillover possibilities. This methodological error can also lead to under investments in this effective program.

[60] When the spillover model is estimated with spatial matrix W1, the spillover effect for physicians is even higher.

CHAPTER 8

CONCLUSIONS AND FURTHER DISCUSSIONS

This paper analyzes how spatial effects may be included in reduced form infant mortality models when contiguous regions are included in the data set. In particular this paper shows that health spillover effects may be very important in evaluating the effectiveness of health policies. Their exclusion in traditional reduced form models can underestimate considerably the effectiveness of health policies and, consequently, diminish additional life-saved-benefits when under-investments in these policies are made.

The empirical analysis investigating variations in infant mortality rate among the municipalities of Ceara State also shows that unemployment, sanitation conditions, urbanization, municipal location and health services availability play an important role in explaining these variations. The relatively low marginal effect of each factor separately suggests that significant reductions in the level of infant mortality rate require multi-sector policies improving jointly socio-economic and health service conditions.

Although this general conclusion is robust with respect to important sensitivity analyses, data limitations and restricted modeling perspectives leave opportunities for complementary future research. First, the timing of the spillover effects may not be contemporaneous with respect to reductions in infant mortality rate. That is, the behavioral change induced by the primary health policy may take a while to also be adopted (or reinforced) in neighboring municipalities. Therefore, different timing patterns (time lagged, for example) for the spillover effect may also be tested in future studies when the program will be more developed, and the data set will be available.

Second, the presence and magnitude of health spillover effects, for example, are shown to vary with respect to the specification of the spatial weighting matrix. Additional patterns of spatial neighborhoods can be used to capture other general transmissions of health service benefits across municipalities. Spatial patterns of special interest are those pre-established micro-regions (set of municipalities) defined by (similar) geographic conditions, or those defined by inter-municipal migration for health services. The same can be attempted for the identification of spatial lag models. In fact, the value for the coefficient greater than one for the spatial pattern based on distance reinforces the necessity to hypothesize different specifications for the spatial structure.

Finally, other health outcomes may also be similarly affected by the community health worker program. As this is a preventive care program for the family as a whole, many other traditional health indicators may also change with the intensity of the program. In particular, one could analyze variations in health outcomes considered avoidable with primary health care (infant mortality by diarrhea, for example).

In sum, this essay shows that the efficacy of health care programs may be understated in traditional reduced form models of infant mortality because they ignore policy spillover possibilities. In this case, a benefit/cost analysis of combating the risk of infant mortality would pay less than necessary attention to these programs. As community participation programs have been advocated extensively by development agencies (World Bank and World Health Organization especially) this essay offers modeling perspectives to give full credit to these programs.

In Ceara State, the community health worker program constitutes good example of development policy combining policy decentralization and local participation. Enthusiastic researchers such as Tendler (1997) and Svitore (2000) provide expositional details on management conditions of the program without a formal framework of analysis. This essay, therefore, complements and reinforces their views by showing how the effectiveness of these programs may be fully appreciated. Otherwise, the possible side effects of under-investments in this program are fewer infant survivors, which is a tremendous burden borne by suffering families and by society as a whole.

Figure 1
Brazil –States
(Ceara State in dark)

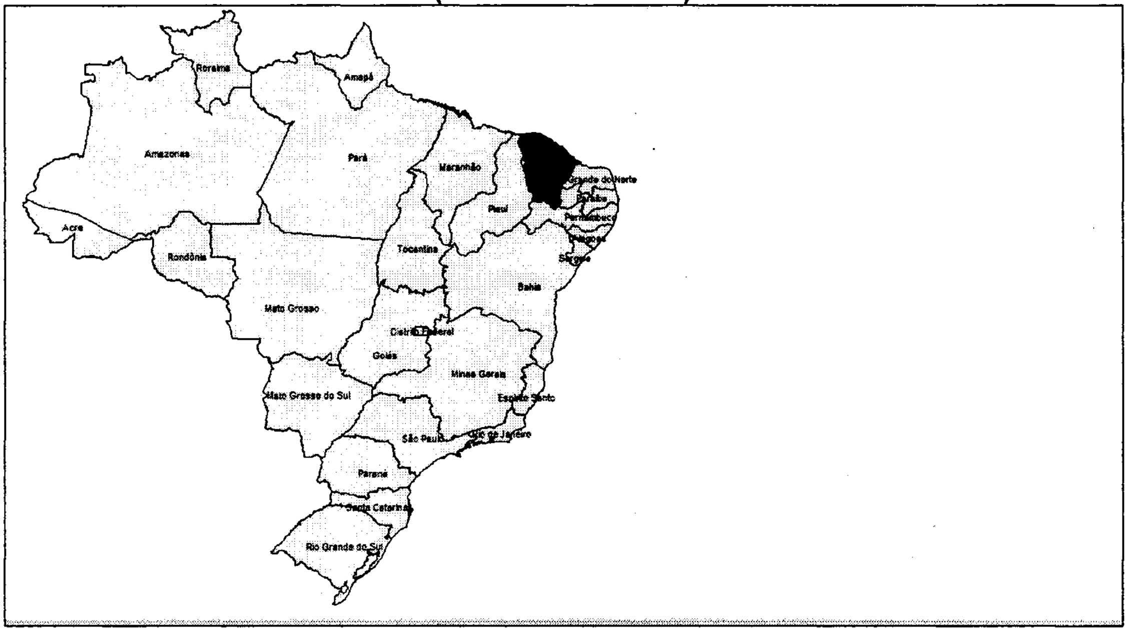


Figure 2
Ceara State – Municipalities
(Fortaleza City in dark)

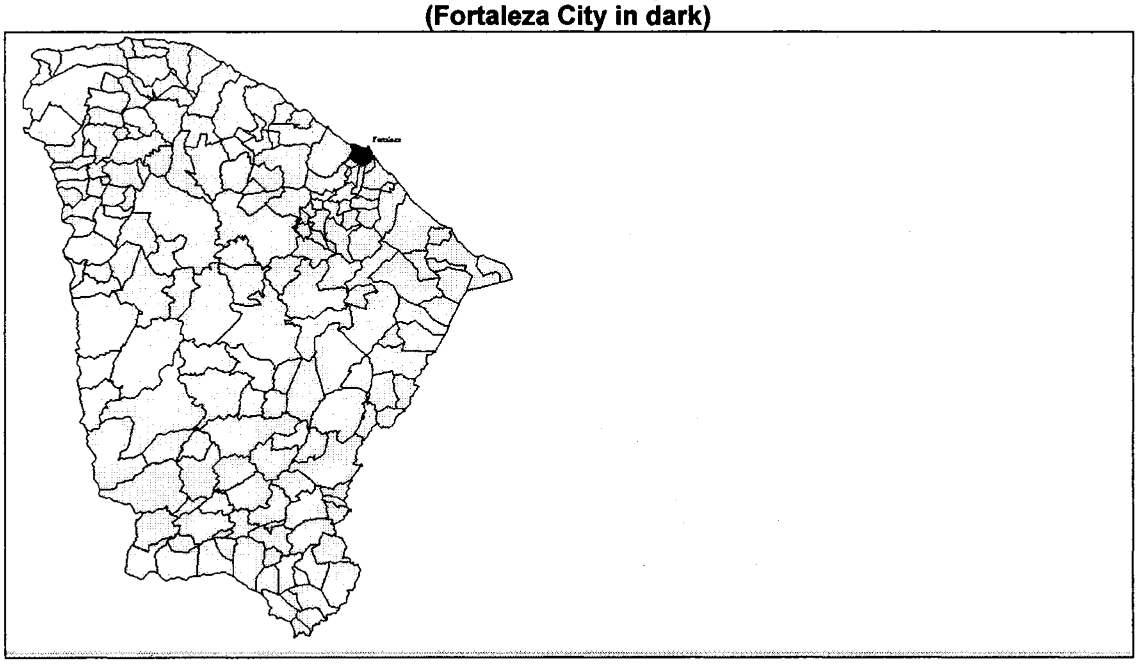


Figure 3

Square Root of Infant Mortality Rate - Municipalities of Ceara State – Brazil 1998/00

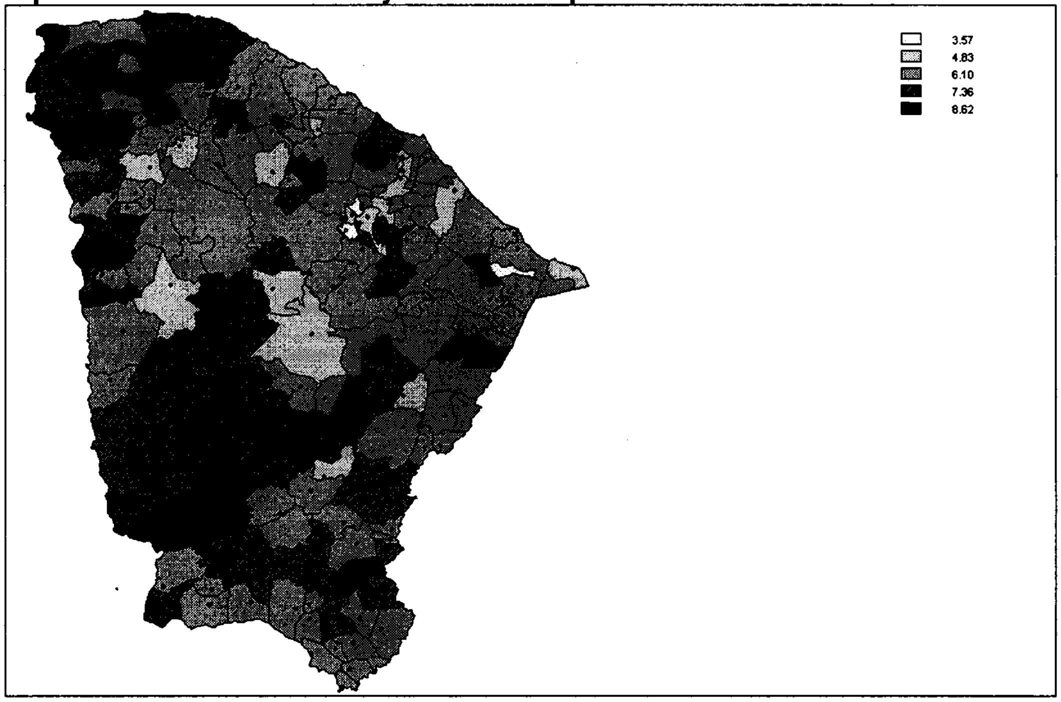
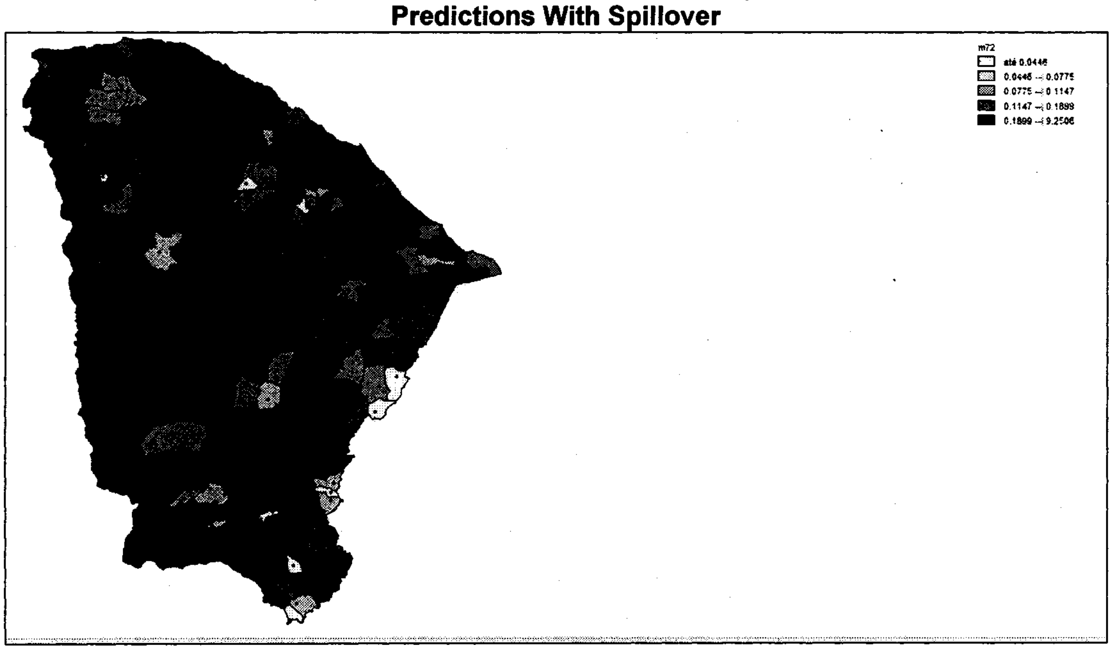
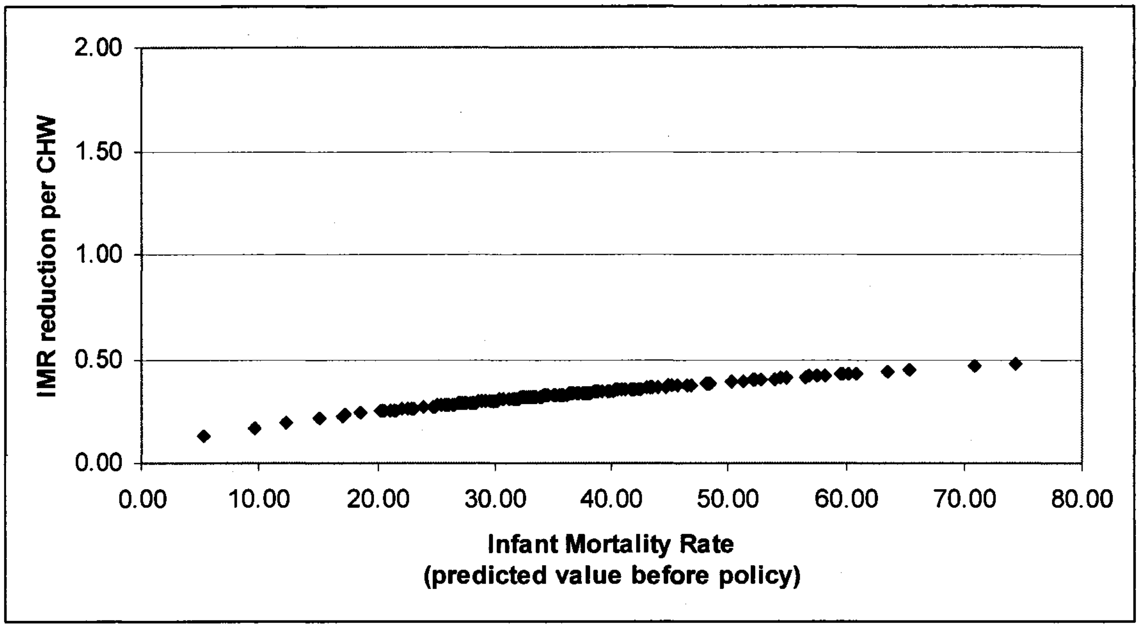


Figure 4
Municipal Policy Effectiveness – Deaths Avoided per Additional CHW
Predictions Without Spillover

Figure 5
Municipal Policy Effectiveness – Deaths Avoided per Additional CHW
Predictions With Spillover



Graph 1
Policy Effectiveness x IMR
Model Without Spillover



Graph 2
Policy EffectvinessProductivity x IMR
Model with spillover

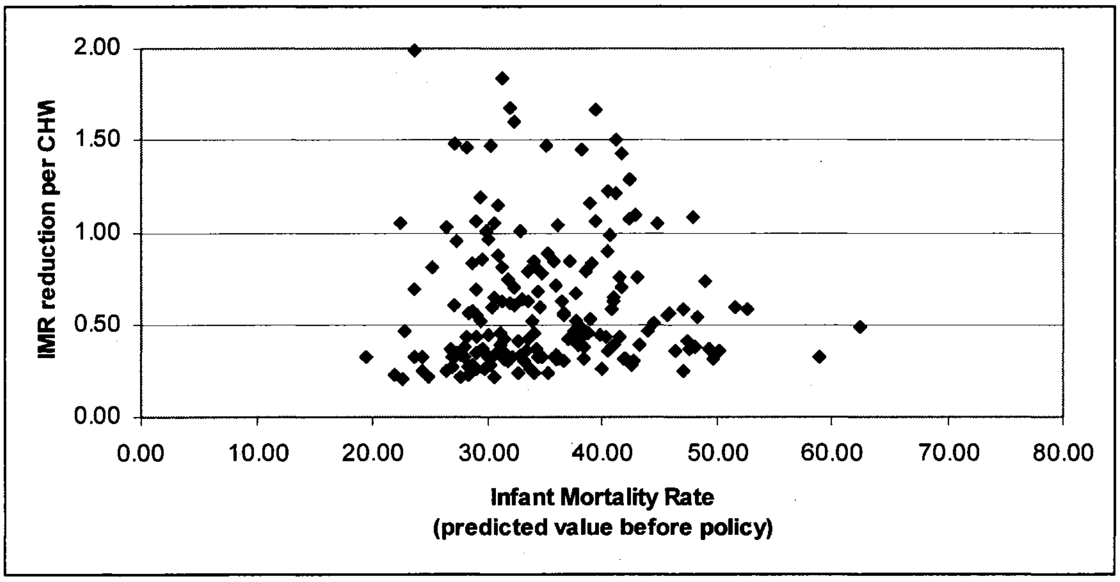


Table 1
Explanatory Variables

Variable name	Definition	Mean	Stdev
IMR9800	The square root of the infant mortality index. This last is calculated by multiplying 1000 to the ratio of infant deaths from 1998 to 2000 to the number of live births during the same period. Source: Annual Statistics of Ceara (IPLANCE)	5.89	0.97
Unemployment	The unemployment rate. Source: Brazilian Census 2000 (IBGE)	14.78	5.68
Sanitation	Percentage of families with adequate garbage removal. Source: Brazilian Census 2000 (IBGE)	38.09	16.40
Illiteracy Rate	Illiteracy rate for women aged 15-49. Source: Brazilian Census 2000 (IBGE)	27.86	5.75
Urban	Percentage of households living in urban areas (2000). Source – 2000 Census (IBGE)	53.35	16.39
Physicians	Number of physicians per group of 10000 individuals in 1999. Source: Annual Statistics of Ceara (IPLANCE)	11.84	6.70
Hospital Beds	Number of hospital beds per group of 10000 individuals in 1999. Source: Annual Statistics of Ceara (IPLANCE)	18.30	14.13
CHW	Number of community health workers per 10000 individuals during 1998/ 2000 period. Source: Annual Statistics of Ceara (IPLANCE)	20.44	35.44
Distance to Capital	Line distance in kilometers from the centroids of the municipalities to the centroid of the Capital of the State. Source: Annual Statistics of Ceara (IPLANCE).	270.51	158.64
Frontier	1 if the municipality has common border with municipalities of other states; 0 otherwise.	0.24	0.43

Table 2
Moran's I – Spatial Correlation Test

Variable	W1	W2	W3	W4
IMR99***,1	0.2988 (6.44)	0.2778 (5.58)	0.1109 (12.63)	0.0841 (7.20)
Unemployment***	0.2954 (6.37)	0.3033 (6.08)	0.0967 (11.08)	0.0322 (3.02)
Sanitation***	0.3326 (7.16)	0.4081 (8.15)	0.0913 (10.50)	0.1900 (15.72)
Illiteracy Rate***	0.4797 (10.28)	0.5780 (11.49)	0.1441 (16.23)	0.2597 (21.32)
Urban***	0.2702 (5.85)	0.3056 (6.13)	0.0820 (9.49)	0.1471 (12.26)
Physicians***	0.2240 (4.86)	0.1346 (2.76)	0.1122 (12.77)	0.0735 (6.34)
Hospital Beds	0.0655 (1.50)	0.0778 (1.64)	0.0247*** (3.27)	0.0385*** (3.53)
CHW***	0.2479 (5.36)	0.2749 (5.52)	0.0714 (8.34)	0.1216 (10.21)

(z value) * , ** and *** - Significant at 1%, 5% and 10% level of significance.
1- Level of significance across matrix specifications unless individually specified.

Table 3
IMR[1] Models Without Spatial Effects
(OLS estimation)

Spatial pattern		Without Spatial Effects	Without Spatial Effects but with time lag
Model		3.1	3.2
Constant		4.2268***	3.0960***
		(5.98)	(4.18)
Unemployment		0.0295**	0.0330***
		(2.60)	(3.01)
Sanitation		-0.0222***	-0.0208***
		(-3.22)	(-3.14)
Illiteracy Rate		0.0425***	0.0265
		(2.64)	(1.65)
Urban		0.0191***	0.0147**
		(3.12)	(2.45)
Physicians		-0.0086	-0.0081
		(-0.71)	(-0.70)
Hospital Beds		-0.0016	-0.0005
		(-0.31)	(-0.10)
CHW		-0.0276**	-0.0168
		(-2.19)	(-1.35)
Distance to capital (km)		0.0021***	0.0019***
		(4.51)	(4.14)
Frontier		-0.0565	-0.1249
		(-0.35)	(-0.81)
IMR 94-96 (square root)		-	0.2069***
			(3.85)
AIC		460.16	448.38
Jarque-Bera normality test		3.74	4.34
(p – value)		(0.154)	(0.11)
Spatial lag tests			
W1 – Contiguity	LM-LAG	6.37**	3.82*
W1 – Contiguity	LM-EL	9.61***	12.57***
W2 - Pop/contiguity	LM-LAG	6.21**	5.09**
W2 – Pop/contiguity	LM-EL	8.07***	10.02***
W3 – Inverse dist.	LM-LAG	7.09***	4.61**
W3 – Inverse dist.	LM-EL	13.60***	12.79***
W4 – Pop/distance	LM-LAG	1.35	1.14
W4 – Pop/distance	LM-EL	8.32***	8.60***
Spatial error tests			
W1 - Contiguity	LM-ERR	1.94	0.17
W1 - Contiguity	LM-LE	5.18**	8.91***
W2 – Pop/contiguity	LM-ERR	1.90	0.82
W2 – Pop/contiguity	LM-LE	3.76*	5.75**
W3 – Inverse dist.	LM-ERR	1.01	0.05
W3 – Inverse dist.	LM-LE	7.54***	8.22***
W4 – Pop/distance	LM-ERR	0.01	0.01
W4 – Pop/distance	LM-LE	6.98***	7.46***

1 – Square Root of Infant mortality Rate. (t-value). *, ** and *** Significant at 10%, 5% and 1% level of significance respectively.

Table 4
IMR[1] Spillover Models
(OLS – Estimation)

Spatial Patterns	Simple Contiguity (W1)	Population Contiguity (W2)	Inverse Distance (W3)	Population Distance (W4)
Models	4.1	4.2	4.3	4.4
Constant	5.4767***	5.6944***	6.2347***	4.7792***
	(7.06)	(7.27)	(3.14)	(6.01)
Unemployment	0.0360***	0.0339***	0.0343***	0.0290**
	(3.25)	(3.05)	(3.03)	(2.55)
Sanitation	-0.0191***	-0.0202***	-0.0185***	-0.0213***
	(-2.87)	(-3.02)	(-2.67)	(-3.01)
Illiteracy Rate	0.0264	0.0284*	0.0290*	0.0368**
	(1.64)	(1.77)	(1.74)	(2.27)
Urban	0.0150**	0.0159***	0.0149**	0.0204***
	(2.51)	(2.66)	(2.35)	(3.40)
Physicians	0.0001	-0.0075	0.0024	-0.0046
	(0.01)	(-0.64)	(0.20)	(-0.38)
Hospital Beds	-0.0044	-0.0026	-0.0051	-0.0041
	(-0.83)	(-0.50)	(-0.90)	(-0.75)
CHW	-0.0220*	-0.0189	-0.0264**	-0.0277**
	(-1.76)	(-1.52)	(-2.08)	(-2.15)
Distance to capital (km)	0.0030***	0.0030***	0.0030***	0.0026***
	(5.67)	(5.53)	(3.82)	(4.02)
Frontier	-0.0959	-0.1638	-0.0655	-0.1261
	(-0.61)	(-1.05)	(-0.40)	(-0.77)
W*Physicians	-0.0550***	-0.0483**	-0.0991	-0.0310*
	(-2.64)	(-2.54)	(-1.25)	(-1.80)
W*CHW	-0.0172	-0.0445**	0.0025	-0.0100
	(-0.85)	(-2.36)	(0.02)	(-0.50)
W*Hospital Beds	-0.0101	-0.0013	-0.0458	0.0025
	(-0.83)	(-0.14)	(-0.57)	(0.36)
AIC	447.01	448.38	456.98	460.35
R^2 – Adjust.	0.3494	0.3371	0.3132	0.3005
Heteroskedasticity (BP)	18.61*	17.32	16.12	21.96**
(p-value)	(0.09)	(0.13)	(0.18)	(0.03)
Spatial correlation tests				
LM - LAG	1.01	1.62	0.49	0.97
LM - LE	0.94	2.97*	3.84**	7.92***
LM – ERR	0.50	0.48	0.02	0.003
LM - EL	0.44	1.80	2.99*	6.95***
F – tests (F-statistics)				
H0: $\delta_{W*Physic.} = \delta_{W*CHW} = \delta_{W*Bed} = 0$	6.25***	5.82***	2.91*	1.83
H0: $\delta_{CHW} = \delta_{W*CHW} = 0$	2.36*	4.85***	2.23	2.77*
H0: $\delta_{Physic} = \delta_{W*Physic} = 0$	3.61**	3.57**	0.78	1.96

1 – Square Root of Infant mortality Rate. (t-value). *, ** and *** Significant at 10%, 5% and 1% level of significance respectively.

Table 5
IMR[1] Spillover Models With Spatial Lag
(IV Estimation – W*IMR = W*Xβ)

Spatial Patterns	Simple Contiguity (W1)	Population Contiguity (W2)	Inverse Distance (W3)	Population Distance (W4)
Models	5.1	5.2	5.3	5.4
Constant	5.4083***	5.9212***	-2.4928	1.8594
	(4.39)	(4.82)	(-0.45)	(1.50)
Unemployment	0.0359***	0.0339***	0.0335***	0.0287**
	(3.23)	(3.05)	(2.98)	(2.59)
Sanitation	-0.0191***	-0.0202***	-0.0185***	-0.0189***
	(-2.86)	(-3.02)	(-2.69)	(-2.80)
Illiteracy Rate	0.0259	0.0299*	0.0290*	0.0247
	(1.51)	(1.74)	(1.75)	(1.51)
Urban	0.0149**	0.0162***	0.0142**	0.0173***
	(2.43)	(2.65)	(2.24)	(2.86)
Physicians	0.0001	-0.0074	0.0030	-0.0011
	(0.02)	(-0.64)	(0.25)	(-0.10)
Hospital Beds	-0.0044	-0.0026	-0.0046	-0.0039
	(-0.83)	(-0.50)	(-0.82)	(-0.73)
CHW	-0.0219*	-0.0189	-0.0261**	-0.0254**
	(-1.76)	(-1.52)	(-2.07)	(-2.02)
Distance to capital (km)	0.0029***	0.0031***	0.0012	0.0017**
	(4.01)	(4.51)	(0.90)	(2.45)
Frontier	-0.0946	-0.1691	0.0057	-0.0125
	(-0.60)	(-1.07)	(0.03)	(-0.08)
W*Physicians	-0.0546**	-0.0493**	0.0092	-0.0273
	(-2.57)	(-2.53)	(0.09)	(-1.62)
W*CHW	-0.0173	-0.0452**	-0.0193	-0.0198
	(-0.85)	(-2.37)	(-0.20)	(-1.01)
W*Hospital Beds	-0.0099	-0.0015	-0.0501	0.0044
	(-0.80)	(-0.17)	(-0.64)	(0.65)
W*IMR (ρ)	0.0146	-0.0476	1.4470*	0.6124***
	(0.03)	(-0.24)	(1.67)	(3.01)
AIC	449.00	450.32	456.00	452.79
Heteroskedasticity (BP)	19.03	18.11	20.68*	17.47
(p – value)	(0.12)	(0.15)	(0.08)	(0.17)
F – tests (F-statistics)				
H0: $\delta_{CHW} = \delta_{W*CHW} = 0$	2.43*	5.36***	2.34*	3.74**
H0: $\delta_{Physic} = : \delta_{W*Physic} = 0$	3.50**	4.02**	0.04	2.19

1 – Square Root of Infant mortality Rate. (t-value). *, ** and *** Significant at 10%, 5% and 1% level of significance respectively.

Table 6
IMR[1] Spatial Lag Models
(IV Estimation – W*IMR = W*Xβ)

Spatial Patterns	Simple Contiguity (W1)	Population Contiguity (W2)	Inverse Distance (W3)	Population Distance (W4)
Models	6.1	6.2	6.3	6.4
Constant	2.8062***	3.1759***	-5.1279*	1.1111
	(2.70)	(3.06)	(-1.77)	(0.96)
Unemployment	0.0295***	0.0302***	0.0329***	0.0296***
	(2.62)	(2.67)	(2.97)	(2.69)
Sanitation	-0.0209***	-0.0220***	-0.0192***	-0.0192***
	(-3.04)	(-3.20)	(-2.84)	(-2.84)
Illiteracy Rate	0.0287	0.0323*	0.0311*	0.0278*
	(1.63)	(1.83)	(1.94)	(1.72)
Urban	0.0160**	0.0171***	0.0155**	0.0165***
	(2.54)	(2.75)	(2.57)	(2.76)
Physicians	-0.0076	-0.0089	0.0016	-0.0041
	(-0.64)	(-0.74)	(0.14)	(-0.35)
Hospital Beds	-0.0021	-0.0018	-0.0036	-0.0021
	(-0.40)	(-0.34)	(-0.69)	(-0.41)
CHW	-0.0259**	-0.0268**	-0.0265**	-0.0275**
	(-2.06)	(-2.12)	(-2.16)	(-2.24)
Distance to capital (km)	0.0014**	0.0017***	0.0003	0.0013**
	(2.42)	(3.06)	(0.53)	(2.53)
Frontier	-0.0276	-0.0423	0.0226	0.0317
	(-0.17)	(-0.27)	(0.14)	(0.20)
W*IMR (ρ)	0.3504*	0.2588	1.7162***	0.6483***
	(1.86)	(1.38)	(3.32)	(3.37)
AIC	458.53	460.14	450.78	450.48
Heteroskedasticity (BP)	14.34	13.23	16.01*	10.60
(p-value)	(0.16)	(0.21)	(0.09)	(0.38)

1 – Square Root of Infant mortality Rate. (t-value). *, ** and *** Significant at 10%, 5% and 1% level of significance respectively.

Table 7
IMR[1] Spatial Spillover Models With Time lag
(OLS Estimation)

Spatial Patterns	Simple Contiguity (W1)	Population Contiguity (W2)	Inverse Distance (W3)	Population Distance (W4)
Models	7.1	7.2	7.3	7.4
Constant	4.3189***	4.5103***	4.1924**	3.5871***
	(5.19)	(5.35)	(2.09)	(4.35)
Unemployment	0.0385***	0.0365***	0.0368***	0.0331***
	(3.56)	(3.37)	(3.36)	(3.01)
Sanitation	-0.0183***	-0.0192***	-0.0180***	-0.0198***
	(-2.82)	(-2.97)	(-2.69)	(-2.99)
Illiteracy Rate	0.0145	0.0163	0.0161	0.0214
	(0.91)	(1.02)	(0.98)	(1.33)
Urban	0.0119**	0.0128**	0.0122*	0.0161***
	(2.04)	(2.18)	(1.97)	(2.69)
Physicians	-0.0003	-0.0074	0.0010	-0.0041
	(-0.03)	(-0.660	(0.09)	(-0.35)
Hospital Beds	-0.0032	-0.0016	-0.0036	-0.0030
	(-0.63)	(-0.32)	(-0.67)	(-0.57)
CHW	-0.0140	-0.0113	-0.0175	-0.0164
	(-1.14)	(-0.92)	(-1.39)	(-1.29)
Distance to capital (km)	0.0026***	0.0027***	0.0025***	0.0021***
	(5.07)	(5.01)	(3.18)	(3.37)
Frontier	-0.1451	-0.2060	-0.1162	-0.1817
	(-0.95)	(-1.35)	(-0.74)	(-1.16)
W*Physicians	-0.0512**	-0.0417**	-0.0922	-0.0356**
	(-2.53)	(-2.25)	(-1.21)	(-2.14)
W*CHW	-0.0114	-0.0350*	0.0360	-0.0082
	(-0.57)	(-1.89)	(0.38)	(-0.43)
W*Hospital Beds	-0.0072	-0.0016	-0.0304	0.0062
	(-0.61)	(-0.19)	(-0.40)	(0.93)
IMR 94-96 (square root)	0.1737***	0.1737***	0.1896***	0.2081***
	(3.28)	(3.26)	(3.50)	(3.85)
AIC	437.72	439.21	446.19	446.96
Heteroskedasticity (BP)	14.79	12.68	12.91	22.08*
(p – value)	(0.32)	(0.47)	(0.45)	(0.05)
F – tests (F-statistics)				
H0: $\delta_{CHW} = \delta_{W*CHW} = 0$	0.98	2.58*	0.98	1.06
H0: $\delta_{Physic} = : \delta_{W*Physic} = 0$	3.35**	2.87*	0.72	2.53*

1 – Square Root of Infant mortality Rate. (t-value). *, ** and *** Significant at 10%, 5% and 1% level of significance respectively.

Table 8
IMR[1] Spillover Models With Time lag and Spatial Lag
(IV Estimation – W*IMR = W*Xβ)

Spatial Patterns	Simple Contiguity (W1)	Population Contiguity (W2)	Inverse Distance (W3)	Population Distance (W4)
Models	8.1	8.2	8.3	8.4
Constant	4.1902***	4.8002***	-4.2581	1.2116
	(3.34)	(3.86)	(-0.78)	(1.00)
Unemployment	0.0384***	0.0364***	0.0360***	0.0325***
	(3.53)	(3.36)	(3.30)	(3.01)
Sanitation	-0.0182***	-0.0192***	-0.0180***	-0.0179***
	(-2.80)	(-2.96)	(-2.71)	(-2.74)
Illiteracy Rate	0.0137	0.0181	0.0161	0.0124
	(0.80)	(1.06)	(0.99)	(0.77)
Urban	0.0117*	0.0132**	0.0115*	0.0139**
	(1.95)	(2.19)	(1.86)	(2.33)
Physicians	-0.0004	-0.0074	0.0016	-0.0011
	(-0.04)	(-0.65)	(0.14)	(-0.10)
Hospital Beds	-0.0032	-0.0016	-0.0032	-0.0029
	(-0.63)	(-0.32)	(-0.58)	(-0.56)
CHW	-0.0139	-0.0113	-0.0172	-0.0155
	(-1.12)	(-0.92)	(-1.38)	(-1.24)
Distance to capital (km)	0.0025***	0.0028***	0.0007	0.0014**
	(3.56)	(4.17)	(0.55)	(2.09)
Frontier	-0.1428	-0.2129	-0.0467	-0.0803
	(-0.93)	(-1.38)	(-0.29)	(-0.50)
W*Physicians	-0.0506**	-0.0431**	0.0128	-0.0319*
	(-2.44)	(-2.26)	(0.13)	(-1.95)
W*CHW	-0.0114	-0.0360*	0.0146	-0.0167
	(-0.57)	(-1.91)	(0.16)	(-0.88)
W*Hospital Beds	-0.0068	-0.0019	-0.0346	0.0075
	(-0.56)	(-0.22)	(-0.45)	(1.14)
W*IMR (ρ)	0.0274	-0.0614	1.4035*	0.5203***
	(0.14)	(-0.32)	(1.67)	(2.62)
IMR 94-96 (square root)	0.1738***	0.1740***	0.1883***	0.1897***
	(3.27)	(3.26)	(3.49)	(3.54)
AIC	439.70	441.10	445.18	441.64
Heteroskedasticity (BP)	15.74	14.84	17.30	17.00
(p – value)	(0.32)	(0.39)	(0.24)	(0.25)
F – tests (F-statistics)				
H0: $\delta_{CHW} = \delta_{W*CHW} = 0$	0.95	2.64*	0.97	4.86***
H0: δ_{Physic} = : $\delta_{W*Physic} = 0$	3.12**	2.92*	0.02	1.99

1 – Square Root of Infant mortality Rate. (t-value). *, ** and *** Significant at 10%, 5% and 1% level of significance respectively.

Table 9
IMR[1] Spillover Models With and Without Time lag
(OLS and IV Estimations – W*IMR = W*Xβ)

Spatial Patterns	Simple Contiguity (W1)	Simple Contiguity (W1)	Population Contiguity (W2)	Population Contiguity (W2)
Models	9.1	9.2	9.3	9.4
Constant	5.4767***	4.3189***	5.6944***	4.5103***
	(-7.06)	(5.19)	(-7.27)	(5.35)
Unemployment	0.0360***	0.0385***	0.0339***	0.0365***
	(-3.25)	(3.56)	(-3.05)	(3.37)
Sanitation	-0.0191***	-0.0183***	-0.0202***	-0.0192***
	(-2.87)	(-2.82)	(-3.02)	(-2.97)
Illiteracy Rate	0.0264	0.0145	0.0284*	0.0163
	(-1.64)	(0.91)	(-1.77)	(1.02)
Urban	0.0150**	0.0119**	0.0159***	0.0128**
	(-2.51)	(2.04)	(-2.66)	(2.18)
Physicians	0.0001	-0.0003	-0.0075	-0.0074
	(-0.01)	(-0.03)	(-0.64)	(-0.66)
Hospital Beds	-0.0044	-0.0032	-0.0026	-0.0016
	(-0.83)	(-0.63)	(-0.50)	(-0.32)
CHW	-0.0220*	-0.0140	-0.0189	-0.0113
	(-1.76)	(-1.14)	(-1.52)	(-0.92)
Distance to capital (km)	0.0030***	0.0026***	0.0030***	0.0027***
	(-5.67)	(5.07)	(-5.53)	(5.01)
Frontier	-0.0959	-0.1451	-0.1638	-0.2060
	(-0.61)	(-0.95)	(-1.05)	(-1.35)
W*Physicians	-0.0550***	-0.0512**	-0.0483**	-0.0417**
	(-2.64)	(-2.53)	(-2.54)	(-2.25)
W*CHW	-0.0172	-0.0114	-0.0445**	-0.0350*
	(-0.85)	(-0.57)	(-2.36)	(-1.89)
W*Hospital Beds	-0.0101	-0.0072	-0.0013	-0.0016
	(-0.83)	(-0.61)	(-0.14)	(-0.19)
IMR 94-96 (square root)	-	0.1737***	-	0.1737***
		(3.28)		(3.26)
AIC	447.01	437.72	448.38	439.21
F – tests (F-statistics)				
H0: $\delta_{CHW} = \delta_{W*CHW} = 0$	2.36*	0.98	4.85***	2.58*
H0: $\delta_{Physic} = : \delta_{W*Physic} = 0$	3.61**	3.35**	3.57**	2.87*

1 – Square Root of Infant mortality Rate. (t-value). *, ** and *** Significant at 10%, 5% and 1% level of significance respectively.

Table 10
IMR Level Models Without Spatial Effects
(OLS estimation)

Spatial pattern		Without Spatial Effects	Without Spatial Effects but with time lag
Model		10.1	10.2
Constant		15.927*	11.623
		(1.90)	(1.43)
Unemployment		0.3414**	0.3864***
		(2.54)	(2.98)
Sanitation		-0.2634***	-0.2476***
		(-3.22)	(-3.15)
Illiteracy Rate		0.5050***	0.2923
		(2.65)	(1.54)
Urban		0.2220***	0.1644**
		(3.06)	(2.32)
Physicians		-0.1173	-0.1043
		(-0.82)	(-0.76)
Hospital Beds		-0.0214	-0.0083
		(-0.33)	(-0.13)
CHW		-0.2893*	-0.1675
		(-1.93)	(-1.14)
Distance to capital (km)		0.0245***	0.0218***
		(4.33)	(3.98)
Frontier		-0.5374	-1.5391
		(-0.28)	(-0.84)
IMR 95 (level)		-	0.1779***
			(4.01)
AIC		1369.56	1355.23
Jarque-Bera normality test		13.61***	13.12***
(p – value)		(0.001)	(0.001)
W1 – Contiguity	LM-LAG	`6.38**	3.63*
W1 – Contiguity	LM-EL	10.54***	12.63***
W2 - Pop/contiguity	LM-LAG	5.89**	4.62**
W2 – Pop/contiguity	LM-EL	8.87***	9.94***
W3 – Inverse dist.	LM-LAG	6.03**	3.65*
W3 – Inverse dist.	LM-EL	11.19***	9.80***
W4 – Pop/distance	LM-LAG	1.20	0.96
W4 – Pop/distance	LM-EL	6.92***	6.79***
Spatial error tests			
W1 - Contiguity	LM-ERR	1.76	0.11
W1 - Contiguity	LM-LE	5.92**	9.11***
W2 – Pop/contiguity	LM-ERR	1.45	0.55
W2 – Pop/contiguity	LM-LE	4.44**	5.87**
W3 – Inverse dist.	LM-ERR	0.88	0.03
W3 – Inverse dist.	LM-LE	6.04**	6.18**
W4 – Pop/distance	LM-ERR	0.01	0.01
W4 – Pop/distance	LM-LE	5.72**	5.84**

(t-value). *, ** and *** Significant at 10%, 5% and 1% level of significance respectively.

Table 11
IMR Level Spillover Models With and Without Time Lag
(OLS – Estimation)

Spatial Patterns	Simple Contiguity (W1)	Simple Contiguity (W1)	Population Contiguity (W2)	Population Contiguity (W2)
Models	11.1	11.2	11.3	11.4
Constant	31.178***	24.965***	34.238***	27.696***
	(3.41)	(2.75)	(3.71)	(3.02)
Unemployment	0.4193***	0.4503***	0.3954***	0.4278***
	(3.20)	(3.53)	(3.02)	(3.36)
Sanitation	-0.2259***	-0.2170***	-0.2379***	-0.2279***
	(-2.87)	(-2.84)	(-3.03)	(-2.98)
Illiteracy Rate	0.3122	0.1565	0.3325*	0.1732
	(1.65)	(0.83)	(1.76)	(0.91)
Urban	0.1726**	0.1328*	0.1822**	0.1418**
	(2.45)	(1.92)	(2.59)	(2.04)
Physicians	-0.0093	-0.0107	-0.1026	-0.0960
	(-0.07)	(-0.08)	(-0.75)	(-0.72)
Hospital Beds	-0.0539	-0.0402	-0.0331	-0.0219
	(-0.86)	(-0.66)	(-0.53)	(-0.36)
CHW	-0.2183	-0.1305	-0.1799	-0.0983
	(-1.49)	(-0.90)	(-1.23)	(-0.68)
Distance to capital (km)	0.0347***	0.0306***	0.0358***	0.0319***
	(5.55)	(4.93)	(5.48)	(4.96)
Frontier	-1.0434	-1.7837	-1.8676	-2.5050
	(-0.57)	(-0.99)	(-1.02)	(-1.40)
W*Physicians	-0.6741***	-0.6252***	-0.6013***	-0.5188**
	(-2.75)	(-2.62)	(-2.69)	(-2.37)
W*CHW	-0.2272	-0.1588	-0.5645**	-0.4459**
	(-0.95)	(-0.68)	(-2.54)	(-2.04)
W*Hospital Beds	-0.1109	-0.0708	-0.0130	-0.0162
	(-0.77)	(-0.51)	(-0.12)	(-0.16)
IMR 95 (level)	-	0.1483***	-	0.1476***
		3.39		(3.37)
AIC	1355.59	1345.55	1355.97	1346.10
Heteroskedasticity (BP) (p – value)				
F – tests (F-statistics)				
H0: $\delta_{CHW} = \delta_{W*CHW} = 0$	1.95	0.79	4.79 ***	2.61*
H0: $\delta_{Physic} = \delta_{W*Physic} = 0$	3.93**	3.60**	4.04**	3.19**

(t-value). *, ** and *** Significant at 10%, 5% and 1% level of significance respectively

Table 12
Policy Simulation - Program Effectiveness

Policy: 1 additional community health worker for each group of 10,000 people for each municipality			
	Panel A - Deaths avoided per CHW		
Place	Without Spillover	With Spillover (W1)	With Spillover (W2)
State Average	2.505	2.410	3.485
	Panel B - Change in IMR per CHW		
Place	Without Spillover	With Spillover (W1)	With Spillover (W2)
State Average	0.240	0.380	1.230
Municipalities			
Fortaleza	0.202 (184th)	0.347 (109th)	1.984 (4th)
Salitre	0.335 (1st)	0.408 (60th)	0.599 (73rd)

State Average Productivity = $\Sigma_i \Delta$deaths (or ΔIMR)$_i$/$\Sigma_i\Delta$CHW$_i$) Municipal Productivity = Δdeaths (or ΔIMR)/ΔCHW

* Municipal ranking of productivity value according to model specification.

BIBLIOGRAPHY

Alves, D. and Belluzo, W. (2004) – Infant Mortality and Child Health in Brazil – *Economics and Human Biology*, 2 pp. 391-410.

Andrade, C. L. T and Szwarcwald, C. L. (2001) – Spatial Analysis of Early Neonatal Mortality in the municipality of Rio de Janeiro, 1995-1996 – *Caderno de Saude Publica*, 17 (5).

Anselin, L. (1986) – Spatial Econometrics Methods and Models – Kluwer Academic Publishers.

Anselin, L. (1988) – Lagrange Multiplier Test Diagnostic for Spatial Dependence and Spatial Heterogeneity – *Geographical Analysis* 20 pp. 1-17.

Anselin, L. (1992) – A Workbook for Using SpaceStat in the Analysis of Spatial Data – University of Illinois.

Anselin L. and Hudak (1992) – Spatial Econometrics in Practice – *Regional Science and Urban Economics* 22 pp. 509-536.

Anselin, L. and Florax, R. (1995) – *New Directions in Spatial Econometrics* – Berlin: Springer-Verlag.

Anselin, L.; Bera, A. K.; Florax, R. and Yoon, M. J. (1996) – Simple Diagnostic Tests for Spatial Dependence - *Regional Science and Urban Economics* 26 pp. 77-104

Anselin L. and Bera, A.K. (1998) – Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics – in Ullah, A. and Giles, D. (Eds.), *Handbook of Applied Economic Statistics,* Marcel Dekker: NY.

Anselin, L. (2003) – Spatial Externalities, Spatial Multipliers and Spatial Econometrics – *International Regional Science Review* – 26 (2) pp. 153-166.

Assuncao et al (1998) – Maps of epidemiological rates: a Bayesian approach – *Caderno de Saude Publica* 14 (4) pp. 713-723.

Barnes-Boyd, C., Norr, K.F. and Nacion, K.W. (2001) – Promoting Infant Health Through Home Visiting By a Nurse-Managed Community Worker Team – *Public Health Nursing* 18 (4) pp. 225-235.

Bell, K.P and Bockstael, N.E. (2000) – Applying the Generalized-Moments Estimation Approach to Spatial Problems Involving Microlevel Data – *Review of Economic and Statistics* 82 (1).

Bithell, J.F.; Dutton, S.J.; Neary, N.M.; and Vincent, T.J. (1995) – Controlling for Socioeconomic Confounding Using Regression Methods – *Journal of Epidemiology and community Health* 49 (2) S15-S19.

Browning, M. et al (1994) – Income and Outcomes: A Structural Model of Intrahousehold Allocation – *Journal of Political Economy* 102 (6).

Brueckner, J. K. (1998) – Testing for Strategic Interaction Among Local Governments: The Case of Growth Controls – *Journal of Urban Economics* 44 pp. 438-467.

Burridge, P. (1980) – On the Cliff-Ord Test for Spatial Autocorrelation – *Journal of the Royal Statistical Society B* 42 pp. 107-108.

Case, A.; Rosen, H.; and Hines, J(1993) – Budget Spillovers and Fiscal Policy Interdependence: Evidence from the States – *Journal of Public Economics*, 52.

Case, A. C. (1991) – Spatial patterns in household demand – *Econometrica* 59 (4) pp. 953-965.

Clayton, D. G., et al (1993) – Spatial Correlation in Ecological Analysis – *International Journal of Epidemiology* 22 (6).

Cliff, A. D. and Ord, J. K. (1981) – Spatial Processes Models & Applications - London, *Pion.*

Conway, K. S. and Rork, J. C. (2004) – Diagnosis Murder - The Death of State 'Death' Taxes – *Economic Inquiry.*

Corman, H. and Grossman, M. (1985) – Determinants of Neonatal Mortality Rates in the U.S. A Reduced Form Model – *Journal of Health Economics* 4 (3).

Corman, H., Joyce, T., and Grossman, M. (1986) – Birth Outcome Production Function in the United Sates – *The Journal of Human Resources*, XXII. 3.

Corman, H., Grossman, M. and Joyce, T.J. (1988) – Demographic Analysis of Birth-Specific Neonatal Mortality – *NBER* Working Paper 2804.

Currie, J. and Gruber, J. (1997) – The Technology of Birth: Health Insurance, Medical Interventions, and Infant Health – *NBER Working Paper* 5985.

Fingleton, B. (2001) – Theoretical Economic Geography and Spatial Econometrics: Dynamic Perspectives – *Journal of Economic Geography* 1.

Florax, R.J.G.M.; Folmer, H. and Rey, S. J. (2003) – Specifications Searches in Spatial Econometrics: The Relevance of Hendry's Methodology – *Regional Science and Urban Economics* 33 pp. 557-579.

Ford, T. (2002) – Blessing and curses: a spatial analysis of the growth effects of foreign direct investment in the United States – Phd Dissertation - University of New Hampshire.

Grossman, M. (1972) – The Demand for Health: A Theoretical and Empirical Investigation – NBER *OccasionalPaper119*

Kalipeni, E. (1993) – Determinants of Infant Mortality in Malawi: A Spatial Perspective – *Social Science and Medicine* 37 (2).

Kelejian, H.H. and Prucha, I. R. (1997) – Estimation of Spatial Regression Models with Autoregressive Errors by Two-Stage Least Squares Procedures: A Serious Problem – *International Regional Science Review* 20, 1& 2 p. 103-111.

Kelejian, H.H. and Prucha, I. R. (1998) – A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances – *Journal of Real Estate Finance and Economics* 17 (1) p. 99-121.

Leal, M. C and Szwarcwald, C. L (1997) – Characteristics of Neonatal Mortality in the State of Rio de Janeiro, Brazil, in the 1980's: a spatio-temporal analysis. – *Revista de Saude Publica* 31 (5).

Lawson, A. (1996) – Use of deprivation indices in small area studies – *Journal of Epidemiology and Community Health,* 50.

Lindsay, A. C. (2002) – Integrating Quantitative and Qualitative methods to Assess the Impact of Child Survival Programs in Developing Countries: The Case of a Program Evaluation in Ceara, Northeast Brazil – *Health Education and Behavior* 29 (5) p. 570-584.

Loran, V. et al (2001) – Deprivation and Mortality: the implications of spatial autocorrelation for health resources allocation - *Social Science and Medicine*, 53.

Moran, P. A. (1948) – The Interpretation of Statistical Maps – *Journal of the Royal Statistical Society*, series B 10.

Preston, Samuel (1975) – The Changing Relations Between Mortality and the Level of Economic Development – *Population Studies* 29 (2).

Rahmanifar, A. (2001) – Indiana Infant Mortality Report: 1999 Period Linked Brith/Infant Death Data Set – Epidemiology Resource Center – Special Report V.1 n.1.

Rosenzweig, M.R. and Schultz, T.P. (1982) – The behavior of mothers as inputs to child health: the determinants of birth weight, gestation, and rate of fetal growth – In *Economics Aspects of Health.* Ed. Victor R. Fuchs, Chicago, IL. University of Chicago Press.

Shimakura, S.E. et al (2001) – Spatial risk distribution: modeling infant mortality in Porto Alegre, Rio Grande do Sul State, Brazil.

Simoes, C. (1996) – A Saude Infantil no Brasil nos Anos 90 – *Infancia Brasileira nos Anos 90* UNICEF/IBGE, Brasilia.

Souza, A. C. et al (1999) – Variations in Infant Mortality Rates Among Municipalities in the State of Ceara, Northeast Brazil: an Ecological Analysis – *International Journal of Epidemiology* 28.

Souza, A. C. et al (2000) – Circumstances of Post-Neonatal Deaths in Ceara, Northeast Brazil: Mother's health Care-Seeking Behaviors During Their infants' Fatal Illness – *Social Science and Medicine* 51 p. 1675-1693.

Sourza, A. C. et al (2001) – Underlying and Proximate Determinants of Diarrhea-Specific Infant Mortality Rates Among the Municipalities in the State of Ceara, North-East Brazil: An Ecological Study – *Journal of Biosocial Science* 3 (2) p. 227-244.

Svitone, E. C. et al (2000) – Primary Health Care Lessons from the Northeast of Brazil: the Agentes de Saude Program – *Pan Am J Public Health* 7(5).

Tendler, J. (1997) – Good Government in the Tropics – Baltimore and London The Johns Hopkins University Press.

World Health Organization (WTO) (1978) – *Health: A Time forJustice* – Geneva: Public Information Division, WHO

World Health Organization (WTO) (2005) – *Redesigning Child Care: Survival, Growht and Development* – Geneva: World Development Report 2005, chapter six.

APPENDICES

APPENDIX A – POLICY PRODUCTIVITY EFFECTS

Health spillover effects in this paper refer to the possibility that the benefits of health service (health structure or health policies) improvements in one region spillover to its neighbors. The magnification of this effect is captured by the weighted average of health service availability in neighbor regions which appear as additional explanatory variables in the reduced form models. Among the three theoretical possibilities of incorporating spatial effects in reduced form models, the data at hand fits better a health spillover models (theoretical model 2). Rewriting this model for each observation and assuming for simplification that there is only one health service variable (H) we have:

$$Y_i = Xi\beta + \gamma H_i + \delta \Sigma_j w_{ij} H_j + \varepsilon_i \qquad (2.1) \qquad \begin{array}{l} i = 1,\ldots,184 \\ j = 1, \ldots,184 \end{array}$$

Where w_{ij} refers to the weight that municipality j has on the municipality i that is given by the standardized version of the corresponding weighting matrix (W) and Yi is the square root of infant mortality rate in each municipality ($Y_i = IMR_i^{1/2}$). The spillover effect refers to the estimated marginal effect that improvements in health service availability in one municipality has on the infant mortality rate of its neighbors. That is, the spillover effect of a municipality j in a neighbor municipality i is given by:

(2.2)
$$\frac{\partial Y_i}{\partial H_j} = \delta w_{ij}$$

Since Y_i is the square root transformation of infant mortality rate, the marginal effect in terms of infant mortality rate becomes:

(2.3)
$$\frac{\partial IMR_i}{\partial H_j} = \frac{\partial IMR_i}{\partial Y_i} x \frac{\partial Y_i}{\partial H_j} = 2IMR_i^{1/2}\delta w_{ij} = 2Y_i \delta w_{ij}$$

It can be seen that the efficacy and extension of the spillover effect provided by improving policy in municipality j depends on the square root of infant mortality rate of the neighbor i (Y_i), the estimated general marginal effect of the neighbors (δ), and the spatial weights designated for this municipality by its neighbors (w_{ij}). As each municipality has usually more than one neighbor, the aggregated spillover effect has to add the impact effect on each of these:

(2.4)
$$\Sigma_i \frac{\partial IMR_i}{\partial H_j} = 2\delta \Sigma_i w_{ij} IMR_i^{1/2}$$

The estimated total spillover effect is an important reference for regional health policies because it gives an estimation case-by-case of the distributional benefits of improving health conditions in one particular place. The total effect of health service improvements in one municipality therefore is given by its in-municipality marginal effect plus its spillover effect:

Total Effect

(2.5)
$$\frac{\partial IMR_j}{\partial H_j} + \Sigma_i \frac{\partial IMR_i}{\partial H_j} = 2\gamma IMR_j^{1/2} + 2\delta \Sigma_i w_{ij} IMR_i^{1/2} = 2\gamma Y_j + 2\delta \Sigma_i w_{ij} Y_i$$

In order to facilitate the understanding of the productivity effect calculated in the policy simulation section, however, the empirical models of reference can be written in their matrix form. To facilitate even more the analysis it is possible to isolate the health variable of interest for which spillover effect is significant. The empirical version of model 2 is given by:

(2.6) $$Y = [\gamma I + \delta W]H$$

With Y representing a n x 1 column matrix where each element corresponds to the predicted value of the square root of the infant mortality rate in each municipality, I is a n x n identity matrix, W is a n x n spatial matrix (spatial matrix W2), H is a n x 1 column matrix with the number of community health workers per group of 10,000 people in each cell, and γ and δ are the estimated parameter (scalar) coefficients. The sum of the elements of the columns of the matrix $[\gamma I + \delta W]$ corresponds to the total effect (2.5) of one additional unit of the health policy variable in a particular municipality. That is, it tell us how this additional unit implemented in this municipality reduces the square root of infant mortality rate in the own municipality and in its neighbors. In a 3 x 3 specification for example these elements are:

(2.7) $$\begin{bmatrix} \gamma & \delta w_{12} & \delta w_{13} \\ \delta w_{21} & \gamma & \delta w_{23} \\ \delta w_{31} & \delta w_{32} & \gamma \end{bmatrix}$$

This is because the spatial matrix W has diagonal elements (w_{ii}) equal to zero. The first column of this matrix shows the marginal effect that the change in the health policy in municipality 1 has in itself (γ), in municipality 2 (δw_{21}), and in municipality 3 (δw_{31}). The sum of these marginal effects is the total effect of municipality 1 in the square root of infant mortality rate statewide. If we want to know the effect in the level of infant mortality rate instead of square root we may differentiate 2.6 with respect to IMR and obtain:

(2.8) $$\frac{\Delta IMR}{\Delta H} = \frac{\Delta IMR}{\Delta Y} x \frac{\Delta Y}{\Delta H} = 2IMR^{1/2}[\gamma I + \delta W] = 2Y^{d}[\gamma I + \delta W]$$

where Y^d is a n x n diagonal matrix having the square root of infant mortality in each municipality (Y_i) as diagonal elements. In a 3 x 3 specification the marginal effect for one municipality in terms of reduction in infant mortality rate statewide is given by the sum of elements of the columns of the following matrix:

(2.9) $$2\begin{bmatrix} \gamma y_1 & \delta w_{12} y_1 & \delta w_{13} y_1 \\ \delta w_{21} y_2 & \gamma y_2 & \delta w_{23} y_2 \\ \delta w_{31} y_3 & \delta w_{32} y_3 & \gamma y_3 \end{bmatrix}$$

where y_i is the square root of infant mortality rate in municipality i.

In the policy simulation section the productivity of the health care program (CHWP) is compared in models with and without spillovers when the government designates one additional community health worker per group of ten thousand individuals for each municipality. In models without spillover δ is equal to zero and equation 2.8 becomes:

(2.10) $$\frac{\Delta IMR}{\Delta H} = 2\gamma Y^d$$

For empirical purpose the discrete difference in the health policy is consider for each model specification assuming that one unit increase in H is equal to the population size (p_i) divided by 10,000[61]. Therefore, when the productivity is measured in terms of reduction in the level of infant mortality per community health worker added the values are calculated from the following models:

(2.11a)	$\Delta IMR = 2\gamma Y^d \Delta H$	Without Spillover Empirical models of Table 3
(2.11b)	$\Delta IMR = 2Y^d[\gamma I + \delta W]\Delta H$	With Spillover Empirical models of Table 9

Where ΔIMR is a n x n matrix with the total predicted changes in the risk of infant death for each municipality given by the sum of row elements, and ΔH is a n x n diagonal matrix where each non-zero element is given by the population size of each municipality divided by 10,000. Using the 3 x 3 example, we have that ΔIMR will be equal to:

(2.12a) Without Spillover
$$2\begin{bmatrix} \gamma y_1(p_1/10{,}000) & 0 & 0 \\ 0 & \gamma y_2(p_2/10{,}000) & 0 \\ 0 & 0 & \gamma y_3(p_3/10{,}000) \end{bmatrix}$$

(2.12b) With Spillover
$$2\begin{bmatrix} \gamma y_1(p_1/10000) & \delta w_{12}y_1(p_2/10{,}000) & \delta w_{13}y_1(p_3/10{,}000) \\ \delta w_{21}y_2(p_1/10{,}000) & \gamma y_2(p_2/10{,}000) & \delta w_{23}y_2(p_3/10{,}000) \\ \delta w_{31}y_3(p_1/10{,}000) & \delta w_{32}y_3(p_2/10{,}000) & \gamma y_3(p_3/10{,}000) \end{bmatrix}$$

[61] H is actually measured as the number of community health workers divided by population size and multiplied by 10,000. Therefore, if we divide the population size by 10,000 we have an approximation on how many additional community health workers should be added in each municipality to increase H (proportion of CHW per 10,000 habitants) by one unity.

Each column element of these matrices represents the marginal effect of the policy improvements in one particular municipality in reducing its own infant mortality rate and the infant mortality rate of its neighbors. It is easy to identify that in models without spillover only the in-municipal marginal effect is considered. The empirical estimates, however, showed that γ is greater for models without spillover which could compensate for the absence of spillover effects in estimating total productivity which for the state as a whole is given by:

(2.13) $$\text{State Average Productivity} = \Sigma_{ij}\Delta IMR/\Sigma_{ij}\Delta H$$

For each municipality the productivity measure is given by:

(2.14) $$\text{Municipal Average Productivity} = \Sigma_{j}\Delta IMR^{j}/\Sigma_{j}\Delta H^{j}$$

where ΔIMR^j and ΔH^j are the column elements of each respective matrices.

When the productivity is measured in infant deaths avoided by community health workers we are assuming that the number of live births remains constant with the policy affecting only the numerator (infant deaths) of the infant mortality index. Although this assumption may be appealing at first sigh it does not compromise the comparative simulations for productivity effects in models with and without spillover given it applies to both.

As only the numerator (child deaths) of the IMR index is assumed to change, equations 2.11a and 2.11b are modified to:

(2.15a) $$\Delta Deaths = 2\gamma Y^{d} LB^{d} \Delta H$$ Without Spillover
Empirical models of Table 3

(2.15b) $$\Delta Deaths = 2Y^{d} LB^{d} [\gamma I + \delta W] \Delta H$$ With Spillover
Empirical models of Table 9

Where LB^d is a n x n diagonal matrix with its elements equal to number of live births[62] in each municipality divided by 1000. This additional term transforms the changes in the risk measure of infant deaths to the actual number of predicted infant deaths avoided per municipality after the policy. ΔDeaths is therefore a n x n matrix of predicted changes in the number of infant deaths, and the productivity measure for the whole state and for each municipality is calculated as:

(2.16a) State Average Productivity = $\Sigma_{ij}\Delta Deaths/\Sigma_{ij}\Delta H$

(2.16b) Municipal Average Productivity = $\Sigma_{j}\Delta Deaths^{j}/\Sigma_{j}\Delta H^{j}$

It was calculated from the empirical estimates that in order to increase one additional community health worker for each group of 10,000 individuals in each municipalities (policy) it was necessary to add a total of approximately 743 community health workers for the whole state ($\Sigma_{ij}\Delta H = 743$). The policy resulted in a predicted reduction of 1858 infant deaths ($\Sigma_{ij}\Delta Deaths=1858$) in models without spillover, and 1790 ($\Sigma_{ij}\Delta Deaths=1790$) or 2594 ($\Sigma_{ij}\Delta Deaths=2594$) in models with significant spillovers defined by W1 and W2 respectively.

[62] The average actual number of life births between 1998 and 2000 for each municipality is applied to both models.

APPENDIX B – SPATAIL CORRELATION DIAGNOSTIC TESTS

Test	Source
$LM-ERR=\left(\frac{e'We}{T}\right)^{2} \sim \chi^{2}_{(1)}$	Burridge (1980)
$LM-EL=\frac{\left(e'We/S^{2}-T\left(R\tilde{J}\right)^{-1}\left(e'Wy/S^{2}\right)\right)^{2}}{\left(T-T^{2}\left(R\tilde{J}\right)^{-1}\right)} \sim \chi^{2}_{(1)}$ $\left(R\tilde{J}\right)^{-1}=\left[T+(WX\beta)'M(WX\beta)/S^{2}\right]^{-1}$	Anselin et al (1996)
$LM-LAG=\frac{\left(e'Wy/S^{2}\right)}{R\tilde{J}} \sim \chi^{2}_{(1)}$	Anselin (1988)
$LM-LE=\frac{\left(e'Wy/S^{2}-e'We/S^{2}\right)}{R\tilde{J}} \sim \chi^{2}_{(1)}$	Anselin et al (1996)

e = R by 1 vector of regression residuals from OLS regression (model 1).
s^2 = e'e/R
T = tr(W'W + W^2), where tr is the matrix trace operator and W is the spatial weighting matrix.
WXβ = Spatial lag of the predicted value from an OLS regression.
M = I – X(X'X)$^{-1}$X is the projection matrix.

PART 2: NEIGHBORHOOD EFFECTS AND INFORMALIZATION OF THE LABOR MARKET – THE CASE OF FORTALEZA, BRAZIL

CHAPTER 1

INTRODUCTION

In developing countries, a large fraction of the working-age population is employed as informal workers. For the last three decades, this fraction has been growing progressively, reaching more than 50% in many Latin American countries and including mostly the self-employed and employees working in small firms without formal labor contract[1]. The growth of informal workers causes many concerns for policy makers. Three main reasons among others have stimulated the study of the so-called informal economy. First, a growing informal sector compromises the reliability of traditional qualitative indicators of the labor market. Low levels of unemployment rate, for example, may be compensated by the accommodation of workers in low quality informal jobs which can induce erroneous diagnostics about labor market conditions. Second, a high informality rate may contribute to the weakening of the tax and social security bases if a vicious cycle of high taxes and consequent escape to informality drives governments and workers respectively. Finally, a large informal sector may be an indication of failures in the labor market usually attested to by the existence of worker segmentation.

[1] The definition of informal worker varies considerably in the literature and more details on this issue will be provided later in this essay. See Table 1 for an overview.

The challenging task of neglecting (or otherwise) labor market segmentation between formal and informal workers has dominated the literature on the informal economy for three decades and persists to this day (Ulyssea, 2005). The traditional view of segmentation is that informality is an involuntary under-employment condition, more associated with the survival of workers than to the accumulation of earnings[2]. Above market clearing wages sustained mostly by institutional constraints (labor legislations, unions and/or wage productivity premiums, for example) induce an excess of supply of workers for which informality is the only alternative left. This involuntary allocation of workers makes them less productive than if they were in the formal sector, which characterizes allocation failures in the labor market.

This residual and marginal character of the informal sector, stated by the segmentation theory, has been contested theoretically and empirically. Conceptually, some authors prefer to understand the informal sector as an unregulated small firm sector for which benefits of informality "...could be linked to the misalignment of implicit and explicit labor taxes with perceived benefits or the desire of workers to retain a degree of independence in their work (Maloney and Cunningham, 2002 p. 5)". In this case, informality is desirable, and informal workers should be seen as an integrated rather than segmented part of the labor market.

[2] The term under-employment identifies those last resort types of jobs. That is, informal workers are unsuccessful formal job seekers. The term, therefore, has a broader definition than working less (part-time, for example) than desirable, and it is difficult to be measured or determined empirically.

Empirically, segmentation tests have been mostly based on wage differences between formal and informal workers. Three main concerns apply to these tests of segmentation. First, it is not unusual to find significant error measurements in the reported earnings of informal workers in labor market surveys. Self-employed workers, for example, commonly do not differentiate human from capital returns from their working services. The worker may report revenues (or losses) rather than estimated wages of his/her one man firm. In addition, in-kind payments and fringe benefits may also be excluded from reported earnings for informal and formal workers respectively. In this case, the (reported) wages of formal and informal workers are not directly comparable.

Another criticism in testing segmentation refers to the aggregation of different types of workers under the same label of informal workers. Within traditional empirical definitions based on firm size or social security contribution it is possible to find informal workers who do not fit the segmentation description of an under-employed involuntary worker. Labor market surveys in Latin America, for example, have shown that many self-employed workers prefer their current situations than working as a formal employee (Maloney and Cunningham, 2002). The dichotomy of formal versus informal workers in investigating segmentation, therefore, may be inadequate as segmentation may exist within each category of worker (Funkenhouser, 1996).

The third concern in testing segmentation with wage differences is that unobserved characteristics of the workers which affect sector choice and earnings are difficult to identify. Traditional remedies for self-selection bias based on two-stage estimation[3] may also be inadequate if the first stage of sector choice is not properly modeled. Maloney (1999) cites Heckman stating that: "...A poor first-stage selection specification may actually induce an unknown bias rather than improve the ordinary least square estimates". In the context of informality this seems to be the case since we don't know *a priori* the conditions under which workers engage in informal activities. Misspecifications in the sector choice model may include a restrictive modeling approach such as ordered probit (Pradhan and Van Soest, 1995) or the exclusion of important determinants of sector choice.

It is interesting to note, for example, that the informal literature has devoted little or no attention to the influence of social interactions on worker's sector decision. Networking and other forms of spontaneous informational channels (observing or hearing about) are responsible for a large portion of new jobs acquired (Topa, 2001). Interactions may occur for formal or informal positions. With respect to the latter, individuals may receive an extra incentive to be informal workers, for example, if they know that others around him/her are doing the same. Scheider and Enste (2000) states that informality represents discredit of one social norm (formal or regulated employment) and the acceptance of another (informal employment). This discredit may increase with the level of informality perceived, which may represent an additional motivation or a breaking stigma factor to leave unemployment in favor of informal activities. Therefore, social interactions may play an important role in workers' labor market

[3] This estimation method is known in the literature as Heckman procedure. See Greene (2000) for further details on this method.

decisions, and their neglect in segmentation analyses may compromise their reliability as misspecification issues arise.

This essay serves as an intermediary input to the analysis of labor market segmentation by investigating if the concerns suggested above apply to the study case of Fortaleza City in Brazil. That is, this essay investigates: i) whether dualisms in the practical definitions of workers (formal versus informal) are appropriate modeling perspectives for the labor market, and ii) whether social interactions are important determinants of worker's sector choice. In order to perform these analyses, this essay modifies traditional multinomial logit models (Gallaway and Bernasek (2002), Pradhan and Van Soest (1995)) in two ways: i) expansion from three to five types of workers by crossing empirical definitions of informal workers based on firm size and social security contributions, and ii) introduction of interaction possibilities in the model.

The analysis is facilitated by a unique pool of cross-sectional data sets which not only allows multiple definition criteria for informal workers, but also permits the confrontation of individual and aggregated data at the neighborhood level. This possibility of neighborhood identification for the workers, together with an emblematic growth of the informal economy during the 1990s - absorbing more than half of the active workers in the labor market – makes the study case of Fortaleza city in Brazil very interesting and exclusive so far.

In sum, this essay connects the informal economy literature and the literature on neighborhood effects in order to investigate informal workers' differentiations and interactions in the labor market. An overview of both types of studies with their main issues and methods is offered in the next section. The third section introduces contextual perspectives and institutions related to the labor market of Fortaleza, Brazil. The fourth section brings the methodology proposed to investigate interactions among different types of workers. Finally, sections five and six respectively discuss the empirical findings supporting heterogeneity among informal workers and the existence of interaction effects in neighborhoods, and present the main conclusions extracted from these general findings.

CHAPTER 2

LITERATURE REVIEW

This essay links ideas and methods of two branches of the economics literature: the informal economy literature and the literature on neighborhood effects. The next two subsections discuss the hypotheses and approaches used in the literature to explain (i) the role of the informal economy, and (ii) the investigation of neighborhood effects in other contexts.

2.1. Informal Economy Literature

Widespread use of the informal sector concept as a way to describe small-scale urban activities in labor markets of developing countries arose in the beginning of the 1970s with the works of Hart (1973) and the International Labor Organization (ILO, 1972). The conceptualization of a unique sector for different types of workers was well accepted by the literature, as it fitted well the description of the disadvantaged sector of segmentation theories.

In general, segmentation theories assert that labor markets are divided into two contrasting sectors. The primary sector contains the most desirable (high-paid) jobs in the labor market as they follow and appropriate the technological progress available. Workers in this sector are fully rewarded according to their productive capacity. Access to this sector is restricted, however, either by efficient wage measures for those already in the sector, by institutional constraints (minimum wages, unions, and labor market regulations, for example) or even by discrimination. The excess of unsuccessful primary sector job seekers is then employed in the secondary sector which develops marginally with a limited productivity capacity of their workers (Deoringer and Piori, 1971). The informal sector, as a synonym for the secondary sector, is therefore considered the main indicator of an inefficient labor market.

With the worldwide persistence and growth of the informal sector during the 1980's and 1990's, especially in Latin American countries, the under-employed (or segmented) character of informal workers began to be more intensively re-examined. Multiple methods and practical definitions are employed to test whether informal workers are segmented. The most traditional methodology involves earning comparisons in a dual labor market where workers are either formal or informal. Differences in wages, in this case, should reflect differences in productivity.

Table 1 shows many studies using comparative earnings equations to state positions with respect to the segmentation hypothesis. It is interesting to note that mixed results are found, but they vary with respect to methodology, definitions, and market places. Gingdling (1991), after controlling for possibilities of self-selection in Costa Rica's labor market and using the Chow test to compare wage structures, favors the segmented view by observing significant and positive differences in earnings for formal workers due to discrimination. Tennen (1991), on the other hand, contradicts segmentation, by verifying that the dummy variable for informal workers becomes insignificant in the wage equation after using Hackman procedure for self-selection in the labor market of Brazil.

The use of wage equations to test segmentation, however, has some caveats. First, reported earnings of the informal workers are not usually comparable to the earnings of formal workers as returns to capital and labor are usually mixed for the former. Second, the dual definition of workers (formal versus informal) can be inadequate. Depending on the practical definition of informal and formal workers, it is possible to find differentiations within each of these categories. The most traditional distinction criteria are based either on social security contributions or on the size and type of the firms. These two criteria may represent different but related types of segmentation sources. It is possible, for example, that within small or large firms, social security contributions may represent an additional differentiator for worker types. The aggregation of different types of workers in this case may induce erroneous generalizations about the role of informal workers in the labor market.

Some authors have recognized these difficulties in testing segmentation with wage equation comparisons and have proposed the use of sector choice models with mobility indicators between sectors as an alternative methodology. That is, segmentation is verified when there are mobility differences between formal and informal positions. In the presence of segmentation, and specially in economic downturns, it is easier for a formal worker to become an informal worker than the reverse. Gong and Van Soest (2002) use panel data to check on how previous working conditions affect the likelihood of switching sectors in

Mexico. They find that "...men's sector choice is not affected by their labor market state in the previous quarter, implying that there is no genuine state of dependence". This finding is consistent with an integrated rather than segmented labor market with no significant barriers to mobility. This study however also uses dual definitions (formal versus informal) without accounting for possible heterogeneity within sectors.

Maloney (1999) also uses panel data to construct transition tables for different types of workers in Mexico. He argues against segmentations by finding that transition probabilities across types of workers are not asymmetric, which suggests that there are no significant barriers to mobility into the formal or informal sector. Maloney (1999) recognizes heterogeneity in the informal sector by differentiating three types of informal workers: self-employed, informal salaried workers, and contract workers. In order to differentiate formal from informal workers he uses alternatively the most traditional definition criteria based on firm size or social security contributions. These two strategies for empirical definitions can also be combined to differentiate informal workers. The advantage in this case is the possibility to identify heterogeneity among self-employed workers and small firm employees who are the core of the informal sector, and whose great range of activities implies qualitative differences among themselves as evidenced in the following section.

2.1.1. Formal versus informal workers: empirical definitions

Mostly influenced by dualistic theories (Lewis, 1954), the informal sector during the 1970's was classified broadly by characteristics such as no barriers to entry, firms with small scale operation, family ownership, and unregulated activities. This broad definition allowed different adaptations for empirical investigation according to what information is available from each country's labor market surveys. Although consensus is still difficult to reach, as shown in Table 1, two main empirical perspectives can be identified: the legalist and the entrepreneurial.

According to the legalist identification, informal activities are those occurring at the margin of the country's working legislation. Social security payments are usually the main mark separating formal from informal workers because additional information on tax evasion or firm's registration is usually not available at the individual level. This type of classification follows Portes et al (1989) which views informal activities as those *"...unregulated by the institutions of society, in a legal and social environment in which similar activities are regulated"* (Portes et al, 1989 p.12).

This definition is usually employed by those who further investigate regulatory issues in the labor market (Neri, 2002) or those who simply aim to have a qualification identifier based on common social rights (Telles, 1992). Informal workers are those who have no protection against impediments to work provoked by unemployment, injuries or aging, and therefore, are more vulnerable than formal workers to these conditions[4].

[4] The protections here refer to the usual rights provided by the social security or unemployment systems, for example.

The focus on labor market legislation however obscures the differentiation between traditional and non-traditional working arrangements that would be related to the scale and type of the business under consideration. In this regard, the entrepreneurial definition of informal workers uses the size of the firm as the point of differentiation. Small firm employees and employers, self-employed and domestic workers are those considered informal[5]. Marcoullier et al (1997) points out that those analysts using entrepreneurial definition of informality usually devote much attention to the capacity of labor markets to generate traditional (formal) working positions.

Table 2 shows a generalization of these two definitions. It is structured for a labor market survey with information on workers' position in their occupation, the size of the worker's firm, and their contribution to social security programs.

Workers can be classified as employees (salaried or wage workers who have a formal or informal contract to work regularly for a particular firm), self-employed (those who work for their own account offering services to one or more firms without formal contract, or to consumers directly – taxi drivers, hairdressers, regular street vendors, computer programmers, dressmakers, etc), liberal professionals (self-employed workers with college degree and registered as a professional in their respective council - lawyers, physicians, accountant, etc)[6],

---

[5] The self-employed and domestic workers are considered a one-man firm and therefore enter under the definition of small firm, for which the maximum size varies according to the characteristics of the labor market and the survey questionnaire.

[6] Those who have a college degree but have unrelated working activities (a lawyer whose main activity is to be the manager of his own restaurant, for example) are classified as self-employed. Liberal professionals are not usually considered informal workers according to the firm size definition because they represent traditional and recognized positions in the labor market. This essay also follows this interpretation.

employer (those who have their own firm and hire employees), family workers (those who work in family business without receive income), and domestic workers (maids or car drivers working for families rather than firms).

The classification of workers according to their position in occupation rather than according to the type of occupation itself is used mostly to identify self-employed workers. These are the core of the informal sector as they represent alternative employment solutions to traditional working contracts and in many cases exemplify the entrepreneurial spirit of the workers. The same taxi driver, for example, can be an employee or a self-employed worker if he works for a specialized urban transportation firm or if he decides to use his own car to work alone respectively. The process of outsourcing production used by the firms more intensively during the 1990s contributed to the rise of small and/or individual (self-employment) firms. The (renewed) question that arises from the segmentation theory is whether the choice of being his/her own boss in small firms is really a voluntary choice for the worker or if it is the last resource in a labor market with constrained entry into traditional (big) firms.

Under this perspective of segmentation, the use of the entrepreneurial definition of informal workers would be more adequate to the analysis. There are, indeed, many intersections between the two definitions criteria presented above as it is shown in Table 2. Some authors use both definitions separately in their analyses to compare results. Marcoullier et al (1997), for example, finds that the share of informal workers when calculated using the legalist perspective is higher than when the firm size criterion is chosen for the labor markets of Mexico, El

Salvador and Peru. Maloney (1999) and Gong and Van Soest (2002) also compare results using one or the other criterion in a panel data analysis of segmentation for the labor market of Mexico. None of these studies, however, have combined both definitions to differentiate workers according to size of the firm and the contribution to the social security system.

The broad classification of informal workers in one category (small firm workers, for example) may hide an important heterogeneity that appears exactly where the definitions do not coincide. The differentiation between those self-employed who contribute or not to the social security system may reflect a significant quality difference between these two types of workers. Telles (1993) shows, for example, that self-employed workers contributing to the social security system receive earnings comparable to those of regular employees. In addition, the choice to be a small firm worker and contribute to the social security system can be correlated by unobservable characteristics of the workers. In this case, an analysis based on either one or other definition becomes inconsistent.

This essay proposes to study differences among workers by combining these two types of definitions. That is, this essay differentiates those who work in small firm or not (entrepreneurial criterion), who are further divided between those who contribute to the social security system or not (legalist criterion). There are, therefore, four types of workers to be compared in their main characteristics: small-firm workers uncovered by the social security program (SF-UCOV), small-firm workers covered by the social-security program (SF-COV), firm workers uncovered by the social security program (F-UCOV) and firm workers covered by the social security programs (F-COV). Small firm workers include: self-employed workers, employees and employers working for a firm with less than three workers[7], and domestic workers[8]. Consequently, firm workers include those liberal professionals and those employees and employers who work for a firm with three or more workers (Table 3).

The classification of a small firm based on a very few number of workers aims to identify those positions where the worker is his/her own boss or an associate. Flexibility in working schedule and self-management are usually pointed out as advantages by surveyed self-employed workers when they report their motives for engaging in this type of employment (Maloney, 1999). On the other hand, self-employed workers are more subjected to considerable variability in earnings when compared to traditional wages (or salaries) received by regular employees.

Segmentation perspectives would be suggested if transition probabilities across the four types of workers (SF-COV, SF-UCOV, F-COV and F-UCOV) were asymmetric after controlling for worker's characteristics (Maloney (1999) or Gong and Van Soest (2002)). As panel data is not available for the labor market of Fortaleza in Brazil yet, this specific analysis of labor market segmentation is left for future research. Alternatively, this essay concentrates efforts to differentiate and qualify those workers with respect to the under-employed

[7] It is assumed that differences between employees or employers in firms with one or two individuals and self-employed workers are due to a taxonomic imprecision from the surveyed worker with respect to his/her own classification. A definition based only on self-employed workers will also be used for comparative purpose in the robustness section.
[8] Family workers are excluded from the analysis as their usual apprentice motivation is not comparable to those of working-for-pay workers.

character of each working position. This under-employed characteristic of each position is investigated in a multinomial logit model by analyzing how education affects their allocation chances when compared to the chances of unemployed workers.

The use of unemployed workers as a comparative group and the expansion from two to four types of workers are modifications of the labor supply model of Gallaway and Bernasek (2002)[9]. It is suggested in segmentation models that informal survival workers cannot afford job search. In this case unemployment arises as a "privileged" position when compared to those employed in the informal sector (Fields, 1975). The last-resort character of one position with virtually no entry cost is symptomatic when those more endowed with human capital prefer to keep looking for other opportunities than to take a lesser position. In order to identify undesirability perspectives for informal workers, this essay compares how educational levels affect the likelihood of working in different positions with the likelihood of being unemployed. The log-odds and marginal effects derived from a multinomial logit model for an average individual are used for this purpose. Although this analysis does not provide a final position about segmentation perspectives in the labor market it serves as an intermediate input to qualify workers according to the assumptions of the segmentation theory.

[9] Gallaway and Bersanek (2002) compare informality likelihoods to either formal activities or not working in general. This last position includes not only unemployed workers but also those out of the labor force.

Finally, the main innovation proposed here is the inclusion of neighborhood effects into the sector choice model. These effects have been generally theorized [10] as the result of interactions (networking or imitation), contextualization (sorting into neighborhoods of individuals with similar characteristics), or local constraints (local determinants of labor demand). In the context of labor supply, neighborhood effects represent additional facility mediators (or constraints) between the potential worker and the different positions in the labor market. Although neighborhood effects have not been included yet in labor supply models with sector choice, a long list of studies analyze these in different contexts as the following section illustrates.

2.2. Neighborhood Effects Literature

The idea that non-market interactions also affect individual decisions has been emphasized in many different aspects of economic theories and is a primary assumption in most sociological models. Interactions in neighborhoods may take the form of peer effects (Powell et al, 2002; Case and Katz, 1991), epidemic theories of ghettos (Crane, 1991), group effects (Grodner, 2002) or local spillovers (Topa, 2001). These theories have been used in many study cases such as smoking, school drop out, negligent behavior of youth, public insurance take-up and unemployment.

The primary assumption in the field is that individual behavior is mostly influenced by one's own characteristics and by one's susceptibility to affecting

[10] See Dietz (2002) for an overview about the general theorization of neighborhood effects in social science studies.

and being affected by social interactions. Aizner and Currie (2002), for example, investigate how neighbors and/or ethnic networks can affect the probability of public health insurance take-up. Case and Katz (1991) show that many youths' negligent behavior (crime and idleness among others) is induced by peer effects in neighborhoods. Crane (1991) evidences that social problems such as teenage childbearing and school drop out may be epidemic, as the likelihood of individual incidence increases exponentially in the poorest neighborhoods. It is interesting to emphasize that the quality of the neighborhoods in this study is defined by the percentage of professional or managerial workers in the neighborhood. The author justifies the option by saying that these types of workers can be role models for the youth, and therefore can affect their behavior.

Interactions in the labor market have been mostly understood as informational channels which improve the chances of unemployed individuals finding a job. Topa (2001) emphasizes, for example, that some evidence suggests that: "...more than 50% of all new jobs are found through friends, relatives, neighbors, or occupational contacts rather than through formal means. This is especially true for low-skill jobs, for less educated workers and for black workers (Topa, 2001 p. 262)"[11].

In observing spatial correlation in unemployment rate at the census track level, Topa (2001) argues that local informational spillovers indeed help individuals to leave unemployment and find a job. The author, however, does not specify what types of workers (formal or informal) are more susceptible to these

[11] Topa (2001) cites Corcoran et al (1980) and Granovetter (1995).

interactions. This essay addresses this question. In addition, if interactions are important factors influencing worker's chances in the labor market they may also affect transitions across worker types. Consequently, segmentation models based on these transitions become missspecified if they exclude the possibility of interactions in the labor market.

Although the interaction idea may sound intuitive, it is much more difficult to test empirically. The three main issues discussed by the literature are the neighborhood sorting, simultaneity, and contextual problems. In trying to assess social interactions in a labor supply model with multiple choices, one could estimate the following multinomial logit model according to Brock and Durlauf (2002):

$$\Pr(C_{in} = j) = \frac{\exp(K_j^{'}Z)}{\sum_s \exp(K_s^{'}Z)} \qquad (1)$$

$$\text{with } K'Z = \delta' X_{in} + \Sigma_s \alpha^s C^s{}_n$$

Where δ is a parameter vector associated with the characteristics of an individual i (X_{in}) who lives in neighborhood n, and $C^s{}_n$ is the percentage of workers in the labor force state s in the neighborhood n. Rejecting the null hypothesis that α^s is equal to zero can be indicative of a social interaction effect. It is very likely, however, that the parameter α^s rather than capturing the interaction among neighbors is rather reflecting their similarities in choosing residential location. This is the so-called sorting (or reflection) problem as analyzed by Manski (1993). Non-randomness in the formation of neighborhoods compromises the inference of α^s as a "pure" social interaction effect (networking or imitation, for example), and identification problems may arise if the workers' characteristics defining labor supply and sector choice coincide with those defining neighborhood sorting.

The other issues in the estimation of social interaction effects refer to simultaneity and contextual problems. The first problem states that social interaction variables ($C^s{}_n$) are potentially endogenous in contemporaneous regressions when individuals are at the same time affecting and being affected by others' decisions. The second problem occurs when there are local conditions important in defining workers' choices in the labor market which are not captured by any of the explanatory variables. In this case we have an omitted variable bias (Grodner, 2002). The use of instrumental variables (Grodner, 2002 or Case and Katz, 1991) and fixed effects (Carman (2004)) are usual solutions for these two types of problems respectively.

On the other hand, the social interaction literature is in general very skeptical about accepting exogenous solutions to the sorting problem when the interacting group (neighborhoods in this case) is not randomly created (Mansky, 1993). Empirical studies evolve attempting to overcome this issue, and some palliative solutions are offered by the literature. Grodner (2002), investigating social interaction on labor supply (number of hours worked), proposes factorial analysis to estimate average characteristics for groups of references as a way to capture the sorting effect and therefore isolate the social interaction effect (captured by the mean working hours in the reference group). Aizer and Currie (2002), on the other hand, interact time and neighborhood fixed effect variables to differentiate network and neighborhood effects on the take-up of health public programs.

Neighborhood sorting and social interactions of workers in neighborhoods are social phenomena for which possibilities depend on their contextualization in time and space. Consequently, in order to differentiate these two possible effects, one must understand the context in which they are more likely to occur. The next section, therefore, identifies some peculiarities of the labor market of Fortaleza City, in Brazil. The empirical methodology section shows how the main issues stated above are addressed in this essay. Advancing a little on this subject, this essay uses a combination of the proposals of Grodner (2002) and Aizer and Currie (2002) to address the issue of endogenous sorting. That is, the sorting effect is aimed to be captured by neighborhood quality indicators and regional fixed effects. This solution was greatly influenced by the specifications of the study site and by the data set available.

CHAPTER 3

STUDY SITE AND DATA SET

With a population of 2.2 million, Fortaleza is the 5th largest urban area in Brazil. The labor market is characterized by a high level of informality for which a local definition is given by the addition of those small-firm uncovered, small-firm covered, and firm-uncovered types of workers according to the definition suggested in this essay[12]. Graphs 1 and 2 show the participation of each type of worker in the labor force during the 1990's and earlier 2000's.

For male workers there is a smooth downward trend to the only category of worker assigned as formal by the local definition (firm covered workers). There is also a downward trend for females that goes from 1993 to 1999, followed by a small increase in the participation of formal workers afterwards. It is interesting to note that this downward trend occurred especially at times of economic growth. This evidence contradicts the traditional segmented view that informality is countercyclical.

It is possible, however, that even in periods of economic growth the labor market has not created enough formal positions to absorb the growing labor

[12] At this point forward in this essay the concept of informal sector will be applied to the aggregation of these three types of workers (small firm uncovered, small firm covered and firm uncovered workers). That is, the general concept of informal sector is based on the Brazilian's definition.

force. On the other hand, it is possible that informal workers (or at least part of them) are adapting themselves to general trends in the labor market such as the growth of the service industry. In this scenario, social interactions may also have contributed to explain this informalization of the labor market by working as a breaking stigma factor or even by the simple replication of activities. The perception that others around (neighbors) are becoming independent workers (a taxi driver, hairdresser, door-to-door sales person, or a small retailer, for example) may represent, therefore, an extra incentive to someone unemployed or even with formal positions to do the same. Social interactions in this case can be driven only by observing (or hearing) and developing what other workers around are doing.

Scheider and Enste (2000) also argue that the informalization of the labor market can also be a sign of decreasing tax morale, where the population becomes less averse to the risk of being caught by fiscal authorities. Together with the already high cost of being a formal worker, this avoidance of contributions to the fiscal system is of great concern to policy makers. As the differentiation of workers is also given by contributions to the social security system, the next section clarifies the costs and the benefits related to this contribution in Brazil.

3.1. Contributions to the Social Security System

Enrollment and contributions to the social security system are mandatory to all types of workers (self-employed or employees) receiving more than a minimum wage[13]. The advantages of contributing to the social security system in Brazil during the 1990s are social protection against impediments to working while the worker is in the labor force (wage compensations for sickness, accidents, or imprisonment; and maternity wages for women), and pensions when the worker is retired by age, disability, or time of contributions[14]. Until 1989 access to public health services in general was limited to workers (and their families) who were enrolled and contributed to the social security system. After 1989, however, with the creation of the Unified System of Health (*Sistema Único de Saúde – SUS)*, the government made public health services available not only to those contributing regularly to the social security system but to all individuals (universality principle). This resolution contributed to an adverse incentive for social security contributions during the 1990's as analyzed by Dart et al (2002).

Contributions are, on average, 20% of the declared contribution wage (for all types of workers). As self-employed workers do not have a regular wage, they have to declare one to the social security system and contribute (20% incidence) accordingly. Contributions of employees are also based on the average of 20% incidence. However, half (10%) are discounted from employees' wage and the other half (10%) is paid by their employer's, who are responsible for transferring

[13] Even if the worker receives less than a minimum wage he/she has to be enrolled in the social security system.
[14] Pensions are also provided to the family of the worker when he/she dies. Time of contributions is usually 35 years for men and 30 for woman.

contributions directly to the system. Contributions to the social security system are only a proxy for legal conditions, as it is the only available information qualifying the labor contract for all types of workers.

Excessive taxation on formal enterprises has often been pointed out as a very important factor inducing informality. Carneiro (1997) estimates payroll taxes costing around 43% of the base salary for an average employee[15] in Brazil. Outsourcing unregistered employees or self-employed workers, therefore, reduces substantially the cost of the labor force for formal firms. Induced also by some discredit toward fiscal authorities, the informality of the labor force became a regular hiring process adopted by firms of all sizes in Fortaleza city in Brazil during the 1990s. Thus, given their high opportunity costs, especially for small firms, social security contributions can be considered an additional and universal qualifier for different types of workers.

The possibility of differentiating workers by type and by social security contributions, as well as the possibility of identifying workers' neighborhoods, is a rare advantage of the data set used in this essay.

3.2. Data Set

The data set used was extracted from the Survey on Unemployment and Under-employment ("Pesquisa Desemprego e Subemprego"), which is a household-based survey conducted monthly by the National Employment System

[15] Social security contributions cost 20% of workers' base salary. The other 23% are related to many other different workers' benefits such as one additional monthly wage per year, unemployment security, and others.

with an office located in Fortaleza city Brazil (SINE/CE). The survey provides specialized labor market information at the individual and household levels, and it also allows aggregation measurements at the neighborhood level, as 101 different surveyed areas (called sectors) can be identified for each family. This spatial sub-division was derived from cluster analysis conducted during the 1980's which involved a set of demographic characteristics in contiguous areas. The survey does not cover all neighborhoods during each month of the year due to the limited amount of questionnaires distributed randomly to the families. Each month, 25% of the neighborhoods are substituted by others, in a way that each neighborhood is surveyed during the same months each consecutive year.

The full data set is composed by pulling all but two monthly cross-section surveys from 1993 to 2002[16]. The surveys from 1991 and 1992 were used only to estimate interaction effect variables for individuals surveyed in 1993 and 1994. The data set includes those 16-64 years old, who at the time of the survey were actively working for pay[17] or looking for a job, and did not have missing information for any variable included in the models. This results in 162,760 males and 117,554 females in the sample.

The available data set, therefore, allows the analysis of differentiation and qualification of workers, as well as the investigation of whether social interactions in neighborhoods are relevant conditionings in the labor market. These analyses

[16] The surveys of July of 1993 and June of 1994 were not included. The former presented different monetary values for the income of different families. This occurred because the whole country was switching the official currency at that time. The second survey was not available from the National System of Employment.

[17] This leaves out family workers who do not receive wages.

and investigations have been excluded from studies of labor market segmentation, which may compromise their reliability in offering policy guidance. If interactions affect labor force positions for workers, for example, they should also affect transitions across them. Consequently segmentation models based on transition rates (Maloney (1999), Neri (2002) or Gong and Van Soest (2002)) should also include the possibility of interactions at the risk of becoming misspecified.

The next section explains the ways in which the main issues of the informal and social interaction literature are addressed in this essay.

CHAPTER 4

EMPIRICAL METHODOLOGY

This essay uses a sector choice model to provide evidence that: i) there are significant differences among the same types of workers such as the usual formal versus informal classification should be avoided, ii) not all types of informal workers can be considered under-employed, and iii) interactions in neighborhoods and spatial segmentation can be important determinants of worker's position in the labor market. Segmentation models based on transition rates across sectors (formal versus informal) should be modified if these statements are verified empirically. These modifications should include at least the differentiation of informal workers and the inclusion of neighborhood effects into these models.

As panel data surveys are still missing for the labor market of Fortaleza, the modeling of transition rates is left for future work. However, the set of evidences suggested in this essay is suitable for investigation with a sector choice framework based on the estimation of multinomial logit models. The multinomial logit model has been widely used for this purpose by the informal sector literature (Hill (1989), Gingdling (1991), Pradhan and Van Soest (1995), Maloney (1999)) and is also referred to by Brock and Durlauf (2002) for the analysis of social interaction perspectives. Its unrestricted character does not assume any ex ante qualitative ordering for the different types of workers[18], which is actually the object of the analysis. In addition, its non-linearity is an important characteristic for the identification of neighborhood effects as shown by Brock and Durlauf (2002). One condition for this identification is that there should be no colinearity between worker's characteristics and the neighborhood effect indicators[19]. The authors argue that in linear and contemporaneous models[20] this condition may not be satisfied, as the aggregation of workers' characteristics at the neighborhood levels may coincide with the expected value for the social interaction variable. In this essay, the multinomial logit model not only brings non-linearity to the analysis but also the neighborhood effect variables ($C^{s}_{n(t-1)}$ and $Y_{n(t-1)}$) are based on non-contemporaneous values.

The graphing strategy for multinomial logit coefficients proposed by Long (1997) also facilitates ordering perspectives for the different types of workers proposed in this essay. This strategy consists of plotting the corresponding log-odds (and odds ratio or marginal effects) of workers' characteristics on the same axis (with the same unit of measurement), allowing magnitude comparisons for the different working positions. That is, it is possible to visualize and identify the

[18] Ordered probit models, for example, already assume a pre-determined ordering for each type of worker. See Pradham and Van Soest (1995) as an example.
[19] The other conditions for identification are that the data must contain: i) sufficient intra-neighborhoods variation for each type of worker, and ii) sufficient inter-neighborhoods variation for the neighborhood effect variables. Both conditions are satisfied in the data set used in this essay. There are 101 neighborhoods each one receiving a different value for each year. The accumulation of two (lagged) years to calculate the participation of each type of worker in the neighborhoods also guaranteed the intra-neighborhood variation of the workers.
[20] Contemporaneous models in this case refer to a social interaction model where the individual decision as well as the aggregated decision of his/her group of reference are taken at the same time (survey) period.

significant differences in the likelihoods of workers' allocation when one particular variable of interest changes.

Therefore, the multinomial logit model serves well for the purpose of analyzing workers' characteristics and the possibility of neighborhood effects. It represents a methodological intersection for the two types of literature cited in this essay (literature on informal economy and social interactions).

4.1. Worker Differentiation and Qualification

In order to verify formally significant differences between workers this essay compares individually the log-odds of different variables and also uses a test proposed by Long (1997) for combining two categories of workers in the multinomial logit model. This is a general Wald test which is based on log-likelihood comparisons of restricted (imposing equal coefficients for two categories of workers) and unrestricted models. The restricted model assumes that all the multinomial logit coefficients (with the exception of the constant) of two types of workers are equal. Long (1997) evidences that this test is statistically more powerful than the likelihood ratio test for multiple binomial logit (or probit) models.

In addition to differentiating workers this essay also qualifies them. The main objective is to observe an ordering that could identify those workers who look more like a segmented under-employed worker. This is done by investigating not only what types of workers are highly represented by less educated workers, but also examining how this representation compares to that of unemployed workers. Education is considered to be the most important filter for formal positions (Telles, 1992). It affects not only potential earnings but it is also an important reference for ordering preferences (Pradhan and Van Soest (1995)). The under-employed character of one position may be evident if more educated workers choose to stay unemployed instead of taking the available (informal) position. This comparison is in particular addressed to those small firm-covered and firm uncovered workers for which no prior expectation can be made with respect to the comparative group of unemployed workers.

Besides education, the other workers' characteristics analyzed are those usually included in traditional labor supply models: age, migration condition, and family characteristics. The main idea is to verify what types of workers also follow regularities associated with informal positions by the segmentation literature. A higher likelihood of informality for rural migrants, for example, is one usual assumption of segmentation models (Fields, 1975). Also following the staging hypothesis of Fields (1975), young workers should be more likely to be part of the informal sector as they lack labor market experience.

In addition, informal positions are also said to be compatible with home activities for females (Maloney, 1999). In this case the presence of children would not represent an almost prohibitive opportunity cost to participate in the labor market as is usually observed in traditional labor supply models. Finally, others' family income is also a common variable present in labor supply models. It is usually expected that the reservation wage for an individual would increase with a higher income for other family members. This variable, therefore, would represent a great indicator of quality for each working position in the labor market. It is, however, subjected to significant measurement errors as informal workers tend to report profits of small business rather than labor earnings. In addition, formal workers do not usually report other types of benefits not included in wages. In these cases, informal wages are overestimated and aggregation at the family income level only amplifies the measurement errors for this variable. Some precaution, therefore, is devoted to the results of this variable.

4.2. Neighborhood Effects and Spatial Segmentation

This essay investigates whether worker's interactions can be important in influencing their working positions. The empirical verification of the importance of social interactions, however, is not as simple as the analysis of α's in theoretical model 1. As stated in the literature review section, three usual problems have to be addressed in order to isolate social interaction perspectives: neighborhood sorting, simultaneity, and contextual effects.

This essay makes use of time-lagged values for the social interaction variables as a solution to avoid the simultaneity (endogeneity) problem when individuals' decisions are at the same time affecting and being affected by others' decisions. The time lag suggests that individuals at the time of their decision are aware of the given social context and act upon it. This solution is only possible as the data set is rich enough to track neighborhoods (but not people) across time, which is not usual in traditional labor market surveys.

With respect to the contextual (local) effect of the neighborhoods, this essay follows Carman (2004) by including fixed effect controls for areas broader than the neighborhood size in order to capture local conditions. At a different level of aggregation, fixed effects do not overcome the interaction effect and yet capture specific local conditions within an urban labor market. The broader areas in this study refer to the six areas covered by the six regional offices of the municipal administration. These areas are proxies for identifying local conditions in the labor market as their populations have common public demands met (or not) by the same office. Although this common characteristic does not necessarily relate to local labor market conditions it can generate spontaneous interactions among their populations.

The sorting problem occurs when the formation of the worker's group of reference (neighborhoods) is not random. In this case, a significant value for α's in model 1 can be reflecting the tendency to workers of the same type cluster themselves into neighborhoods rather than capture interaction perspectives. In order to control for this possibility, this essay assumes that most of the sorting across neighborhoods in the labor market of Fortaleza is induced by income determinants. This is because ethnicity is considered to be of less relevance than it is in developed countries with relatively high inflows of foreign workers forming ethnic neighborhoods. Therefore, the median value of neighborhoods' family income lagged one period is included in the model to capture the sorting effect and isolate the interaction effect.

The final multinomial logit model of labor supply and sector choice is therefore defined as:

(2) $$\Pr(C_{int} = j) = \frac{\exp(K_j' Z)}{\sum_s \exp(K_s' Z)}$$

$$\text{with } K'Z = \delta' X_{int} + \Sigma_s \alpha^s C^s_{n(t-1)} + \theta Y_{n(t-1)} + \lambda' R_N$$

Where $Y_{n(t-1)}$ is the median value for neighborhood's family income, R_n are fixed effects dummy variables for the six administrative regions, and the other variables are defined as before. The vector of worker's characteristics (X_{int}) includes: education variables, age, age squared, rural migration indicators, family characteristics, family income, and year dummy variables. The definition of each variable included in the model is shown in Table 4.

The inclusion of $Y_{n(t-1)}$ and R_n in the model is an attempt to control the sorting and contextual problems discussed above, and consequently extract a "clean" interaction effect from α^s. However, this essay recognizes in accordance with the literature, that this task may not be accomplished completely. This is because the process of sorting into neighborhoods of one type of worker is not random, and its modeling alone deserves further examinations. In this case, the empirical findings are interpreted as suggestive rather than confirmatory of the existence of interaction effects in the labor market. In addition, as this essay aims to provide insights to expand labor market segmentation models, it is important to evidence that neighborhood effects in general are significant factors explaining worker's position in the labor market. If this is the case, their exclusion in segmentation models may compromise the findings of these. In addition, it is possible that neighborhood sorting and interactions reinforce each other, driving a segmentation process that occurs at the spatial level.

In order to investigate if neighborhood effects in general are important in defining workers' position in the labor market and if social interactions are part of such effects, this essay compares the estimations of α in theoretical models 1 and 2. In model 1, α captures neighborhood effects in general as there are no controls for sorting or local effects. That is, α is the mixed effect of workers' sorting, interactions, and local constraints, which form the general (or net) neighborhood effect. If α is significant in models 1 and 2, this is an indication that neighborhood effects are important in influencing workers' position in the labor market, and that social interactions are probably an important influential factor.

The existence of four different types of workers plus those unemployed in the multinomial logit model allows a very interesting analysis of interactions between workers of the same and different types. Although the interpretation of multinomial logit coefficients is not straightforward for this purpose, their corresponding marginal effects for an average worker help to clarify their meanings (Long, 1997). There are sixteen corresponding marginal effects of interest (μ^s_j), each one capturing the effect that the relative prevalence of one type of worker j over those unemployed (omitted category) in the neighborhood has on the likelihood of the worker be of type s. Table 5 shows the μ^s_j for which significant values suggests neighborhood effects (model 1) or interactions (model 2) between the same and different types of workers. In the main diagonal the effects occur between the same types of workers, and off diagonal they occur between different types of workers.

Therefore, it is possible to analyze if different types of workers tend to segment or integrate themselves into neighborhoods. Spatial segmentation or spatial integration depends on the sign and value of the marginal effects. When the estimates of the main diagonal have a positive sign and the elements off diagonal are negative and significant there is a trend for spatial segmentation. When the marginal effects off diagonal have a positive sign also the neighborhoods tend to become more diversified and consequently there is more spatial integration between the different types of workers (Table 5).

Although the multinomial logit model provides good reference for comparing working positions and testing neighborhood effects, it relies on the assumption of independence of irrelevant alternatives (IIA). This assumption says that the comparative odds of any two alternatives are independent of the existence or not of other alternative choices. With some degree of controversy, the hypothesis of IIA is said to be used only in cases where the choices are plausible differentiated by the researcher (McFadden, 1984). Given this is actually what has to be tested; it is thus recommendable to estimate alternative approaches to providing robustness to the findings of the multinomial logit model.

This essay offers two approaches to reinforce the findings of the multinomial logit model with five types of workers. In the first, the multinomial logit model is used but only for four types of workers. Unemployed workers are excluded from the sample. The objective is to observe if the ordering found with five types of workers is preserved if one type is excluded. Also, restricting the sample to workers, this essay also uses the bivariate probit model as a complementary specification model to compare the different definitions of workers. The main advantage of this specification is that it provides a basic test to compare if the two most important definition criteria for informal workers (firm size versus social security contributions) can be used separately as they have been used in the literature. The test consists of verifying if the residuals of each set of working choices (small versus firm workers and covered versus uncovered workers) are significantly correlated. In positive cases, it is inadequate to estimate the model using one or other definition. Consequently, aggregations should be avoided in segmentation models. The bivariate probit model with neighborhood effects is also an innovation with respect to Devaney and Chien (2000) which confront participation in retirement plans for self-employed and wage workers using a bivariate probit model. In this respect, he finds that unobservables that affect the self-employment choice also affect participation in retirement plans. This essay verifies if the same can be said to the labor market of Fortaleza, Brazil.

CHAPTER 5

EMPIRICAL FINDINGS

The empirical analysis that follows is divided into two parts. The first is more descriptive and identifies in isolation the characteristics of the different types of workers in the labor market and their evolution during the 1990's. The second part identifies if the results found in the descriptive analysis persist in a multivariate framework. Each part is also divided between the analyses of workers' differentiation and interactions.

5.1. Descriptive Analysis

5.1.1. Worker's differentiation and qualification

The reduction in the share of firm covered workers[21] during the 1990s (Graphs 1 and 2) was compensated by the increase in the share of unemployed and small-firm uncovered workers. This is to say that the informalization process in the 1990s was induced by one type of informal worker that looks like an under-employed worker. This conclusion comes from the definition of under-employment suggested in this essay which involves those workers with education achievements lower than those of unemployed workers. This is verified

[21] It is important to re-emphasize that this is the only type of worker considered formal according to local differentiation of formal and informal workers. Therefore, small-firm uncovered, small-firm covered, and firm uncovered workers are all considered informal.

in Graphs 3, 4, 5 and 6 where those less (more) educated workers are relatively more (less) represented by those small-firm uncovered workers. A closer analysis of these graphs also shows two interesting results.

First, the percentage of workers with secondary education (no education) increased (reduced) for all types of workers and for both genders. This is an indication that a comparative reduction in the "quality" of informal workers in the labor market was not responsible for inducing its informalization during the 1990's.

Second, male and female workers present different comparative orderings when small-firm covered and firm uncovered workers are considered. That is, for males, small-firm covered workers should not be considered an under-employment position according to the definition used here. And following the same definition, firm covered workers should be considered in general an under-employment position. For females, it is the reverse. While small-firm covered workers should be considered under-employed, firm uncovered workers should not be considered under-employed. This gender difference may suggest that whereas females may have higher preferences for employment opportunities in traditional firms, males may be more willing to work in positions contributing to the social security system independent of the size of the firm. That is, if one should aggregate workers in two sectors (formal versus informal), the firm size definition would work better for females and the social security contribution would be more precise for males. However, this essay suggests that more types of workers' should be differentiated to provide more precise and/or reliable evidence of labor market segmentations.

The aggregation of workers in only two categories would not be referred if there were significant differences between them. This is what is shown in Tables 6 and 7. Almost all workers' characteristics representing human capital and behavioral variables present significantly different mean values across all types of workers. The three categories of informal workers according to the local definition (SF-UCOV, SF-COV and F-COV) also present significant differences for most of their characteristics. In addition, focusing only on the educational variables, Tables 6 and 7 reinforce that male and female workers differ with respect to the "under-employed" character of their middle positions (SF-COV and F-UCOV). That is, for male workers with no education, the participation of SF-COV workers in this educational category (50%) is lower than those of unemployed workers (53%), which is lower than those of F-UCOV workers (62%). For female workers this ordering is inverted (SF-COV (48.1%), unemployed (37.0%) and F-UCOV (36.5%)).

Other patterns of worker characteristics are worthy of mention. When compared to the unemployed, those firm covered workers follow traditional patterns of labor supply. That is, the comparative likelihood of having work (and consequently not looking for a job) is higher if one is the main provider of the family and if he/she is older. The presence of children reduces the likelihood of labor supply as a firm covered worker for females and increase for males. With respect to the worker's migration conditions there is no clear pattern that makes labor supply more likely if the worker is in any specific stage of adaptation in the labor market.

Significant differences for the average value of the income of other family members can also be observed in Tables 6 and 7. This result is also an indication that labor supply decisions within the family are different for the different types of workers.

Both genders had a similar pattern of labor supply with respect to age and being the main provider of the family. That is, on average, there are more older and household head workers working as SF-COV, SF-UCOV, or F-UCOV, than being unemployed. However, their differences in likelihood under the descriptive analysis are not sufficient to assert that these types of (informal) workers have to be treated differently. Therefore, a confirmatory analysis using a multivariate set up is needed to reinforce such a conclusion.

5.1.2. Neighborhood effects and spatial segmentation

In order to investigate spatial aspects of the distribution of workers in this descriptive analysis, the shares of workers for the neighborhoods were used as units of analysis, and spatial correlation tests were performed with the use of Moran's I index. In the present context, this index measures the degree to which neighborhood areas present similarities in the prevalence of each type of worker. That is, it is a proxy for the clustering of the different types of workers into areas formed by contiguous neighborhoods. Higher and significant values for the Moran's I index mean higher levels of clustering[22].

When the shares of each type of worker are calculated for the whole period for each neighborhood, a significant level of spatial correlation was observed (Table 8)[23]. This is an indication that the same types of workers tend to live in nearby neighborhoods. In order to observe how the prevalence of informal workers (or part of it) relates to their spatial concentration this essay compares the evolution of the Moran's I and the evolution of the share of informal workers for the whole city during the 1990's[24].

It is interesting to note that the prevalence of informality (share of SF-COV and SF-UCOV and F-UCOV workers) and its spatial concentration had similar time trends over the period (Graph 7). That is, when the share of informal workers increased they became more spatially concentrated. The same can be said when only the share of SF-UCOV workers is considered (Graph 8)[25]. Although this evidence occurs at the neighborhood level, it can also be an indication that neighborhood effects (clustering and/or interactions) play a role in workers' decisions in the labor market.

---

[22] The Moran's I index range in values from -1 (negative perfect spatial correlation) to 0 (no spatial correlation) to 1 (positive perfect spatial correlation).

[23] It is important to emphasize that some conditions were imposed in constructing the spatial matrix of neighborhoods. It was not possible to identify the exact positions of each of the 101 neighborhoods in the city. But it was possible to identify the barrios in which the neighborhoods were located, and these geographical areas (barrios) can be identified in maps. In this case neighborhoods were treated as barrios to generate a simple contiguity spatial matrix. The worker's share by neighborhood for the whole period goes from 1993 to 2002.

[24] It is important to re-emphasize here that the use of the concept of informal workers at this stage follows the local definition which is based on the aggregation of small-firm covered, small-firm uncovered, and firm uncovered workers. The neighborhood shares and the workers shares were calculated using a three year moving average for each one. The three year accumulation was used to avoid great variations in the shares of workers for small neighborhoods.

[25] The share of SF-UCOV workers alone was analyzed because it is the most dynamic part of the informal sector as it was seen in Graphs 1 and 2.

In this descriptive analysis, therefore, it was observed that: (1) there are significant differences in the characteristics of the different types of (informal) workers, (2) males and females present differences in terms of "under-employment" conditions regarding the SF-COV and the F-UCOV positions in the labor market, and (3) at the neighborhood level, there is evidence of spatial segmentation as the level of clustering increases with higher prevalence of informal workers. The next section investigates if such conclusions stand up in a multivariate approach.

5.2. Multivariate Analysis

5.2.1. Worker's differentiation and qualification

The multinomial logit model produces multiple coefficients representing many relative comparisons between the log odds of two different categories of workers. One base category serves as standardized reference for these comparisons, for parameters identification, and also for expositional purposes. As this essay qualifies workers by comparing their educational achievements with those of unemployed workers, it is very convenient therefore to use these last types of workers as the reference base. This way it is possible not only to qualify workers with respect to their under-employment situation, but also the logit coefficients can be interpreted as a labor supply likelihood which is traditionally referred in the literature (Pradhan and Van Soest, A. (1995), Magnac (1991), Gallaway and Bersanek (2002)).

In order to provide ordering references and marginal effects for the different types of workers this essay uses Long (1997) graphing strategy for multinomial logit models[26]. The graphs compare significant differences between any two categories of workers by linking the labeled categories if their coefficients are not significantly different.

The log odds coefficients and their corresponding marginal effects are shown in Tables 9 and 10 for males and females respectively. There are two models specifications in each table. Models with and without neighborhood effects are estimated in order to analyze if there are significant differences in the coefficients of workers' characteristics[27].

The general differentiation of workers is analyzed with the Wald test shown at the end of the Tables 9 and 10. The significant values show that the differences in the characteristics of the workers are such that would not be adequate to join any two pair of workers' categories in only one. In this case, neither the size of the firm nor the contribution to the social security system is a complete criterion to separate formal and informal workers as it has been done in the literature (Marcoullier et al (1997), Pradhan and Van Soest (1995), Gong and Van Soest (2002)).

The differences between workers' characteristics are clear when analyzed separately. Figures 1 to 12 present Long's strategy of plotting odds ratio estimates related to those of Tables 9 and 10. Each row in the figures represents

[26] The "ado" command "mlogplot" prepared by Long (1997) was used for this graphing strategy. This "ado" command is public available for Stata users.
[27] The coefficients for the social interaction variables in the model with neighborhood effects are analyzed separately in Tables 11 (males) and 12 (females) respectively.

the factor change in the comparative odds of unemployment with respect to any other particular working position, when the selected (row) variable changes by one unit (or by one category in case of dummy). The comparative odds are negative (positive) for the positions located to the left (right) of the unemployment state of reference (0). In this way, when a male worker gets primary education (as compared to those with no education – the omitted category) his odds to be a small-firm uncovered (SF-UCOV = 1) or a firm uncovered worker (F-UCOV = 2) reduce significantly by a factor of 0.742 and 0.738 respectively, when compared to the possibility to stay unemployed (Figure 1). The line linking those two categories of workers (SF-UCOV =1 and F-UCOV = 2) show that having a primary education does not make their own comparative odds significantly different. In the opposite side, the comparative odds to get a position as a small-firm covered or firm covered worker increases by a factor 1.36 or 1.28 respectively when the worker gets primary education.

From the analysis of educational variable for males and females (Figures 1 to 4) three important findings arise. First, there are no significant differences in the estimates for models with or without neighborhood effects. This analysis is the same for males (Figures 1 and 2) and females (Figures 3 and 4), implying that neighborhood effects are complementary explanations for the allocation of workers in different working positions.

Second, as the level of education increases it is possible to observe bigger disparities between the comparative odds for the different types of workers. The impact of education on the comparative odds can be visualized by the distance between each position given the figures present the same scale in terms of factor changes in odds ratio for all variables. As the workers acquire primary, secondary and college education, the distance between the positions, and especially between F-COV (4) and SF-UCOV (1), increases. This means that the education is a clear differentiator of workers, in accordance with the qualitative criterion proposed in this essay.

Finally, the ordering and the qualification of workers differ for males and females with respect to the intermediate positions (F-UCOV (2) e SF-COV (3)). For males, small-firm uncovered (1) or firm uncovered (2) workers can be considered under-employed positions as more education makes workers more likely to stay unemployed than to consider one of these working positions[28]. Small-firm covered workers have a positive likelihood of labor supply when their educational level increases. This is to say that not all types of informal workers according to the local definition have matching characteristics with segmented involuntary workers. The aggregation of different types of workers under one definition/premise in this case is not adequate when the size of the informal economy is a proxy for the distortions in the labor market (segmentation theory).

For females, aggregation is also inappropriate but the ordering of positions with respect to education is different when compared to male workers. Education is positively correlated to labor supply for firm workers (covered (4) or uncovered (3)) and negatively correlated with respect to small firm workers (covered (3) or uncovered (1)). This different ordering for females also suggests that education

[28] The exception occurs when male workers get college degree. In this case, a positive likelihood of labor supply also applies to firm uncovered workers.

improves comparatively the chances of getting positions in traditional firms more than working positions, which contribute to the social security system. If this comparative restriction is driven by stigma and/or discrimination is an analysis that goes beyond the objective of this essay and is left for future works.

With respect to the analyses of family characteristics for males (Figures 5 and 6) and females (Figures 7 and 8) workers, they present usual features. That is, labor supply in any position (with exception of Small-Firm Covered (3)) is more likely to occur for males and less likely to occur for females when there are children at home. This is a usual regularity in labor supply models (Pradhan and Van Soest (1995) or Gallaway and Bersanek (2002)). There is however, a positive significant likelihood of labor supply for small-firm uncovered female workers in the presence of young children. This is consistent with the view that this type of informality may be desirable for females as it allows working schedule flexibility to be compatible with home activities.

Labor supply for males is very likely to occur for any type of working activity when the male is the household head. Females have the same qualitative result with the exception that labor supply as firm covered workers is not likely to occur for females heading their homes. This may also be an indication of preference for informal positions (SF-COV, SF-UCOV, or F-UCOV) as they can generate a comparative lower opportunity cost for family care.

No definite pattern could be extracted from the analysis of migration. Neither males (Figures 9 and 10) nor females (Figures 11 and 12) presented a consistent pattern where working positions would be more likely to occur than unemployment for recent migrants. This result is not in accordance with the staging hypothesis of Fields (1975) where the transition from informal to formal jobs comes with a better knowledge of the labor market.

It is common to assume that the income of other family members has a negative and significant impact on the likelihood of becoming a worker especially for females[29]. This result, however, may vary with the level of the income and the role play by the females in the family. Pradhan and Van Soest (1997) shows that labor supply probabilities for females can be higher for low and high levels of family income as their willingness to work can represent respectively a necessary income complement or a personal matching of couples' characteristics [30]. Interesting to notice (Table 10) that only small-firm uncovered female worker had a significant and negative likelihood of labor supply when the family income increases. This unexpected result may also be caused by error measurements in the income variables for informal workers as they may report revenues rather than wages for their services/activities. A relatively higher informal wage may increase the reservation wage for females to work, reducing the likelihood of labor supply. Due to this possibility of error measurement in income, however, this essay skips further interpretations for these related coefficients and focuses attention only on the differentiated effects among the different types of workers, which is consistent with the plea for a more expansive classification of workers.

---

[29] Traditional here refers to models where the distinction between formal and informal jobs is not considered, and only the first type is included in the sample.
[30] High educated workers tend to married themselves, which makes the likelihood of labor supply for females positive and significant.

In sum, the suggestive findings of the descriptive analysis of workers' characteristics also stand in a multivariate framework. First, there are substantial differences between workers considered to be informal by local definition, or by general classifications in the literature. Secondly, males and female workers differ with respect to how education affects ranking positions and their under-employed character. These findings should be taken into consideration in order to lend more robustness to the analysis of labor market segmentation.

5.2.2. Neighborhood effects and spatial segmentation

The analyses of neighborhood effect indicators in worker allocation are based on theoretical models 1 and 2 shown in Tables 11 and 12 for males and females respectively. The coefficients show the comparative likelihood that a worker will be in a particular working position if the prevalence of workers of determined type increases in his/her neighborhood. In theoretical model 1 this neighborhood effect is unspecified and is driven by the possibilities of social interactions, sorting of workers, and local conditions. Theoretical model 2 is an attempt to isolate the social interaction effect by including controls for sorting and local effects. Although this essay shares Manski's (1993) skepticism about this possibility, the comparison of models at least provides evidence favoring (or not) the existence of interaction effects, and also serves as a robustness check for this traditional (variable) indicator of social interaction.

The coefficients for the social interaction variables are shown in Tables 11 and 12 with own and crossed interaction effects[31]. The elements in the main diagonal (in bold) represent own effects. In models without controls for sorting and local effect (model 1), all types of effects between workers of the same type are significant with the exception of firm covered male workers. The same pattern is also observed for females meaning that the comparative likelihood of leaving unemployment in favor of any working position increases if the prevalence of workers of the same type expands in the neighborhood.

Looking at the marginal effects (in brackets) it is clear that small-firm uncovered workers have the highest own neighborhood effect. If the percentage of small-firm uncovered workers in the neighborhood increased by 1 point in the last two years[32] the probability of finding a worker of the same type (SF-UCOV) this year increases by 0.0035. The smallest effect for theoretical model 1 occurs for firm workers (-0.0009) which is consistent with Granovetter (1995) who observed that interactions are usually stronger among less educated workers.

Most of the crossed neighborhood effects are positive when the logit coefficients are considered. This indicates that workers prefer to work in any position than staying unemployed when the general neighborhood level of occupation rate increases. When the marginal effects are the reference guide, however, it is possible to identify a negative relationship between small-firm

---

[31] The other coefficients and marginal effects for theoretical model 1 are shown in the following tables 12 and 13 for males and females respectively. The other coefficients and marginal effects for theoretical model 2 are those corresponding to the model with neighborhood effects of tables 1 and 2.

[32] The marginal effect for each variable is calculated holding all other variables at their mean value.

uncovered workers and firm covered workers[33]. This spatial pattern seems to be following the informalization trend of the 1990's (Graphs 1 and 2) where these two types of working positions are negatively correlated. That is, segmentation rather than integration at the spatial dimension is occurring for these two types of workers.

With respect to the relationship between the other types of workers, it is more likely that spatial integration rather than segmentation be the norm, especially between small firm covered and uncovered workers. This is easier to observe in Figures 13 and 14, which show the change in predicted marginal effects for each type of worker when the prevalence of workers change by one unit. To the right (left) of the division line (at zero value) we have positive (negative) marginal effects for each row variable. Small firm uncovered workers, firm uncovered workers, and small firm covered workers are on the same (positive) side of the figure for each of their respective neighborhood effect variables. This suggests that more workers of these types tend to cluster themselves into neighborhoods.

When theoretical model 2 is considered, two important results arise. First, many (own and crossed) coefficients considerably reduce their magnitude or even become insignificant. This is an indication that most of the neighborhood

[33] It is important to reinforce here the difference between the logit coefficients and the predicted marginal effects. The log-odds indicate how the likelihood of each working position with respect to unemployment change when each explanatory variable change. This effect hides the direct relationship between each type of worker and their prevalence in the neighborhoods. In this case, the marginal effects seem to be appropriate for this task as they provide this direct relationship. Their computation however depends on all individual characteristics (taken at their mean values) and the magnitude of the change for the explanatory variables (assumed to be 1 unit). Because of this dependence the marginal effects do not necessarily have the same sign of the logit coefficients (Long, 1997).

effects are more related to the contextual effects of sorting and local conditions than to the possibility of social interactions. The marginal effects are also smaller for model specification 2. Figures 15 and 16, for example, show that the predicted marginal effects are comparatively closer to 0 for this modeling perspective, attesting to the reduced effect of social interaction variables after controlling for characteristics of neighborhoods.

Secondly, with respect to own effects, it is observed that those related to small-firm uncovered and covered workers continue to be significant after contextual control, which is an indication that social interactions may be important for these types of workers. Their crossed relationship also confirms spatial integration between these two types of workers. This result applies to males and females.

In sum, the analysis of neighborhood effects in the labor market of Fortaleza showed that: i) neighborhood effects are important determinants of workers' allocation in the labor market, ii) spatial segmentation is more likely to occur between small-firm uncovered and firm covered workers, and iii) spatial integration as well as social interactions are more likely to occur between small firm workers (covered or uncovered).

Thus, labor market segmentation models based on transition probabilities should not only disaggregate formal or informal workers at different levels, but should also consider the possibility that neighborhood effects are important determinants of such transitions.

5.3. Robustness Tests

The findings of workers' differentiation and significant neighborhood effects can be subjected to the particular definition of workers used so far and to modeling perspectives. This section, therefore, provides some robustness tests against some of these possibilities.

5.3.1. Workers' Definition

Small firm workers have been defined here as the union of those self-employed workers, employees and employers in firms with 1 or 2 workers, and domestic workers. On the other hand, firm workers are those employees and employers working for firms with 3 or more workers, plus those liberal professionals. This section analyzes whether or not the results found based on these definitions are sustained if only self-employed workers are considered small-firm workers. Liberal professionals, domestic workers and firms with less than three workers are left out of the sample in order to provide a direct comparison between self-employed and traditional wage workers.

Aggregation is still inadequate for this restrictive definition of small-firm workers as can be seen at the bottom of Tables 15 and 16. With respect to ordering, small-firm covered workers increased their comparative log-odds as educational level increases (Figures 17 and 18). This is even more evident for females. This could be expected as most of the excluded domestic workers are females and have low educational levels. Thus, there is even stronger evidence that not all informal positions as locally defined can be considered under-employment.

The results regarding significant neighborhood effects are also maintained, with spatial segmentation perspectives occurring especially between small-firm uncovered and firm covered workers (Tables 15 and 16). Figures 19 and 20 confirm that the marginal effects for the social interaction variables do not change significantly when compared to the models with all types of workers (Figures 15 and 16). Therefore, by comparing only self-employed workers to traditional wage workers this essay reaches the same demanding conclusions that more differentiations for formal and informal workers and inclusion of neighborhood effects are necessary actions for one to test labor market segmentation.

5.3.2. Modeling perspectives

In this section two different modeling perspectives are used for robustness tests. In the first, a multinomial logit model is estimated excluding unemployed workers from the sample. As the multinomial logit model assumes that the comparative odds of two options are independent of the existence of other options (IIA assumption) it is expected that the results will hold. The second modeling perspective is a bivariate probit model where the decisions regarding the type of worker (small-firm or firm worker) and the contribution to the social security system (covered or uncovered workers) are estimated as simultaneous and correlated decisions.

The ordering and qualitative analysis for male workers remains approximately the same when unemployed workers are taken out of the sample (Figure 21). The same can be said with respect to females (Figure 22). As the values of the log-odds are approximately the same for models excluding (Panels B) or not (Panels A) unemployed workers, this is also an evidence in favor of the IIA assumption and the respective use of the multinomial logit model[34]. The possibility of social interactions and the kind of relationship between the workers at spatial level are also consistent with the previous findings. As can be seen in Tables 17 and 18, as well as on Figures 23 and 24, spatial segmentation perspectives are more likely to occur between firm covered workers and small firm uncovered workers.

The results for the bivariate probit model are shown in Tables 19 and 20 for males and females respectively. Two important tests were performed and shown at the end of each table. The first tests if the two decisions (worker's type and social security contributions) can be taken separately. In other terms, this is a test to verify if the two usual conceptualizations of (informal) workers can be used separately. The significant value for ρ means that there are correlated and unobservable characteristics for the workers, which makes the two decisions inseparable. In this case, a dichotomy of workers (formal versus informal) is not appropriate.

[34] It is important to mention that the results for two tests performed to verify if the IIA assumption holds for model specification 2 were contradictory. One based on Hausman and McFadden (1984) was for the IIA and the other based on Small and Hsiao (1985) was against the IIA assumption. The tests were performed on Stata 8 with command "mlogtest".

The qualitative analysis for the four types of worker cannot be performed with the bivariate probit model. However, it is possible to test if education represents a similar constraint for being a firm worker or for contributing to the social security system. That is, it is possible to verify if the comparative log-odds for being a firm worker (versus a small firm worker) are the same for being a worker contributing to the social security system (versus a non-contributor) when the educational level changes. The significant and progressive coefficients for the educational variables in Tables 19 and 20 confirm that education is an important factor for finding positions as firm workers and also for finding positions which pays social security contributions. Comparing the educational coefficients for each position and gender it is interesting to note that males and females have different results. The tests at the end of the tables show that education for males improves comparatively more the odds for being a social security contributor than the odds for working in traditional firms. The exception is for a college graduated worker for which both comparative likelihoods of finding positions are equivalent. For females, the reverse occurs. With more education, it is comparatively easier to find a traditional firm worker female, than to find a female worker contributing to the social security system. This result also confirms the anterior findings related to the multinomial logit model.

Some social interaction indicators are also significant in the bivariate probit model. Their negative coefficients sustain the evidences in favor of spatial segmentation between small firm and traditional firm workers, and for higher possibilities of social interactions among neighbor workers with low level of human capital. This is again consistent with the earlier findings of theoretical model 2, and reinforces the necessity for further investigations on the transmission mechanisms of the neighborhood effects.

In sum, worker's differentiation and neighborhood effects are important pre-requisites for any attempt to test labor market segmentation between formal and informal workers in Fortaleza City. In addition, the local definition of informal sector has aggregated different types of workers who do not necessarily fit the concept of under-employed worker. Erroneous generalization of workers may also distort policy perspectives when the labor market looks worse than it is.

CHAPTER 6

CONCLUSION AND FURTHER DISCUSSIONS

The rise of the informal economy worldwide has intensified discussions about the possibility and consequences of labor market segmentation. Although new methodologies and data sets have improved the empirical debate on testing segmentation in different labor markets, some old and new concerns have passed almost unattended. First, the differentiation between formal and informal workers has always been controversial. Although different approaches have been used, the dichotomy on the definition of workers is still present in many studies.

This essay showed that the dual definition of workers (either formal or informal) may not be appropriate as those consider formal or informal present significant differentiations among themselves. For the labor market of Fortaleza, the combination of dual definitions based on firm size or social security contributions produced four types of workers (small-firm uncovered, small-firm covered, firm-uncovered and firm-uncovered) for which allocation likelihoods varied significantly with respect to their own, family and local characteristics.

In addition, when unemployed workers were taken as reference it was observed that not all types of informal workers according to the local definition can be considered under-employed, as they have comparatively higher allocation likelihoods with higher levels of education. This is the case for small-firm covered males and firm uncovered females. Consequently, segmentation studies, especially using transition likelihoods between formal and informal working positions, should consider additional qualitative differentiators for those positions, which can be linked to different sources of segmentation.

This essay provided evidences that the size of workers' firm and their social security contributions represent two different but correlated criteria of differentiations. Their separate use to qualify workers as have been done in the literature (Marcouiller et al (1997), Funkenhourser (1997), Gong and Van Soest (2002)) showed to be inappropriate for the labor market of Fortaleza City in Brazil.

Finally, worker's allocation and transition by extension may also be affected by the clustering and interactions of workers into neighborhoods. This essay raised this possibility by finding that the likelihood of a worker being of a certain type is significantly affected by a higher prevalence of workers of the same type in his/her neighborhood. This effect, however, was more prominent and positive among informal workers with the lowest educational level, which is consistent with Granovetter (1995) who observed that networking is usually stronger for low quality jobs.

A negative correlation was also observed between the prevalence of informal workers (small-firm uncovered workers) and the likelihood of finding traditional formal workers (firm covered workers) in neighborhoods. This result favors the possibility of segmentation at least at the spatial (neighborhood) level.

The type and existence of a relationship between spatial segmentation and labor market segmentation is an analysis left for futures studies as it goes beyond the possibilities of this essay. However, as the basic objective of this essay was to provide additional methodological subsidies to labor market segmentation studies some concerns can be advanced here. It is important re-emphasize, for example, that social interaction effects are difficult to identify when the interactive group of reference (neighbors) is not randomly formed. This essay used neighborhood aggregated variables (family per capita income for each neighborhood) and local fixed effect proxies to control for the sorting of workers of the same type into neighborhoods. Making the sorting process endogenous is another possibility that has been suggested by the social interaction literature (Brock and Durlauf (2002)), in order to allow the identification of social interaction effects among other neighborhood effects.

In segmentation studies based on transition probabilities between formal and informal workers it may be interesting to construct interaction variables also based on the other's transitions rather than on the prevalence of determined type of worker. That is, the interaction variable may be constructed based on the percentage of workers changing positions rather than on the percentage of workers in a determined position. This type of segmentation study however demands panel or at least retrospective data[35] sets which are not available so far for the labor market of Fortaleza City, Brazil.

[35] Retrospective data set are those where at time t the individual answer questions about his/her condition on period t – 1.

The presence of social interactions among workers at the neighborhood level also demands investigation about the mechanism through which the interaction may occur. If it is informational, for example, it is important to verify if this interaction is a substitute or a complement for imperfect information about opening positions across the whole city. When the main vehicle of information about job openings for an informal worker is another informal worker in the neighborhood then one can expect informality to become contagious. The consequences of this proliferation of informality depend on the role that informal workers have in the labor market for which qualification comes with the segmentation analysis. If the segmentation of the labor market occurs together with a spatial segmentation then it is important that corrective policies also include local (or spatial) perspectives.

Finally, the interesting finding of spatial segmentation of informal workers also opens possibilities for further investigations into its relationship with other social outcomes such as criminality, corruption, and other deviations from social norms.

Table 1 - Segmentation Studies Using Earnings Comparison Methodology

Study	Methodology	Worker Categories	Data selection/age group / exclusions/data type/sites data Type/Site(s)	Main Findings (Segmentation theory)
Hill (1989)	3sls for earnings and hours of work with self-selection correction from labor supply and sector choice model	- Informal: Family workers. - Formal: Employees. - Not working.	Married women (with spouse present)/ 20 to 59 years old/ N.S./Cross-Section / Tokyo Metropolitan Area (Japan)	"...this analysis clearly confirms the importance of treating separately (...) the two market sectors for countries (...) characterized by large informal sectors."
Tannen (1991)	Earnings equation with dummy for formal workers. Self-selection correction from sector choice model	-Informal: Self-employed + employees without social-security contributions. -Formal: Employees with social security contributions.	All workers with positive earnings / 15-65 years old/ N.S./Cross Section / Northeast States of Brazil.	"But contrary to the dual market model, controlling for the skills of workers serves to eliminate completely de difference in earnings between the formal and informal sectors of the private nonfarm urban economy."
Magnac (1991)	Potential wage differences from a generalized bivariate Tobit model.	- Informal: Self-employed. - Formal: Wage workers. - Non-participants. - Unemployed.	Married women (with spouse present)/ 18-60 years old/Employers and family worker/Cross-section/ 7 Urban Areas of Colombia.	"...the assumption of competitive markets seems to be an accurate description of the labor market."

Cont. Table 1

Gingdling (1991)	Earning equation comparisons. Chow test for expected wage decomposition with sector selectivity control.	- Informal private: worker not in formal private or public sectors, working for a nonservice firm with manual or no-machinery, working in a house or street, or is a domestic worker. - Formal private: Professionals in occupations associated with government sponsorship or not + workers belonging to cooperatives, unions or any professional organization. - Public: workers directly employed by the central government or working for semiautonomous enterprises.	Workers with positive monthly income/N.S./N.S./ Cross-section/ Metropolitan Area of San Jose (Costa Rica).	"All the results presented in this article are consistent with the hypothesis of labor market segmentation between the public and private-formal sector (...).The evidence in this article also supports the conclusion of labor market segmentation between the private-formal and informal sectors (...)."
Telles (1993)	Earnings differences without self-selection control from choice models	-Informal protected self-employed: the self-employed who contributed to social security system. - Informal unprotected: employee or self-employed without contributions to the social security system. - Informal domestic employees. - Formal employees.	Workers w/ positive income/ 10-64 years old/ Employers and liberal professional/ Cross Section/ 9 largest Metropolitan Areas of Brazil.	"Formal sector workers earn monthly incomes that are substantially lower than those of the protected self-employed. With rare exceptions, they also earn incomes that are somewhat higher than those of unprotected workers, the extent of which varies by occupation."
Funkenhouser (1996)	Sector choice and earnings equations (focus on human capital returns in the informal sector)	- Informal: Self-employed + Domestic + Family + Small Firm employee (<5 employees) (All with non-professional occupation). - Formal: Others.	Workers w/ positive income/18-65 years old/N.S. Cross-section /Guatemala, El Salvador, Honduras, Nicaragua and Costa Rica.	"...the finding of high returns to human capital provides evidence in favor of the existence of a dynamic informal sector."

Cont. Table 1

Funkenhouser (1997a)	Mobility between sectors with predicted wages from transition	- Informal: Self-employed + Domestic + Family + Small Firm employee (<6 employees) (All with non-professional occupation). - Formal: Others.	All workers/ 10-65 years old/ Employers/ Cross-section/El Salvador (urban centers).	"Segmentation within the labor market may be the result of restricted access to pre-labor-market characteristics."
Funkenhouser (1997b)	Earnings structures for mobility constrain hypotheses	- Informal: Self-employed + Domestic + Family + Small Firm employee (<5 employees) (All with non-professional occupation). - Formal: Others.	All workers/10 years or older/N.S./ Cross-section / Guatemala (urban centers).	"...show the Guatemalan labor market to have different reward structures across sectors.... Demand side rigidities...do not explain all of the earnings patterns across sectors".
Marcoullier et al (1997)	Earning equations with self-selection corrections. Decomposition and comparisons of predicted earnings.	- Informal*: (1) Self-employed + small size firm employee or employer (< 5 workers) or (2) workers not-paying for social security protection (independent of occupation). - Formal*: (1) Professionals + non-small size firm employee or employer (5 or more workers) or (2) workers w/ social security contributions.	Workers w/ positive income/N.S./ Domestic and family worker/Cross-section/ Urban areas of El Salvador, Mexico and Peru.	"Substantial returns to schooling in the informal sector do not fit the image of a secondary sector as sketched by Dickens and Lang. (...) Our research (...) casts doubt on the received wisdom that the informal sector, always and everywhere, is a poorly paid but easily entered refuge for those who have no other option."
Meng (2001)	Earning equations comparisons.	- Informal wage-earners: worked on construction sites + worked for private firms + worked for collective-owned service sector. - Formal wage-earner: worked for public + joint venture + foreign-owned + collectively-owned industrial firms. - Self-employed.	Migrant workers /N.S./N.S./Cross-section/Jinan City (China).	"...among all migrants those who worked in the self-employed group are the ones who felt the most satisfied with their current situation while the wage earners in the informal sector and migrants who worked in the formal sector felt less satisfied than the self-employed."

Table 2 - Identification of Informal Workers According to Different Empirical Strategies

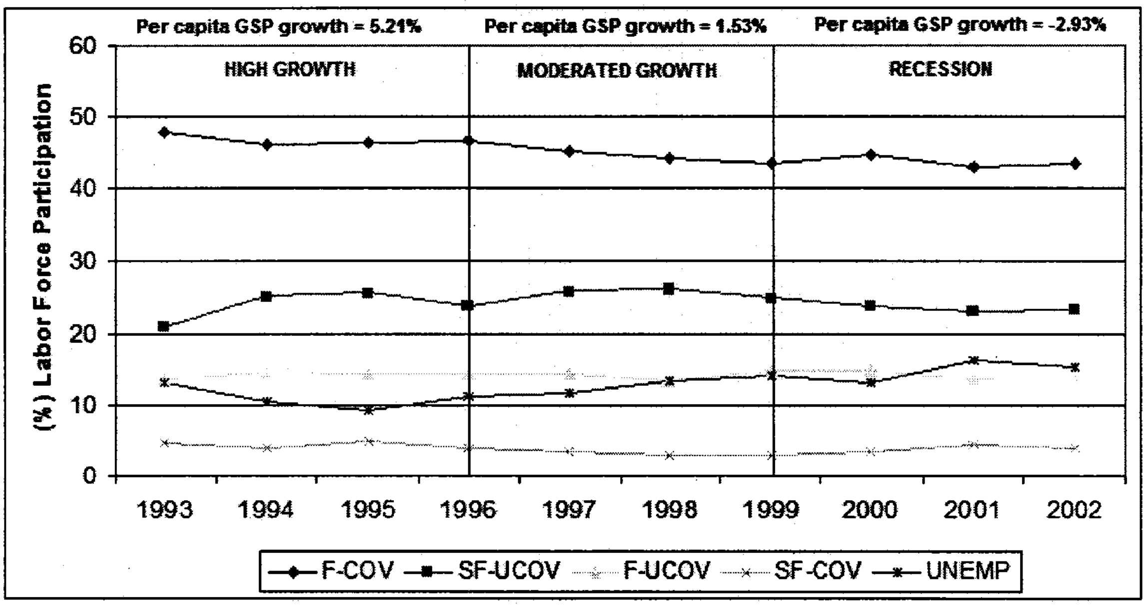
Legal Protection (Social Security)	Position in the Occupation							
	Employee		Liberal Professional	Self Employed	Employer		Family Worker	Domestic Worker
	Big Firm	Small Firm			Big Firm	Small Firm		
Covered		Ie		Ie		Ie	Ie	Ie
Uncovered	II	II, Ie	II	II, Ie	II	II, Ie	II, Ie	II, Ie

II – Informal worker according to the legalist definition.
Ie – Informal worker according to the entrepreneurial definition.

Table 3 – Worker Definition and Examples

Position on Occupation	Definition	Examples
Self-Employed	Those who work for their own account offering services to one or more firms or consumers without an employment contract. The eventuality of the services is what differentiates them with respect to employees.	Taxi drivers, hairdressers, street vendors, computer programmers, dressmakers, repairers, etc.
Liberal Professional	Those self-employed workers with college degree, registered as a professional in their respective council.	Lawyers, physicians, accountant, etc.
Employees/employers	Salaried or wage workers who have a formal or informal contract to work regularly for a particular firm.	Secretary, janitor, professors, etc.
Domestic	Employees who work for families rather than firms.	Maids, car drivers, gardeners, etc.
Small-Firm Worker	Self-Employed + employees/employers of firms with one or two workers + Domestic	-
Firm Worker	Employees/employers of firms with three or more workers + liberal professional	-
Small-Firm Uncovered Worker (SF-UCOV)	Small-Firm Worker not contributing to the Social Security System	-
Firm Uncovered Worker (F-UCOV)	Firm Worker not contributing to the Social Security System	-
Small-Firm Covered Worker (SF-COV)	Small-Firm Worker contributing to the Social Security System	-
Firm Covered Worker (F-COV)	Firm Worker contributing to the Social Security System	-

Graph 1 - Labor Force Participation by Type of Worker – Males



Graph 2 - Labor Force Participation by Type of Worker – Females

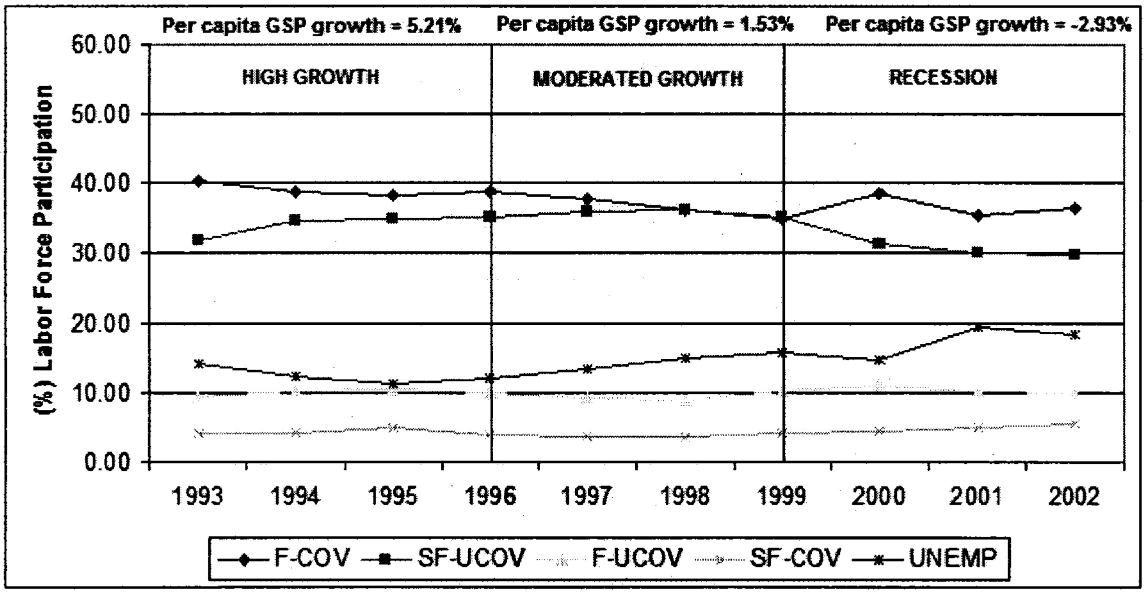


Table 4 - Explanatory Variables

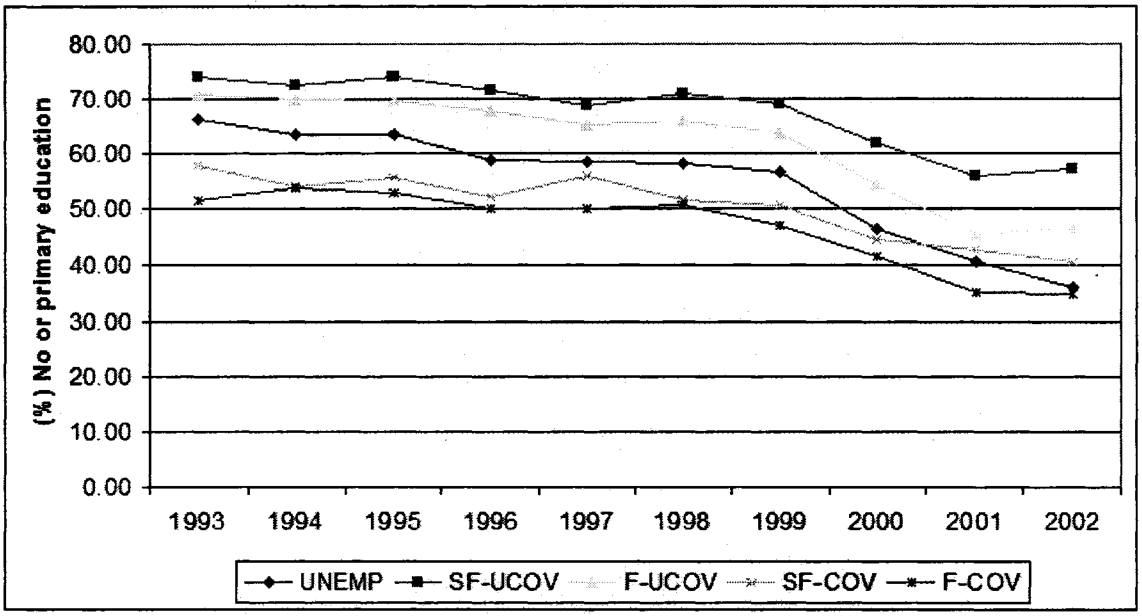
Variable	Definition
No Educ.	Dummy variable equal to 1 if the individual has no education or incomplete primary level, 0 otherwise
Primary Educ.	Dummy variable equal to 1 if the individual has complete primary level or incomplete secondary level, 0 otherwise
Second. Educ.	Dummy variable equal to 1 if the individual has complete secondary level or incomplete college level, 0 otherwise
College Educ.	Dummy variable equal to 1 if the individual has complete college level or more, 0 otherwise
Household head	Dummy variable equal to 1 if the individual is the household head, 0 otherwise
Migrantes Up to 2 years	Dummy Variable equal to 1 if the individual migrated from other municipalities within the state and has up to two years of residency, 0 otherwise.
Migrantes 2 to 5 years	Dummy Variable equal to 1 if the individual migrated from other municipalities within the state and has between 2 and 5 years of residency, 0 otherwise
Migrantes > 5 years	Dummy Variable equal to 1 if the individual migrated from other municipalities within the state and has more than 5 years of residency, 0 otherwise
Migrants outside state	Dummy Variable equal to 1 if the individual migrated from outside of the state.
Age	Individuals' age
# Children under 5	Number of children less than 5 years old.
Per capita family income (others)	Real Per capita family income excluding the reported income of the individual of reference (R$ 2000).
%F-COV (neigh)	Percentage of individuals in the neighborhood who were firm covered workers in the last two years.
%SF-UCOV (neigh)	Percentage of individuals in the neighborhood who were small-firm uncovered workers in the last two years.
%F-COV (neigh)	Percentage of individuals in the neighborhood who were firm uncovered workers in the last two years.
%SF-COV (neigh)	Percentage of individuals in the neighborhood who were small-firm covered workers in the last two years.
INCNEIGH	Median of the per capita family income in the neighborhood in the last year
Region 1	Dummy variable equal to 1 if the individual lives in municipal administrative area 1, 0 otherwise
Region 3	Dummy variable equal to 1 if the individual lives in municipal administrative area 3, 0 otherwise
Region 4	Dummy variable equal to 1 if the individual lives in municipal administrative area 4, 0 otherwise
Region 5	Dummy variable equal to 1 if the individual lives in municipal administrative area 5, 0 otherwise
Region 6	Dummy variable equal to 1 if the individual lives in municipal administrative area 6, 0 otherwise

Table 5 – Spatial Segmentation versus Spatial Integration

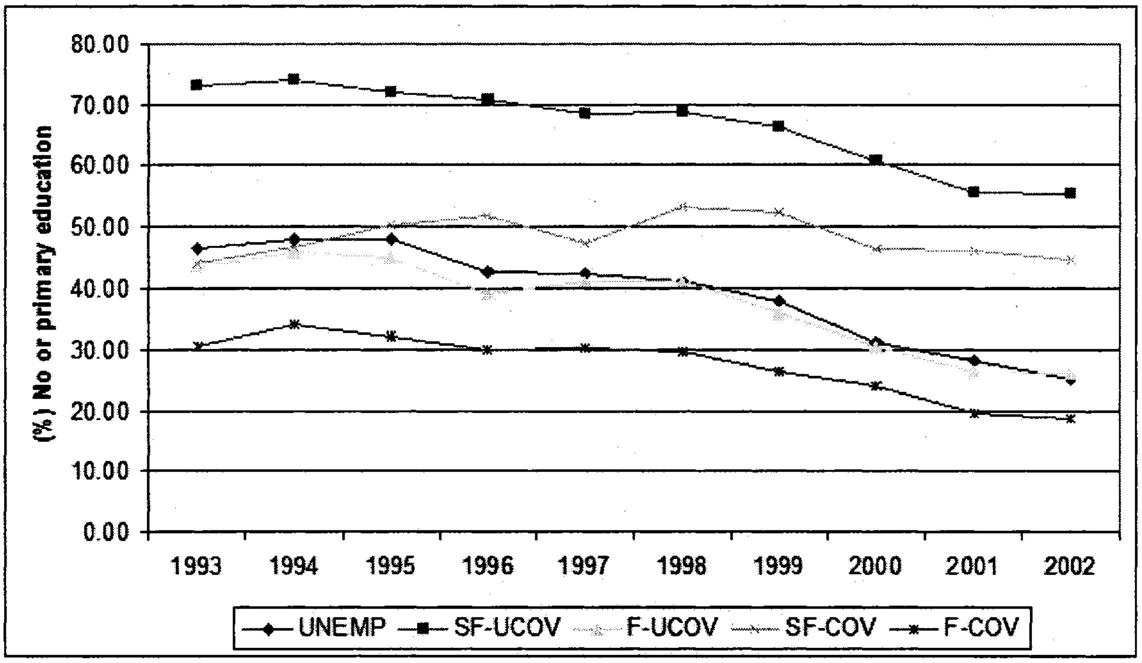
	SF - UCOV	F - UCOV	SF – COV	F – COV
% SF – UCOV	$\mu^{1}_{sf\text{-}u}$			
% F – UCOV		$\mu^{2}_{f\text{-}u}$	μ^{s}_{j}	
% SF – COV		μ^{j}_{s}	$\mu^{3}_{sf\text{-}c}$	
% F – COV				$\mu^{4}_{f\text{-}c}$

Spatial segmentation: μ^{j}_{j} (+), μ^{s}_{j} (-) Spatial integration: μ^{j}_{j} (+), μ^{s}_{j} (+)

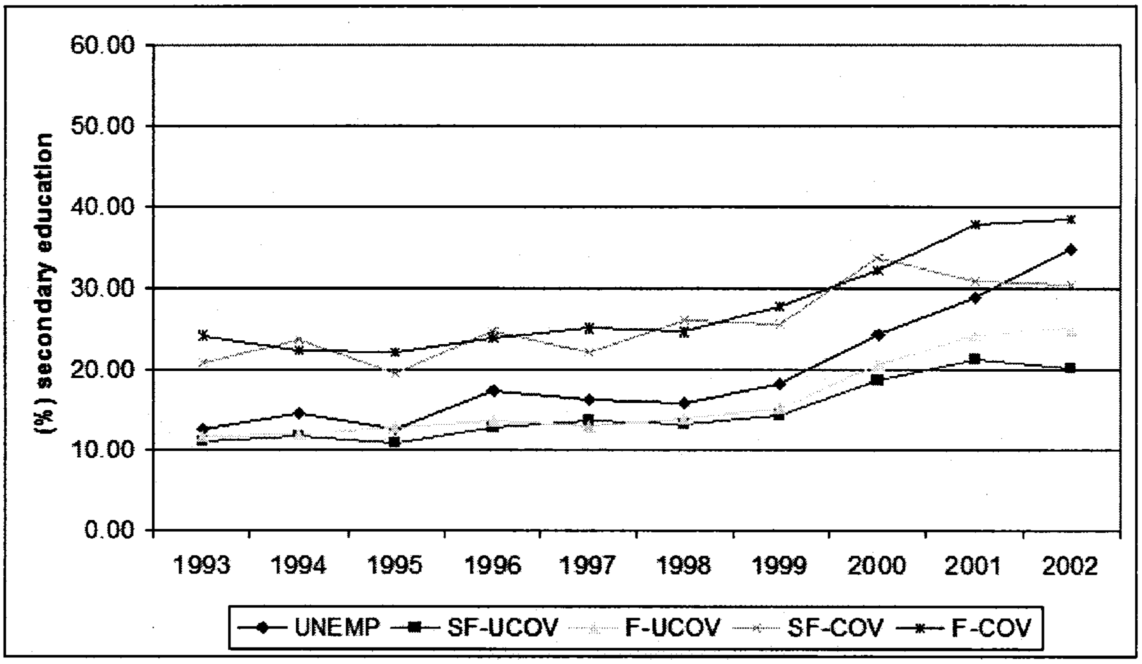
Graph 3 - Percentage of Workers With None or Incomplete Primary Education – Males



Graph 4 - Percentage of Workers With None or Incomplete Primary Education – Females



Graph 5 - Percentage of Workers with Secondary Education – Males



Graph 6 - Percentage of Workers with Secondary Education – Females

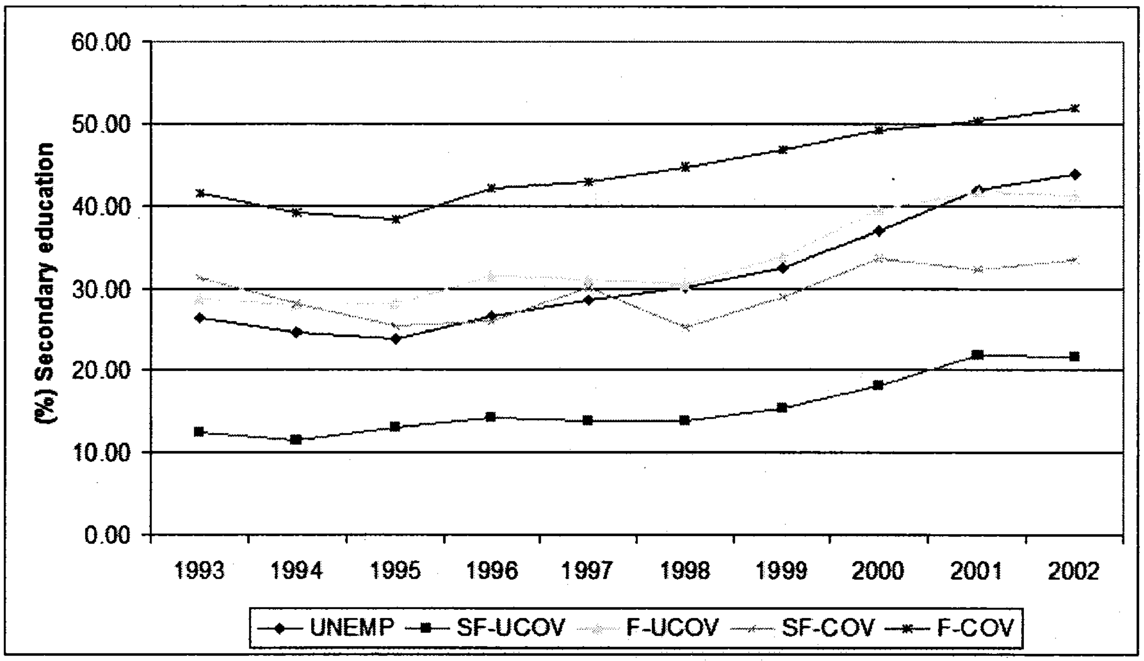


Table 6 - Mean Values by Labor Force State - Males

	Unemployed	Small Firm - Uncovered	Firm - Uncovered	Small Firm - Covered	Firm - Covered
Individual Characteristics					
Age	28.64	37.18**/**	30.16**/**	40.15**/**	33.80
No Educ.	0.53	0.67**/**	0.62**/**	0.50**/**	0.47
Primary Educ.	0.25	0.17**/**	0.20**/**	0.20**/**	0.21
Second. Educ.	0.21	0.15**/**	0.16**/**	0.26**/**	0.28
College Educ.	0.01	0.01 */**	0.02**/**	0.04**/+	0.04
Household head	0.31	0.67**/**	0.45**/**	0.73**/**	0.61
Migrantes Up to 2 years	0.04	0.03**/**	0.05**/**	0.02**/**	0.03
Migrantes 2 to 5 years	0.01	0.01**/**	0.01**/**	0.01 /**	0.01
Migrantes > 5 years	0.17	0.29**/**	0.20**/**	0.30**/**	0.24
Migrants outside state	0.08	0.08 /**	0.07**/	0.10**/**	0.07
# Children under 5	0.51	0.56**/**	0.59**/**	0.43**/**	0.56
Per capita familiy income (others)	158.59	105.88**/**	127.65 /	148.84**/**	121.09
Neighborhood Characteristics					
%F-COV	42.37	42.22**/+	42.41 +/**	42.96**/**	42.84
%SF-UCOV	28.21	28.60**/**	28.33 */	27.85**/**	28.03
%F-UCOV	12.36	12.38 /**	12.39 /**	12.24**/*	12.30
%SF-COV	3.87	3.90**/**	3.92**/	4.16**/**	3.93
INCNEIGH	93.52	92.21 /	92.37 /	96.52**/**	92.87
Region 1	0.19	0.16**/**	0.17**/**	0.16**/**	0.17
Region 3	0.16	0.15*/**	0.15+/**	0.14**/**	0.16
Region 4	0.16	0.17**/**	0.18**/**	0.20**/**	0.18
Region 5	0.18	0.19**/**	0.19**/**	0.16**/**	0.18
Region 6	0.14	0.14**/**	0.14**/**	0.12**/**	0.13
Observations	20,854 (12.81%)	39,443 (24.23%)	23,046 (14.16%)	6,147 (3.78%)	73,270 (45.32%)

/ (p-value < 0.01) comparing to unemployed / (p-value < 0.01) comparing to formal covered

*/** (p-value < 0.05) comparing to unemployed / (p-value < 0.01) comparing to formal covered

**/* (p-value < 0.05) comparing to unemployed / (p-value < 0.05) comparing to formal covered

**/+ (p-value < 0.01) comparing to unemployed / (p-value < 0.10) comparing to formal covered

+/** (p-value < 0.10) comparing to unemployed / (p-value < 0.01) comparing to formal covered

/** (p-value > 0.10) comparing to unemployed / (p-value < 0.01) comparing to formal covered

**/ (p-value < 0.01) comparing to unemployed / (p-value > 0.10) comparing to formal covered

*/ (p-value < 0.05) comparing to unemployed / (p-value > 0.10) comparing to formal covered

/ (p-value > 0.10) comparing to unemployed / (p-value > 0.10) comparing to formal covered

Table 7 - Mean Values by Labor Force State - Females

	Unemployed	Small Firm - Uncovered	Firm - Uncovered	Small Firm - Covered	Firm - Covered
Individual Characteristics					
Age	26.62	37.08**/**	29.62**/*	38.07**/**	33.39
No Educ.	0.37	0.66**/**	0.37+/**	0.48**/**	0.27
Primary Educ.	0.28	0.17**/**	0.24**/**	0.19**/**	0.18
Second. Educ.	0.33	0.16**/**	0.34**/**	0.30**/**	0.45
College Educ.	0.02	0.01**/**	0.05**/**	0.03**/**	0.10
Household head	0.09	0.25**/**	0.13**/**	0.25**/**	0.16
Migrantes Up to 2 years	0.04	0.03**/**	0.03**/**	0.02**/	0.02
Migrantes 2 to 5 years	0.01	0.01 /**	0.01 /*	0.01 /*	0.01
Migrantes > 5 years	0.16	0.34**/**	0.19**/**	0.33**/**	0.23
Migrants outside state	0.08	0.08 /**	0.07**/**	0.09**/**	0.06
# Children under 5	0.55	0.53 /**	0.45**/**	0.38**/	0.39
Per capita familiy income (others)	183.08	147.38**/**	204.60**/*	212.63**/	220.47
Neighborhood Characteristics					
%F-COV	42.57	42.24**/	42.76**/	42.59**/**	43.53
%SF-UCOV	27.88	28.62**/**	28.03*/	28.05**/**	27.45
%F-UCOV	12.32	12.33 /**	12.29 /**	12.18**/	12.14
%SF-COV	3.88	3.98**/**	3.99**/**	4.27**/*	4.06
INCNEIGH	96.01	93.83 /**	95.82*/**	99.15**/	99.66
Region 1	0.21	0.15**/**	0.16**/**	0.15**/**	0.18
Region 3	0.16	0.16 /**	0.16 /**	0.14**/**	0.17
Region 4	0.17	0.18**/**	0.20**/**	0.18**/**	0.22
Region 5	0.15	0.18**/**	0.17**/**	0.14**/	0.14
Region 6	0.13	0.14**/**	0.12**/**	0.12**/**	0.11
Observations	17429 (14.83%)	39228 (33.37%)	11740 (9.99%)	5133 (4.37%)	44024 (37.45%)

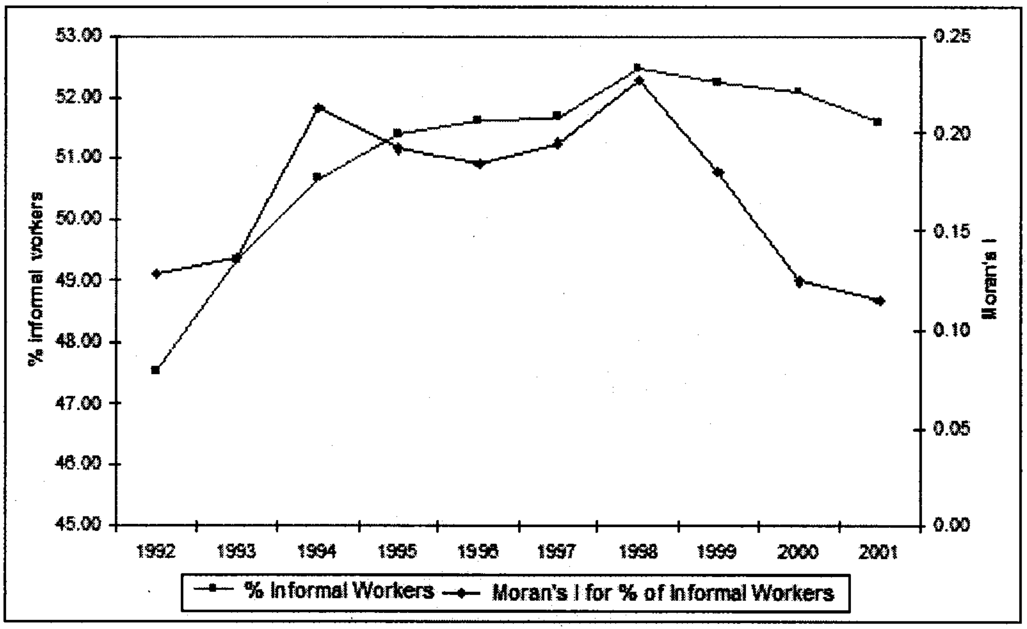
/ (p-value < 0.01) comparing to unemployed / (p-value < 0.01) comparing to formal covered
*/** (p-value < 0.05) comparing to unemployed / (p-value < 0.01) comparing to formal covered
**/* (p-value < 0.05) comparing to unemployed / (p-value < 0.05) comparing to formal covered
**/+ (p-value < 0.01) comparing to unemployed / (p-value < 0.10) comparing to formal covered
+/** (p-value < 0.10) comparing to unemployed / (p-value < 0.01) comparing to formal covered
/** (p-value > 0.10) comparing to unemployed / (p-value < 0.01) comparing to formal covered
**/ (p-value < 0.01) comparing to unemployed / (p-value > 0.10) comparing to formal covered
*/ (p-value < 0.05) comparing to unemployed / (p-value > 0.10) comparing to formal covered
/ (p-value > 0.10) comparing to unemployed / (p-value > 0.10) comparing to formal covered

Table 8 - Spatial Correlation for Workers' Share in Neighborhoods

	Unemployed	Small Firm - Uncovered	Firm - Uncovered	Small Firm - Covered	Firm - Covered
Moran's I	0.305** (7.75)	0.195** (5.85)	0.180** (4.62)	0.531** (13.11)	0.231** (5.85)

Z statistics in parentheses.
+ significant at 10%; * significant at 5%; ** significant at 1%

Graph 7- Percentage of Informal Workers versus Moran's I for the Share of Informal Workers in the Neighborhoods



Graph 8 - Percentage of Informal Uncovered Workers versus Moran's I for the Share of SF-UCOV Workers in the Neighborhoods

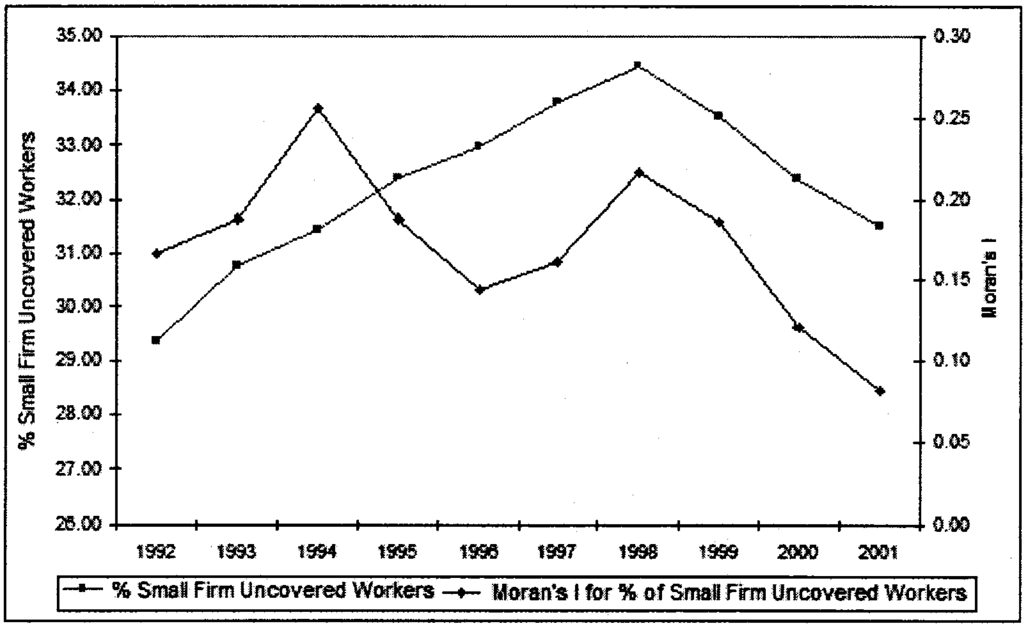


Table 9 - Multinomial Logit Coefficients and Marginal Effects – Workers' Types - Males

Variables	Small Firm - Uncovered		Firm - Uncovered		Small Firm - Covered		Firm - Covered	
	Without Neigh. Effect	With Neigh. Effect	Without Neigh. Effect	With Neigh. Effect	Without Neigh. Effect	With Neigh. Effect	Without Neigh. Effect	With Neigh. Effect
age	0.1049**	0.1049**	-0.0628**	-0.0634**	0.1201**	0.1195**	0.1162**	0.1158**
	(15.43)	(15.53)	(9.37)	(9.46)	(13.49)	(13.45)	(23.49)	(23.14)
	[0.0071]	[0.0072]	[-0.0194]	[-0.0194]	[0.0015]	[0.0015]	[0.0193]	[0.0193]
age2	-0.0008**	-0.0008**	0.0008**	0.0008**	-0.0008**	-0.0008**	-0.0013**	-0.0013**
	(9.40)	(9.47)	(10.09)	(10.08)	(7.00)	(7.13)	(19.58)	(19.40)
	[0.0000]	[0.0000]	[0.0002]	[0.0002]	[0.0000]	[0.0000]	[-0.0003]	[-0.0003]
Primary	-0.2990**	-0.3028**	-0.2782**	-0.2883**	0.3104**	0.2916**	0.2516**	0.2478**
Educ.	(10.39)	(10.82)	(9.28)	(9.81)	(5.45)	(5.13)	(10.00)	(9.91)
	[-0.0732]	[-0.0729]	[-0.0402]	[-0.0409]	[0.0099]	[0.0093]	[0.1076]	[0.1082]
Second.	-0.2601**	-0.2796**	-0.1783**	-0.2034**	0.6653**	0.6136**	0.6988**	0.6926**
Educ.	(7.25)	(8.12)	(4.99)	(5.96)	(11.82)	(11.32)	(25.22)	(25.16)
	[-0.1181]	[-0.1193]	[-0.0602]	[-0.0617]	[0.0127]	[0.0110]	[0.1981]	[0.2014]
College	-0.2667**	-0.3237**	0.9221**	0.8649**	1.1837**	1.0480**	1.5359**	1.5216**
Educ.	(2.60)	(3.10)	(10.63)	(9.88)	(10.46)	(9.16)	(18.18)	(18.31)
	[-0.1828]	[-0.1848]	[-0.0193]	[-0.0231]	[0.0040]	[0.0001]	[0.2744]	[0.2829]
# Children	0.0502**	0.0517**	0.0715**	0.0744**	-0.0917**	-0.0811**	0.0336**	0.0349**
under 5	(3.35)	(3.47)	(4.59)	(4.83)	(3.59)	(3.12)	(2.75)	(2.86)
	[0.0037]	[0.0036]	[0.0051]	[0.0053]	[-0.0042]	[-0.0039]	[-0.0006]	[-0.0008]
Household	0.8451**	0.8528**	0.6341**	0.6435**	1.0296**	1.0634**	0.9255**	0.9305**
head	(30.00)	(29.94)	(22.17)	(22.78)	(26.93)	(27.81)	(41.33)	(41.47)
	[0.0212]	[0.0215]	[-0.0170]	[-0.0166]	[0.0089]	[0.0097]	[0.0785]	[0.0779]
Migrantes	-0.0175	-0.0114	0.1434**	0.1489**	-0.1310	-0.0945	0.0415	0.0477
Up to 2	(0.31)	(0.20)	(2.78)	(2.92)	(1.05)	(0.75)	(0.86)	(0.98)
years	[-0.0118]	[-0.0119]	[0.0164]	[0.0163]	[-0.0051]	[-0.0042]	[0.0042]	[0.0041]
Migrantes	0.4564**	0.4609**	0.4943**	0.4987**	0.3838*	0.4214*	0.5180**	0.5246**
2 to 5	(4.25)	(4.27)	(4.66)	(4.72)	(2.06)	(2.26)	(5.04)	(5.11)
years	[0.0021]	[0.0017]	[0.0067]	[0.0064]	[-0.0021]	[-0.0011]	[0.0344]	[0.0346]
Migrantes	0.1741**	0.1814**	0.0677*	0.0748**	0.2478**	0.2796**	0.1519**	0.1572**
> 5 years	(6.32)	(6.83)	(2.42)	(2.75)	(5.88)	(6.60)	(5.92)	(6.14)
	[0.0102]	[0.0104]	[-0.0089]	[-0.0088]	[0.0040]	[0.0049]	[0.0093]	[0.0088]
Migrants	-0.0619	-0.0580	-0.1337**	-0.1327**	0.0123	0.0125	-0.3177**	-0.3146**
outside	(1.64)	(1.53)	(3.27)	(3.25)	(0.19)	(0.20)	(8.00)	(7.98)
state	[0.0288]	[0.0291]	[0.0060]	[0.0058]	[0.0068]	[0.0067]	[-0.0631]	[-0.0628]
Per capita	-0.0001**	-0.0001**	-0.0001**	-0.0001**	0.0000	0.0000	-0.0001**	-0.0001**
family inc.	(3.32)	(3.45)	(3.20)	(3.35)	(0.51)	(0.11)	(6.96)	(6.97)
(others)	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
D1994	0.4169**	0.4339**	0.2894**	0.2924**	0.0378	0.0581	0.1948**	0.2030**
	(6.30)	(6.17)	(3.86)	(3.76)	(0.43)	(0.63)	(3.35)	(3.40)
	[0.0456]	[0.0477]	[0.0069]	[0.0060]	[-0.0063]	[-0.0059]	[-0.0214]	[-0.0220]
D1995	0.5524**	0.5369**	0.4124**	0.4009**	0.3774**	0.3654**	0.3364**	0.3496**
	(7.80)	(7.03)	(5.04)	(4.60)	(3.47)	(3.32)	(5.79)	(5.71)
	[0.0469]	[0.0426]	[0.0058]	[0.0042]	[0.0002]	[-0.0002]	[-0.0164]	[-0.0102]
D1996	0.3096**	0.1979**	0.2291**	0.1691*	-0.0717	-0.1824+	0.1451**	0.1815**
	(5.29)	(3.07)	(3.58)	(2.30)	(0.77)	(1.78)	(2.67)	(3.05)
	[0.0337]	[0.0108]	[0.0075]	[0.0022]	[-0.0076]	[-0.0098]	[-0.0148]	[0.0133]
D1997	0.3210**	0.1871**	0.1937**	0.1221	-0.2685**	-0.4026**	0.0507	0.0975+
	(5.24)	(2.73)	(2.78)	(1.55)	(3.38)	(4.44)	(0.91)	(1.68)
	[0.0504]	[0.0220]	[0.0096]	[0.0033]	[-0.0114]	[-0.0138]	[-0.0349]	[-0.0006]

Cont. Table 9

D1998	0.1333*	0.0020	-0.0369	-0.0960	-0.6154**	-0.7318**	-0.1515**	-0.1032+
	(2.02)	(0.03)	(0.53)	(1.20)	(7.61)	(8.34)	(2.76)	(1.81)
	[0.0492]	[0.0205]	[0.0027]	[-0.0022]	[-0.0152]	[-0.0170]	[-0.0433]	[-0.0107]
D1999	0.0459	-0.0776	0.0397	-0.0000	-0.6925**	-0.7789**	-0.2245**	-0.1808**
	(0.73)	(1.11)	(0.68)	(0.00)	(8.49)	(8.82)	(4.51)	(3.46)
	[0.0378]	[0.0111]	[0.0210]	[0.0181]	[-0.0159]	[-0.0171]	[-0.0550]	[-0.0268]
D2000	0.1481*	0.0361	0.1552*	0.1206	-0.3993**	-0.4611**	-0.1274*	-0.0740
	(2.21)	(0.52)	(2.08)	(1.43)	(4.31)	(4.41)	(2.25)	(1.28)
	[0.0399]	[0.0139]	[0.0244]	[0.0210]	[-0.0113]	[-0.0124]	[-0.0539]	[-0.0248]
D2001	-0.0498	-0.1513*	-0.1202+	-0.1661*	-0.4185**	-0.4851**	-0.3813**	-0.3125**
	(0.83)	(2.27)	(1.85)	(2.02)	(5.66)	(5.58)	(7.55)	(5.71)
	[0.0416]	[0.0165]	[0.0133]	[0.0074]	[-0.0064]	[-0.0080]	[-0.0747]	[-0.0430]
D2002	0.0372	-0.0520	-0.0144	-0.0612	-0.4440**	-0.5049**	-0.3051**	-0.2355**
	(0.62)	(0.81)	(0.20)	(0.72)	(5.53)	(5.25)	(5.66)	(3.90)
	[0.0460]	[0.0229]	[0.0186]	[0.0120]	[-0.0090]	[-0.0104]	[-0.0726]	[-0.0418]
Region 1	-	-0.0185	-	-0.0422	-	-0.2550**	-	-0.0688+
		(0.26)		(0.68)		(2.95)		(1.71)
		[0.0078]		[0.0012]		[-0.0063]		[-0.0085]
Region 3	-	0.0047	-	0.0063	-	-0.3658**	-	-0.0592
		(0.07)		(0.10)		(4.60)		(1.58)
		[0.0101]		[0.0061]		[-0.0098]		[-0.0106]
Region 4	-	0.1015	-	0.1201*	-	-0.1044	-	0.0558
		(1.60)		(2.04)		(1.45)		(1.29)
		[0.0089]		[0.0079]		[-0.0053]		[-0.0043]
Region 5	-	-0.0088	-	0.0176	-	-0.3204**	-	-0.0385
		(0.13)		(0.27)		(4.85)		(0.97)
		[0.0045]		[0.0064]		[-0.0089]		[-0.0052]
Region 6	-	-0.0474	-	-0.0761	-	-0.3784**	-	-0.0987**
		(0.72)		(1.02)		(3.20)		(2.64)
		[0.0079]		[0.0005]		[-0.0089]		[-0.0089]
INCNEIGH	-	0.0021**	-	0.0011*	-	0.0022**	-	-0.0009+
		(4.18)		(2.20)		(3.06)		(1.93)
		[0.0004]		[0.0001]		[0.0001]		[-0.0006]
Wald Test for Combining Labor Force States (Chi-Square value)								
H0:$\beta_{Unemp} = \beta_{column}$	9362**	12154**	1095**	1331**	3617**	4387**	6308**	7647**
H0:$\beta_{sf\text{-}ucov} = \beta_{column}$		-	5863**	6715**	1101**	1430**	8466**	11215**
H0:$\beta_{F\text{-}ucov} = \beta_{column}$		-	-	-	1938**	2330**	2568**	3088**
H0:$\beta_{sf\text{-}cov} = \beta_{column}$		-	-	-	-	-	1236**	1571**

Robust t statistics in parentheses + significant at 10%; * significant at 5%; ** significant at 1%.
Marginal evaluated at population means are given in square brackets.
Unemployed is the base category.

Table 10 - Multinomial Logit Coefficients and Marginal Effects – Workers' Types - Females

Variables	Small Firm - Uncovered		Firm - Uncovered		Small Firm - Covered		Firm - Covered	
	Without Neigh. Effect	With Neigh. Effect	Without Neigh. Effect	With Neigh. Effect	Without Neigh. Effect	With Neigh. Effect	Without Neigh. Effect	With Neigh. Effect
age	0.1615**	0.1631**	-0.0123	-0.0114	0.1610**	0.1652**	0.1598**	0.1600**
	(22.82)	(23.13)	(1.56)	(1.46)	(16.74)	(16.79)	(20.73)	(20.78)
	[0.0129]	[0.0131]	[-0.0144]	[-0.0144]	[0.0018]	[0.0019]	[0.0154]	[0.0151]
age2	-0.0010**	-0.0010**	0.0006**	0.0006**	-0.0009**	-0.0009**	-0.0013**	-0.0013**
	(9.86)	(10.14)	(5.45)	(5.46)	(6.49)	(6.89)	(10.96)	(11.03)
	[-0.0001]	[-0.0001]	[0.0002]	[0.0002]	[0.0000]	[0.0000]	[-0.0002]	[-0.0002]
Primary	-0.5600**	-0.5567**	0.0179	0.0195	-0.0999+	-0.1053*	0.2304**	0.2242**
Educ.	(19.28)	(20.00)	(0.50)	(0.54)	(1.79)	(2.03)	(7.20)	(7.09)
	[-0.1383]	[-0.1369]	[0.0083]	[0.0087]	[-0.0019]	[-0.0020]	[0.1243]	[0.1224]
Second.	-0.9730**	-0.9797**	0.1497**	0.1460**	-0.0234	-0.0485	0.8253**	0.8119**
Educ.	(25.35)	(27.10)	(3.59)	(3.37)	(0.37)	(0.90)	(22.24)	(22.48)
	[-0.2848]	[-0.2842]	[0.0024]	[0.0030]	[-0.0062]	[-0.0068]	[0.3037]	[0.3021]
College	-1.5989**	-1.6456**	0.8245**	0.8098**	-0.3549**	-0.4648**	1.5798**	1.5527**
Educ.	(19.16)	(20.10)	(9.86)	(9.38)	(3.17)	(4.36)	(22.77)	(21.98)
	[-0.3172]	[-0.3179]	[0.0005]	[0.0016]	[-0.0322]	[-0.0329]	[0.4220]	[0.4208]
# Children	0.0348**	0.0355**	-0.1158**	-0.1155**	-0.1383**	-0.1330**	-0.1266**	-0.1246**
under 5	(2.77)	(2.89)	(5.90)	(5.91)	(5.30)	(5.04)	(9.62)	(9.45)
	[0.0296]	[0.0294]	[-0.0062]	[-0.0063]	[-0.0037]	[-0.0034]	[-0.0273]	[-0.0270]
Household	0.1228**	0.1241**	0.0917*	0.0904*	0.1711**	0.1768**	0.0240	0.0238
head	(3.25)	(3.29)	(2.24)	(2.22)	(3.76)	(3.85)	(0.71)	(0.70)
	[0.0180]	[0.0183]	[0.0026]	[0.0025]	[0.0049]	[0.0051]	[-0.0170]	[-0.0173]
Migrantes	0.2667**	0.2597**	-0.1332+	-0.1450*	-0.1094	-0.0781	-0.3827**	-0.3852**
Up to 2	(4.29)	(4.36)	(1.89)	(2.10)	(0.79)	(0.55)	(6.73)	(6.79)
years	[0.1180]	[0.1165]	[-0.0086]	[-0.0094]	[-0.0026]	[-0.0011]	[-0.1134]	[-0.1132]
Migrantes	0.3685**	0.3570**	-0.0133	-0.0252	0.4974*	0.5243*	0.0650	0.0684
2 to 5	(3.32)	(3.27)	(0.11)	(0.21)	(2.39)	(2.49)	(0.66)	(0.70)
years	[0.0664]	[0.0631]	[-0.0189]	[-0.0197]	[0.0170]	[0.0185]	[-0.0431]	[-0.0409]
Migrantes	0.3578**	0.3537**	-0.0044	-0.0138	0.3569**	0.3846**	0.0352	0.0357
> 5 years	(11.07)	(11.79)	(0.13)	(0.40)	(7.93)	(8.56)	(1.46)	(1.46)
	[0.0696]	[0.0685]	[-0.0160]	[-0.0169]	[0.0099]	[0.0111]	[-0.0446]	[-0.0440]
Migrants	0.0308	0.0263	-0.2559**	-0.2579**	0.0557	0.0488	-0.4932**	-0.4890**
outside	(0.75)	(0.67)	(4.74)	(4.85)	(0.79)	(0.73)	(11.94)	(11.91)
state	[0.0761]	[0.0749]	[-0.0073]	[-0.0073]	[0.0122]	[0.0117]	[-0.1072]	[-0.1057]
Per capita	-0.0001**	-0.0001**	0.0000	0.0000	0.0000*	0.0000+	-0.0000	-0.0000
family inc.	(2.71)	(2.79)	(0.07)	(0.06)	(2.27)	(1.91)	(1.04)	(1.16)
(others)	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
D1994	0.1370*	0.1460*	0.1923*	0.1932*	0.0611	0.0945	0.0827	0.0908
	(2.05)	(2.01)	(2.45)	(2.38)	(0.51)	(0.74)	(1.23)	(1.31)
	[0.0116]	[0.0120]	[0.0102]	[0.0094]	[-0.0018]	[-0.0006]	[-0.0073]	[-0.0072]
D1995	0.2897**	0.2591**	0.3403**	0.3256**	0.3832**	0.3785**	0.2030**	0.2172**
	(4.23)	(3.44)	(4.06)	(3.76)	(2.81)	(2.68)	(2.96)	(3.02)
	[0.0190]	[0.0111]	[0.0122]	[0.0113]	[0.0073]	[0.0073]	[-0.0112]	[-0.0031]
D1996	0.2150**	0.1099+	0.2063**	0.1845*	0.0411	-0.0343	0.0978+	0.1295*
	(3.61)	(1.73)	(2.95)	(2.43)	(0.41)	(0.35)	(1.71)	(2.19)
	[0.0267]	[0.0012]	[0.0080]	[0.0087]	[-0.0041]	[-0.0060]	[-0.0142]	[0.0093]
D1997	0.1036	-0.0070	0.0189	0.0138	-0.1627	-0.2291*	-0.0375	0.0036
	(1.50)	(0.10)	(0.21)	(0.15)	(1.46)	(2.04)	(0.53)	(0.05)
	[0.0294]	[0.0005]	[0.0004]	[0.0025]	[-0.0075]	[-0.0091]	[-0.0204]	[0.0049]

Cont. Table 10

D1998	-0.0217	-0.1235+	-0.1431+	-0.1323	-0.3213**	-0.3479**	-0.2117**	-0.1607*
	(0.33)	(1.74)	(1.77)	(1.45)	(3.12)	(3.20)	(3.28)	(2.24)
	[0.0317]	[0.0025]	[-0.0028]	[-0.0001]	[-0.0086]	[-0.0089]	[-0.0362]	[-0.0115]
D1999	-0.0566	-0.1521*	-0.0558	-0.0410	-0.2299*	-0.2348*	-0.3154**	-0.2594**
	(0.95)	(2.26)	(0.86)	(0.57)	(2.26)	(2.21)	(6.17)	(4.48)
	[0.0320]	[0.0036]	[0.0109]	[0.0138]	[-0.0034]	[-0.0031]	[-0.0606]	[-0.0369]
D2000	-0.0750	-0.1516*	0.0702	0.1045	-0.0799	-0.0430	-0.2299**	-0.1545*
	(1.26)	(2.21)	(0.92)	(1.19)	(0.75)	(0.39)	(4.09)	(2.55)
	[0.0104]	[-0.0170]	[0.0204]	[0.0236]	[0.0012]	[0.0025]	[-0.0465]	[-0.0222]
D2001	-0.3548**	-0.4138**	-0.3247**	-0.2744**	-0.2579*	-0.2094+	-0.6387**	-0.5538**
	(6.09)	(6.02)	(4.08)	(3.20)	(2.42)	(1.84)	(10.83)	(8.02)
	[0.0134]	[-0.0112]	[0.0079]	[0.0116]	[0.0066]	[0.0080]	[-0.0879]	[-0.0654]
D2002	-0.3004**	-0.3340**	-0.2945**	-0.2270*	-0.1385	-0.0772	-0.5939**	-0.5034**
	(4.83)	(4.55)	(3.65)	(2.46)	(1.28)	(0.65)	(8.86)	(6.52)
	[0.0173]	[-0.0035]	[0.0065]	[0.0107]	[0.0106]	[0.0121]	[-0.0871]	[-0.0670]
Region 1	-	-0.3777**	-	-0.1956**	-	-0.5303**	-	-0.0265
		(7.56)		(3.11)		(7.34)		(0.49)
		[-0.0631]		[-0.0032]		[-0.0145]		[0.0580]
Region 3	-	-0.0966+	-	-0.0102	-	-0.4633**	-	0.0527
		(1.76)		(0.15)		(4.79)		(0.85)
		[-0.0215]		[0.0019]		[-0.0169]		[0.0329]
Region 4	-	-0.0047	-	0.0917	-	-0.2862**	-	0.1433*
		(0.09)		(1.45)		(3.53)		(2.46)
		[-0.0189]		[0.0039]		[-0.0138]		[0.0359]
Region 5	-	-0.0137	-	0.1455*	-	-0.4097**	-	0.0202
		(0.22)		(2.40)		(4.41)		(0.34)
		[-0.0057]		[0.0160]		[-0.0162]		[0.0064]
Region 6	-	-0.0873+	-	-0.0863	-	-0.3971**	-	-0.0128
		(1.79)		(1.52)		(3.43)		(0.23)
		[-0.0093]		[-0.0031]		[-0.0134]		[0.0182]
INCNEIGH	-	0.0008+	-	-0.0009	-	-0.0000	-	-0.0007
		(1.69)		(1.56)		(0.07)		(1.38)
		[0.0003]		[-0.0001]		[0.0000]		[-0.0002]
Wald Test for Combining Labor Force States (Chi-Square value)								
H0:$\beta_{Unemp} = \beta_{column}$	9362**	12154**	1095**	1331**	3617**	4387**	6308**	7647**
H0:$\beta_{sf\text{-}ucov} = \beta_{column}$	-	-	5863**	6715**	1101**	1430**	8466**	11215**
H0:$\beta_{F\text{-}ucov} = \beta_{column}$	-	-	-	-	1938**	2330**	2568**	3088**
H0:$\beta_{sf\text{-}cov} = \beta_{column}$	-	-	-	-	-	-	1236**	1571**

Robust t statistics in parentheses + significant at 10%; * significant at 5%; ** significant at 1%.
Marginal evaluated at population means are given in square brackets.
Unemployed is the base category.

Figure 1 – Logit Coefficients and Odds Ratio for Educational Variables – Model Without Neighborhood Effects – Males

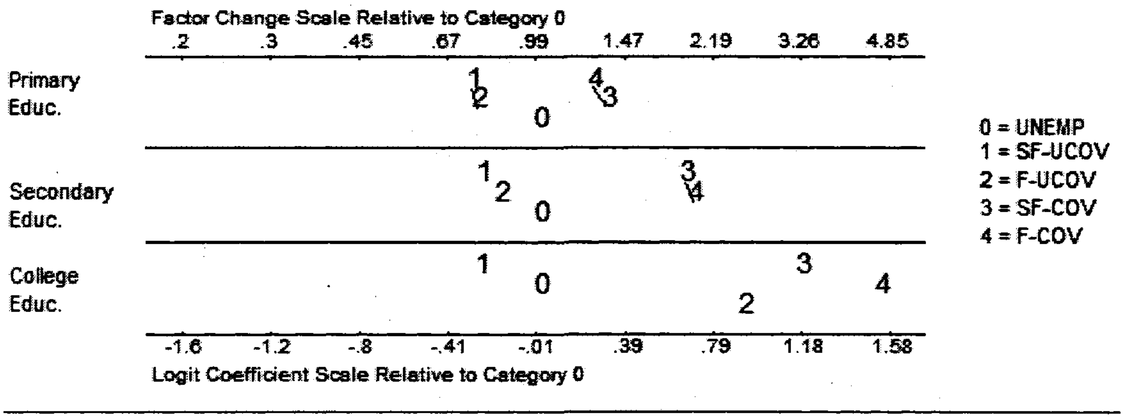


Figure 2 – Logit Coefficients and Odds Ratio for Educational Variables – Model With Neighborhood Effects – Males

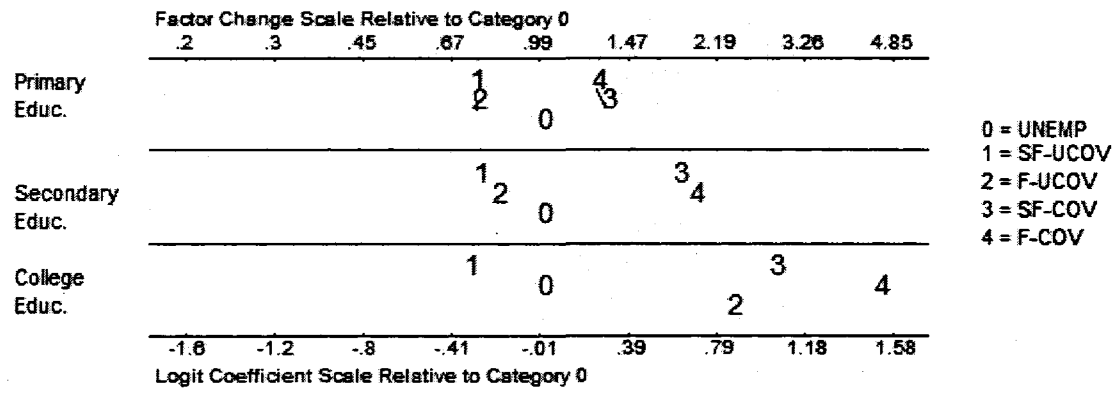


Figure 3 – Logit Coefficients and Odds Ratio for Educational Variables – Model Without Neighborhood Effects – Females

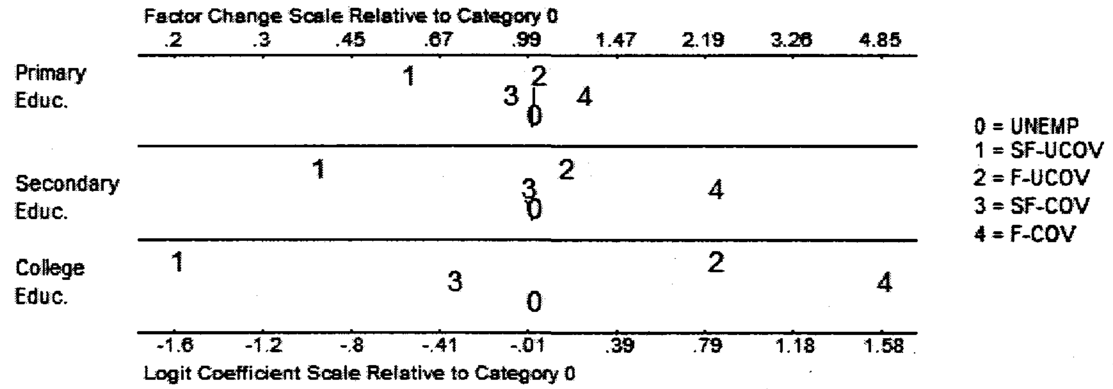


Figure 4 – Logit Coefficients and Odds Ratio for Educational Variables – Model With Neighborhood Effects – Females

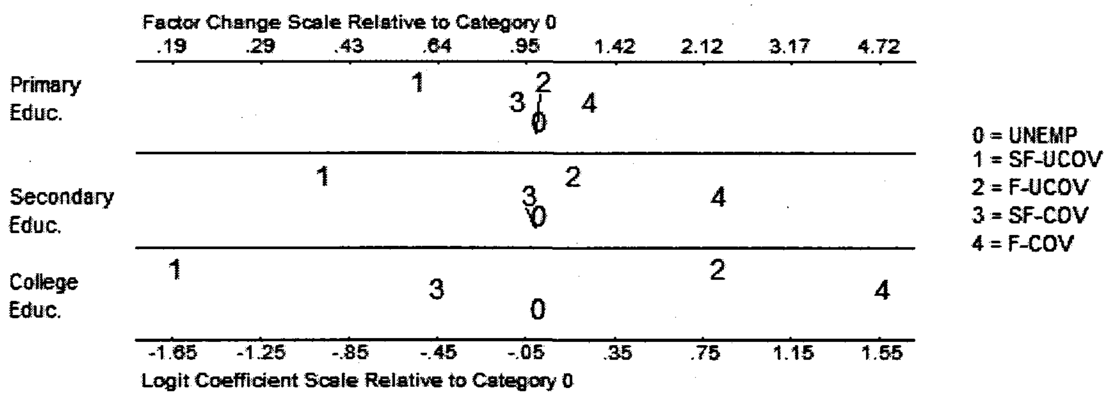


Figure 5 - Logit Coefficients and Odds Ratio for Family Characteristics – Model Without Neighborhood Effects – Males

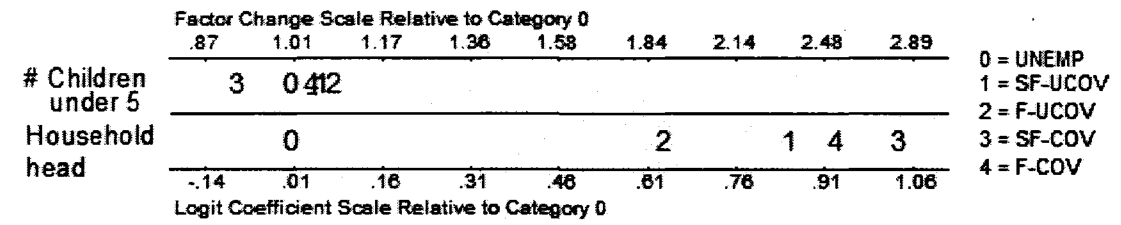


Figure 6 - Logit Coefficients and Odds Ratio for Family Characteristics – Model With Neighborhood Effects – Males

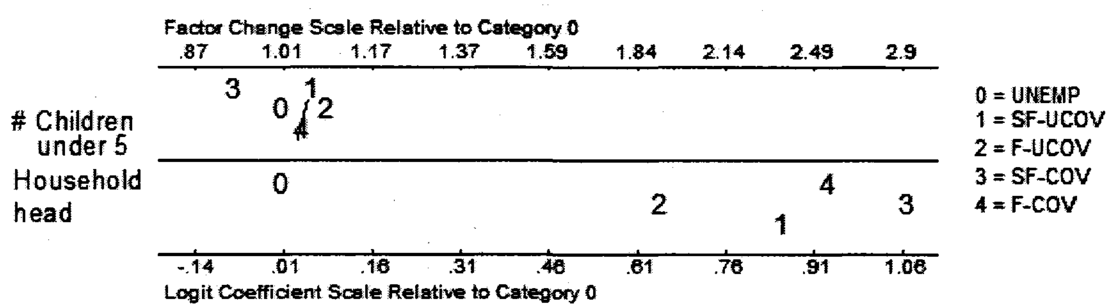


Figure 7 - Logit Coefficients and Odds Ratio for Family Characteristics – Model Without Neighborhood – Females

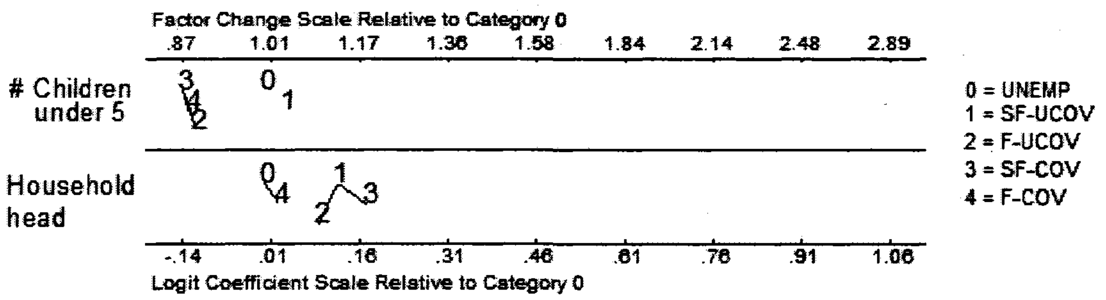


Figure 8 - Logit Coefficients and Odds Ratio for Family Characteristics – Model With Neighborhood Effects – Females

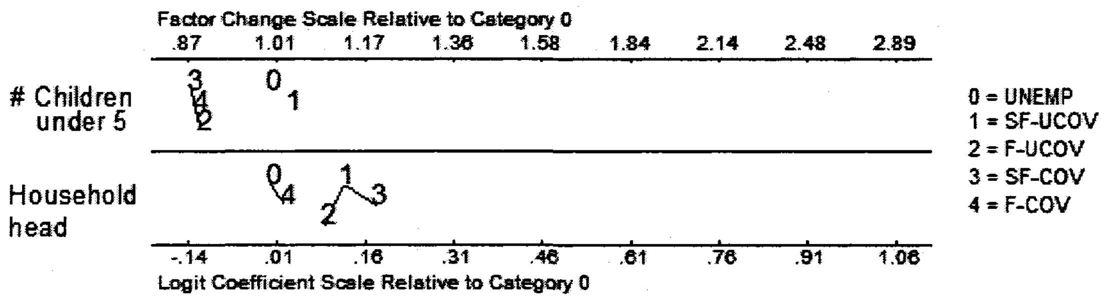


Figure 9 - Logit Coefficients and Odds Ratio for Migration Variables – Model Without Neighborhood Effects – Males

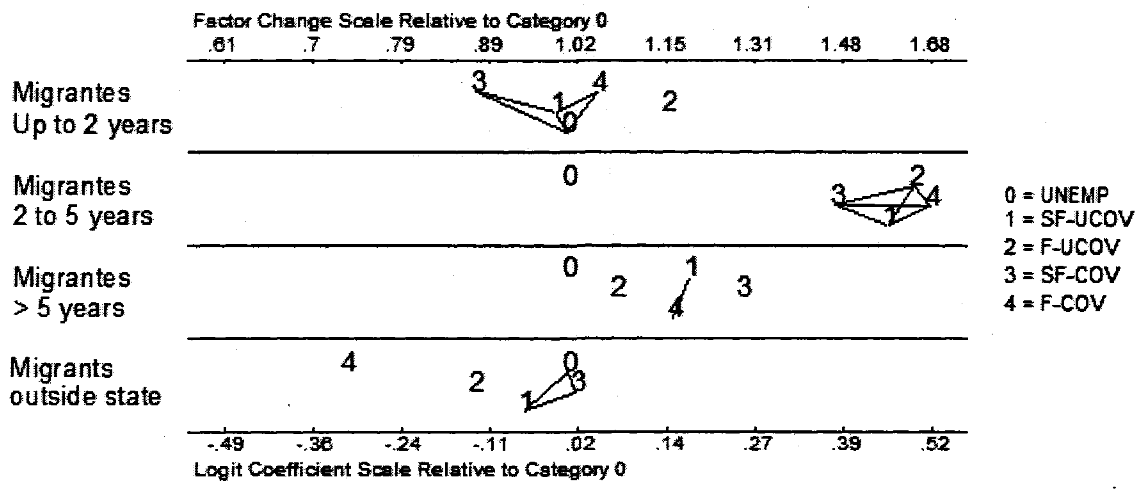


Figure 10 - Logit Coefficients and Odds Ratio for Migration Variables – Model With Neighborhood Effects – Males

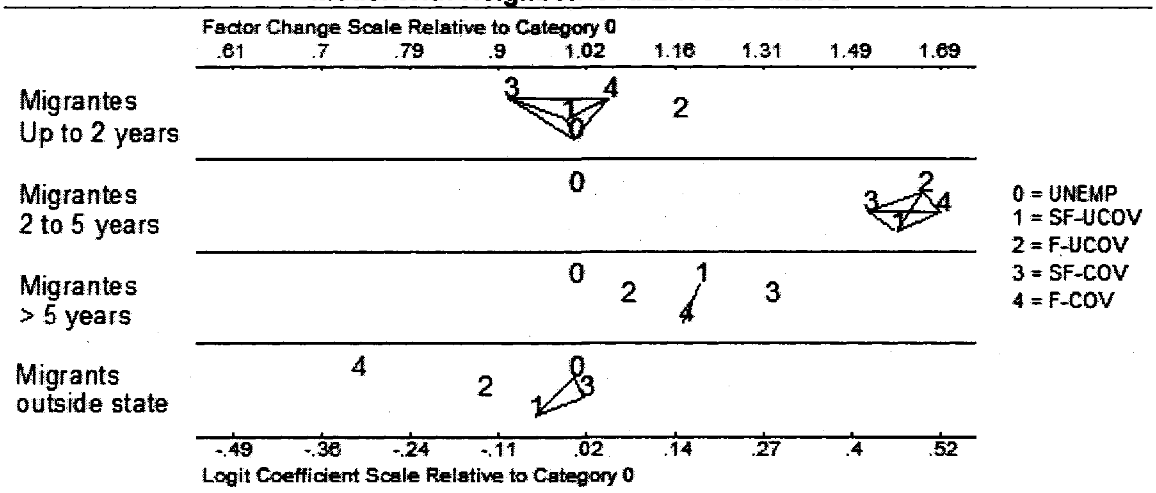


Figure 11 - Logit Coefficients and Odds Ratio for Migration Variables – Model Without Neighborhood Effects – Females

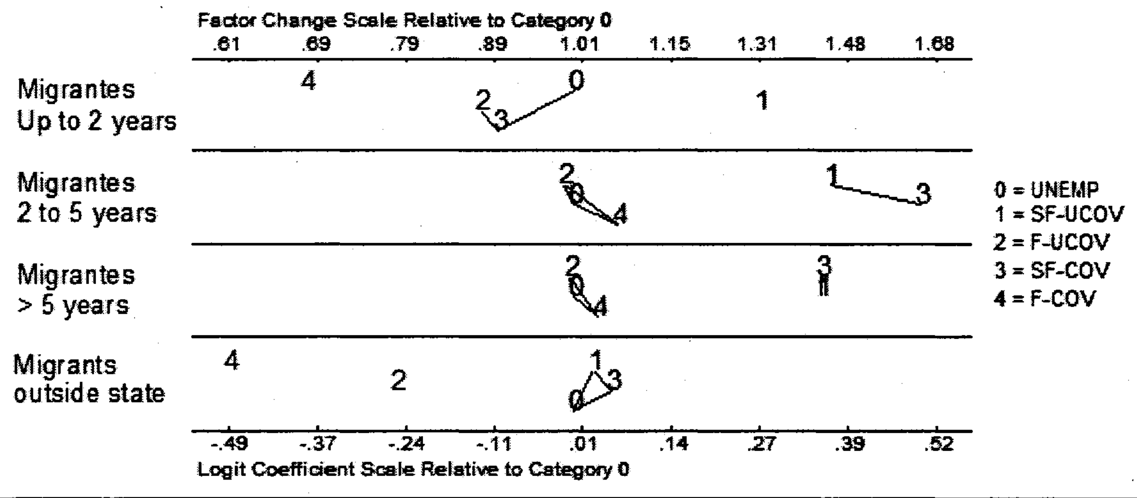


Figure 12 - Logit Coefficients and Odds Ratio for Migration Variables – Model With Neighborhood Effects – Females

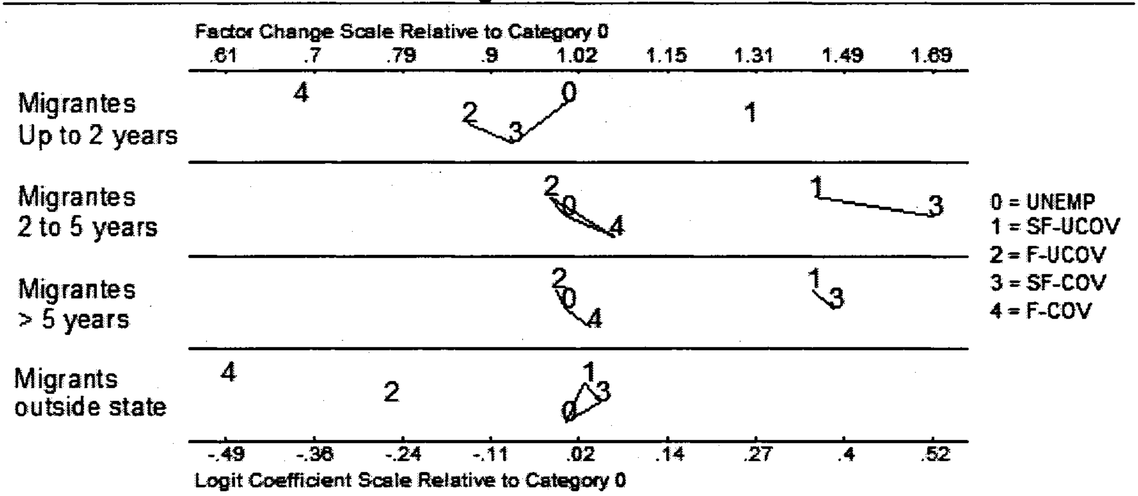


Table 11 - Multinomial Logit Coefficients and Marginal Effects – Social Interaction Variables - Males

Variables	Small Firm Uncovered		Firm Uncovered		Small Firm Covered		Firm Covered	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
%SF-UCOV	**0.0257****	**0.0201****	0.0129*	0.0070	0.0235**	0.0197*	0.0053	0.0038
(neigh)	**(4.32)**	**(3.24)**	(2.31)	(1.10)	(2.84)	(2.29)	(1.07)	(0.77)
	[0.0035]	**[0.0028]**	[0.0002]	[-0.0002]	[0.0004]	[0.0004]	[-0.0028]	[-0.0021]
%F-UCOV	0.0040	-0.0061	**0.0164***	**0.0058**	0.0113	0.0017	-0.0020	-0.0048
(neigh)	(0.51)	(0.80)	**(1.97)**	**(0.72)**	(0.94)	(0.15)	(0.33)	(0.83)
	[0.0003]	[-0.0008]	**[0.0019]**	**[0.0012]**	[0.0003]	[0.0002]	[-0.0022]	[-0.0009]
%SF-COV	0.0358**	0.0135	0.0346**	0.0179+	**0.0845****	**0.0298+**	0.0106	0.0077
(neigh)	(3.72)	(1.36)	(3.41)	(1.69)	**(6.51)**	**(1.91)**	(1.41)	(0.88)
	[0.0035]	[0.0007]	[0.0019]	[0.0011]	**[0.0021]**	**[0.0006]**	[-0.0051]	[-0.0013]
%F-COV	0.0126*	-0.0010	0.0130**	0.0018	0.0189*	0.0062	**0.0067**	**0.0060**
(neigh)	(2.45)	(0.16)	(2.74)	(0.30)	(2.44)	(0.75)	**(1.36)**	**(1.12)**
	[0.0009]	[-0.0010]	[0.0006]	[-0.0002]	[0.0003]	[0.0001]	**[-0.0009]**	**[0.0014]**

Robust t statistics in parentheses + significant at 10%; * significant at 5%; ** significant at 1%.
Marginal evaluated at population means are given in square brackets.
Unemployed is the base category.

Table 12 - Multinomial Logit Coefficients and Marginal Effects – Social Interaction Variables - Females

Variables	Small Firm Uncovered		Firm Uncovered		Small Firm Covered		Firm Covered	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
%SF-UCOV	**0.0493****	**0.0292****	0.0328**	0.0196**	0.0474**	0.0362**	0.0121*	0.0078
(neigh)	**(6.09)**	**(3.92)**	(4.50)	(2.72)	(4.13)	(3.41)	(2.09)	(1.22)
	[0.0074]	**[0.0042]**	[0.0007]	[0.0004]	[0.0010]	[0.0009]	[-0.0056]	[-0.0033]
%F-UCOV	0.0240**	0.0044	**0.0248***	**0.0096**	0.0151	0.0081	0.0039	-0.0011
(neigh)	(2.66)	(0.52)	**(2.45)**	**(0.95)**	(1.09)	(0.60)	(0.47)	(0.13)
	[0.0037]	[0.0007]	**[0.0013]**	**[0.0008]**	[0.0001]	[0.0003]	[-0.0035]	[-0.0014]
%SF-COV	0.0659**	0.0308**	0.0382**	0.0306*	**0.1334****	**0.0753****	0.0133	0.0158
(neigh)	(7.03)	(3.17)	(4.06)	(2.51)	**(9.39)**	**(5.28)**	(1.49)	(1.60)
	[0.0094]	[0.0026]	[0.0002]	[0.0008]	**[0.0044]**	**[0.0023]**	[-0.0092]	[-0.0028]
%F-COV	0.0201**	0.0054	0.0205**	0.0137+	0.0213+	0.0170	**0.0159***	**0.0120+**
(neigh)	(2.67)	(0.73)	(3.10)	(1.78)	(1.89)	(1.62)	**(2.49)**	**(1.66)**
	[0.0013]	[-0.0011]	[0.0005]	[0.0005]	[0.0002]	[0.0004]	**[0.0000]**	**[0.0013]**

Robust t statistics in parentheses + significant at 10%; * significant at 5%; ** significant at 1%.
Marginal evaluated at population means are given in square brackets.
Unemployed is the base category.

Table 13 - Multinomial Logit Coefficients and Marginal Effects – Males (Model 1)

Variables	Small Firm - Uncovered	Firm - Uncovered	Small Firm - Covered	Firm - Covered
age	0.1054**	-0.0630**	0.1200**	0.1159**
	(15.59)	(9.41)	(13.46)	(23.21)
	[0.0073]	[-0.0194]	[0.0015]	[0.0192]
age2	-0.0008**	0.0008**	-0.0008**	-0.0013**
	(9.46)	(10.08)	(7.06)	(19.46)
	[0.0000]	[0.0002]	[0.0000]	[-0.0003]
Primary	-0.2941**	-0.2822**	0.2994**	0.2474**
Educ.	(10.47)	(9.59)	(5.19)	(9.87)
	[-0.0718]	[-0.0405]	[0.0095]	[0.1066]
Second.	-0.2569**	-0.1883**	0.6380**	0.6908**
Educ.	(7.28)	(5.48)	(11.62)	(24.72)
	[-0.1164]	[-0.0606]	[0.0118]	[0.1972]
College	-0.2795**	0.8942**	1.1123**	1.5201**
Educ.	(2.71)	(10.53)	(9.62)	(18.28)
	[-0.1823]	[-0.0205]	[0.0021]	[0.2763]
# Children	0.0495**	0.0726**	-0.0872**	0.0345**
under 5	(3.33)	(4.70)	(3.35)	(2.84)
	[0.0034]	[0.0052]	[-0.0041]	[-0.0005]
Household	0.8457**	0.6390**	1.0452**	0.9294**
head	(29.88)	(22.56)	(27.20)	(41.09)
	[0.0206]	[-0.0168]	[0.0093]	[0.0789]
Migrantes	-0.0140	0.1477**	-0.1131	0.0445
Up to 2	(0.25)	(2.87)	(0.90)	(0.92)
years	[-0.0118]	[0.0165]	[-0.0047]	[0.0040]
Migrantes	0.4575**	0.4969**	0.4002*	0.5206**
2 to 5	(4.21)	(4.68)	(2.13)	(5.06)
years	[0.0018]	[0.0067]	[-0.0016]	[0.0344]
Migrantes	0.1747**	0.0709*	0.2592**	0.1544**
> 5 years	(6.53)	(2.55)	(6.09)	(6.08)
	[0.0098]	[-0.0088]	[0.0043]	[0.0094]
Migrants	-0.0580	-0.1331**	0.0154	-0.3162**
outside	(1.54)	(3.26)	(0.24)	(7.90)
state	[0.0293]	[0.0058]	[0.0068]	[-0.0632]
Per capita	-0.0001**	-0.0001**	0.0000	-0.0001**
family inc.	(3.37)	(3.32)	(0.23)	(7.06)
(others)	[0.0000]	[0.0000]	[0.0000]	[0.0000]
D1994	0.4335**	0.2939**	0.0631	0.2057**
	(6.12)	(3.70)	(0.67)	(3.43)
	[0.0471]	[0.0060]	[-0.0058]	[-0.0214]
D1995	0.5380**	0.3992**	0.3767**	0.3486**
	(6.87)	(4.47)	(3.21)	(5.59)
	[0.0429]	[0.0040]	[0.0002]	[-0.0106]
D1996	0.2477**	0.1874**	-0.1248	0.1501*
	(3.79)	(2.61)	(1.26)	(2.47)
	[0.0230]	[0.0045]	[-0.0084]	[-0.0025]
D1997	0.2639**	0.1563*	-0.3120**	0.0557
	(4.02)	(2.10)	(3.83)	(0.95)
	[0.0402]	[0.0071]	[-0.0120]	[-0.0237]

Cont Table 13

D1998	0.0959	-0.0471	-0.6019**	-0.1397*
	(1.38)	(0.64)	(7.26)	(2.47)
	[0.0406]	[0.0022]	[-0.0147]	[-0.0353]
D1999	0.0170	0.0557	-0.6380**	-0.2058**
	(0.25)	(0.91)	(7.58)	(3.90)
	[0.0289]	[0.0229]	[-0.0147]	[-0.0487]
D2000	0.1454*	0.1888*	-0.3075**	-0.0974
	(2.09)	(2.39)	(3.20)	(1.63)
	[0.0338]	[0.0266]	[-0.0094]	[-0.0496]
D2001	-0.0358	-0.0988	-0.3416**	-0.3478**
	(0.57)	(1.38)	(4.32)	(6.34)
	[0.0391]	[0.0131]	[-0.0048]	[-0.0705]
D2002	0.0686	0.0132	-0.3789**	-0.2705**
	(1.14)	(0.17)	(4.11)	(4.71)
	[0.0467]	[0.0184]	[-0.0080]	[-0.0703]

Robust t statistics in parentheses + significant at 10%; * significant at 5%; ** significant at 1%.
Marginal evaluated at population means are given in square brackets.
Unemployed is the base category.

Table 14 - Multinomial Logit Coefficients and Marginal Effects – Females (Model 1)

Variables	Small Firm - Uncovered	Firm - Uncovered	Small Firm - Covered	Firm - Covered
age	0.1626**	-0.0116	-0.0116	-0.0116
	(22.97)	(1.49)	(1.49)	(1.49)
	[0.0131]	[-0.0144]	[0.0019]	[0.0152]
age2	-0.0010**	0.0006**	0.0006**	0.0006**
	(10.00)	(5.44)	(5.44)	(5.44)
	[-0.0001]	[0.0002]	[0.0000]	[-0.0002]
Primary Educ.	-0.5496**	0.0209	0.0209	0.0209
	(19.40)	(0.58)	(0.58)	(0.58)
	[-0.1357]	[0.0087]	[-0.0020]	[0.1213]
Second. Educ.	-0.9603**	0.1500**	0.1500**	0.1500**
	(26.14)	(3.55)	(3.55)	(3.55)
	[-0.2812]	[0.0029]	[-0.0066]	[0.2995]
College Educ.	-1.6100**	0.8093**	0.8093**	0.8093**
	(19.39)	(9.46)	(9.46)	(9.46)
	[-0.3166]	[0.0013]	[-0.0327]	[0.4197]
# Children under 5	0.0329**	-0.1161**	-0.1161**	-0.1161**
	(2.65)	(5.94)	(5.94)	(5.94)
	[0.0290]	[-0.0063]	[-0.0035]	[-0.0267]
Household head	0.1236**	0.0923*	0.0923*	0.0923*
	(3.28)	(2.27)	(2.27)	(2.27)
	[0.0180]	[0.0026]	[0.0049]	[-0.0169]
Migrantes Up to 2 years	0.2664**	-0.1354+	-0.1354+	-0.1354+
	(4.39)	(1.94)	(1.94)	(1.94)
	[0.1177]	[-0.0089]	[-0.0019]	[-0.1135]
Migrantes 2 to 5 years	0.3635**	-0.0162	-0.0162	-0.0162
	(3.32)	(0.14)	(0.14)	(0.14)
	[0.0646]	[-0.0192]	[0.0183]	[-0.0423]
Migrantes > 5 years	0.3563**	-0.0055	-0.0055	-0.0055
	(11.82)	(0.16)	(0.16)	(0.16)
	[0.0687]	[-0.0163]	[0.0104]	[-0.0439]
Migrants outside state	0.0300	-0.2576**	-0.2576**	-0.2576**
	(0.75)	(4.80)	(4.80)	(4.80)
	[0.0760]	[-0.0074]	[0.0120]	[-0.1068]
Per capita family inc. (others)	-0.0001**	-0.0000	-0.0000	-0.0000
	(2.73)	(0.01)	(0.01)	(0.01)
	[0.0000]	[0.0000]	[0.0000]	[0.0000]
D1994	0.1503+	0.1993*	0.1993*	0.1993*
	(1.95)	(2.41)	(2.41)	(2.41)
	[0.0119]	[0.0096]	[-0.0005]	[-0.0069]
D1995	0.2282**	0.3005**	0.3005**	0.3005**
	(2.84)	(3.44)	(3.44)	(3.44)
	[0.0061]	[0.0102]	[0.0073]	[0.0013]
D1996	0.0598	0.1057	0.1057	0.1057
	(0.87)	(1.48)	(1.48)	(1.48)
	[-0.0017]	[0.0044]	[-0.0064]	[0.0119]
D1997	-0.0425	-0.0747	-0.0747	-0.0747
	(0.56)	(0.86)	(0.86)	(0.86)
	[0.0021]	[-0.0028]	[-0.0093]	[0.0035]

Cont. Table 14

D1998	-0.1294+	-0.2082*	-0.3511**	-0.1995**
	(1.81)	(2.47)	(3.24)	(3.01)
	[0.0088]	[-0.0055]	[-0.0081]	[-0.0168]
D1999	-0.1460*	-0.0990	-0.2175*	-0.2871**
	(2.32)	(1.52)	(2.16)	(5.53)
	[0.0102]	[0.0087]	[-0.0018]	[-0.0418]
D2000	-0.1246+	0.0574	-0.0151	-0.1769**
	(1.85)	(0.73)	(0.14)	(3.05)
	[-0.0072]	[0.0183]	[0.0040]	[-0.0290]
D2001	-0.3853**	-0.3326**	-0.1975+	-0.5831**
	(5.83)	(4.16)	(1.77)	(9.15)
	[0.0000]	[0.0058]	[0.0091]	[-0.0730]
D2002	-0.2850**	-0.2626**	-0.0817	-0.5240**
	(3.90)	(3.22)	(0.66)	(7.37)
	[0.0108]	[0.0061]	[0.0117]	[-0.0755]

Robust t statistics in parentheses + significant at 10%; * significant at 5%; ** significant at 1%.
Marginal evaluated at population means are given in square brackets.
Unemployed is the base category.

Figure 13 – Marginal Effects for Social Interaction Variables – Model 1 - Males

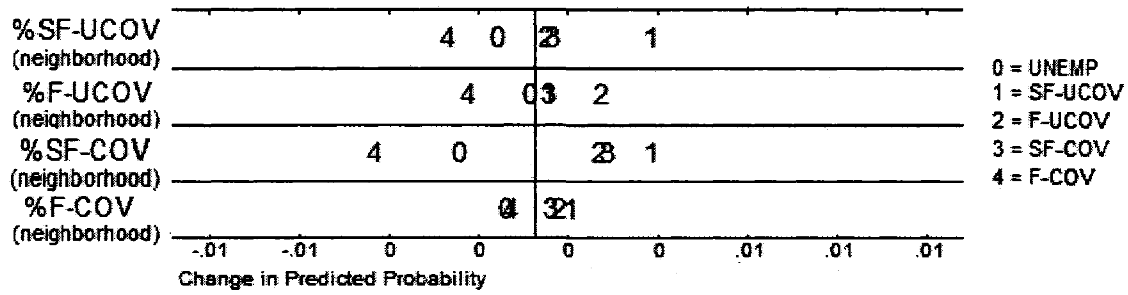


Figure 14 – Marginal Effects for Social Interaction Variables – Model 1 - Females

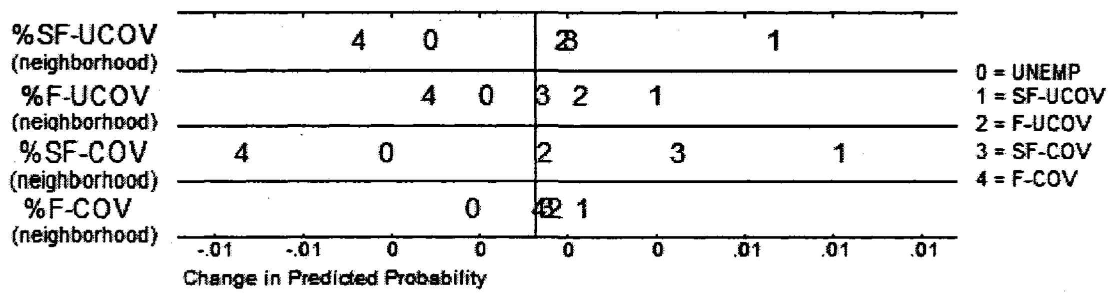


Figure 15 – Marginal Effects for Social Interaction Variables – Model 2 - Males

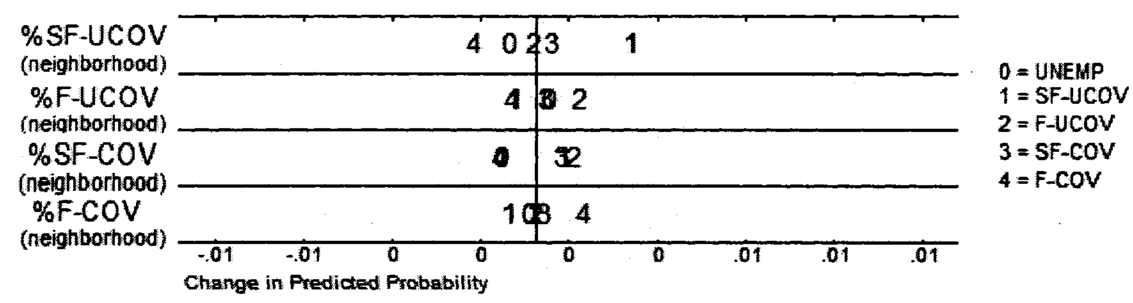


Figure 16 – Marginal Effects for Social Interaction Variables – Model 2 – Females

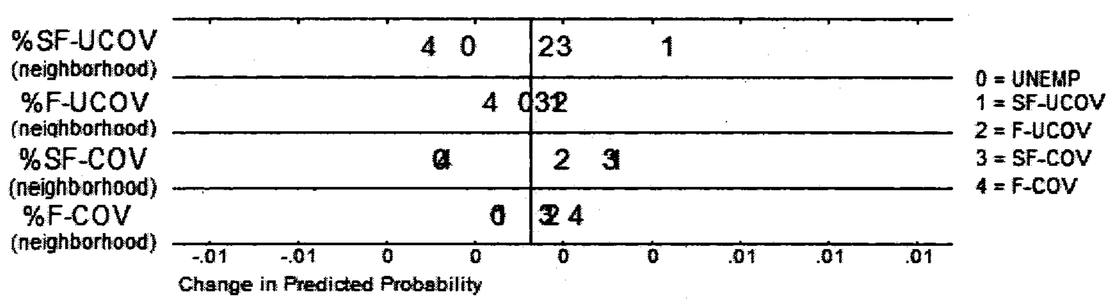


Table 15 - Multinomial Logit Coefficients and Marginal Effects – Self-Employed as Small Firm Workers - Males (Model 2)

Variables	Small Firm - Uncovered	Firm - Uncovered	Small Firm - Covered	Firm - Covered
age	0.1628**	-0.0612**	0.2337**	0.1213**
	(22.02)	(9.10)	(20.58)	(24.34)
	[0.0153]	[-0.0223]	[0.0023]	[0.0157]
age2	-0.0015**	0.0008**	-0.0019**	-0.0014**
	(15.90)	(9.74)	(14.16)	(20.68)
	[-0.0001]	[0.0003]	[0.0000]	[-0.0002]
Primary Educ.	-0.2721**	-0.2868**	0.4848**	0.2581**
	(8.85)	(9.76)	(7.25)	(10.19)
	[-0.0629]	[-0.0458]	[0.0077]	[0.1077]
Second. Educ.	-0.2079**	-0.2040**	0.9108**	0.7062**
	(5.95)	(5.97)	(15.33)	(25.53)
	[-0.1003]	[-0.0697]	[0.0103]	[0.1974]
College Educ.	-0.2913**	0.6093**	1.2893**	1.4923**
	(2.74)	(6.93)	(10.19)	(18.00)
	[-0.1593]	[-0.0516]	[0.0045]	[0.2858]
# Children under 5	0.0608**	0.0739**	-0.1853**	0.0348**
	(4.11)	(4.78)	(5.71)	(2.84)
	[0.0047]	[0.0052]	[-0.0035]	[-0.0018]
Household head	0.9404**	0.6345**	1.2679**	0.9299**
	(33.46)	(22.69)	(24.26)	(41.89)
	[0.0346]	[-0.0203]	[0.0075]	[0.0778]
Migrantes Up to 2 years	-0.1103+	0.1476**	-0.3439*	0.0311
	(1.92)	(2.92)	(2.19)	(0.65)
	[-0.0245]	[0.0212]	[-0.0048]	[0.0095]
Migrantes 2 to 5 years	0.4690**	0.4991**	0.5376*	0.5361**
	(4.19)	(4.72)	(2.21)	(5.19)
	[0.0010]	[0.0052]	[0.0012]	[0.0375]
Migrantes > 5 years	0.1890**	0.0780**	0.3539**	0.1586**
	(6.77)	(2.86)	(6.87)	(6.20)
	[0.0108]	[-0.0087]	[0.0036]	[0.0104]
Migrants outside state	-0.0422	-0.1354**	0.1469*	-0.3102**
	(1.07)	(3.34)	(2.04)	(7.83)
	[0.0297]	[0.0060]	[0.0058]	[-0.0644]
Per capita family inc. (others)	-0.0001**	-0.0001**	0.0000+	-0.0001**
	(3.43)	(3.38)	(1.88)	(7.00)
	[0.0000]	[0.0000]	[0.0000]	[0.0000]
D1994	0.4594**	0.2904**	0.1180	0.2077**
	(6.45)	(3.75)	(1.22)	(3.49)
	[0.0472]	[0.0054]	[-0.0020]	[-0.0230]
D1995	0.5350**	0.4019**	0.4586**	0.3458**
	(6.86)	(4.61)	(3.77)	(5.69)
	[0.0389]	[0.0058]	[0.0016]	[-0.0085]
D1996	0.2203**	0.1688*	-0.1475	0.1814**
	(3.32)	(2.29)	(1.29)	(3.04)
	[0.0124]	[0.0009]	[-0.0044]	[0.0095]
D1997	0.1989**	0.1229	-0.5268**	0.0973+
	(2.80)	(1.56)	(5.25)	(1.68)
	[0.0207]	[0.0028]	[-0.0079]	[-0.0035]

Cont. Table 15

D1998	-0.0260	-0.0919	-0.9065**	-0.1064+
	(0.36)	(1.15)	(10.41)	(1.88)
	[0.0123]	[-0.0013]	[-0.0096]	[-0.0117]
D1999	-0.1231+	-0.0007	-0.9885**	-0.1840**
	(1.76)	(0.01)	(8.63)	(3.55)
	[0.0012]	[0.0199]	[-0.0099]	[-0.0274]
D2000	0.0082	0.1145	-0.6970**	-0.0765
	(0.11)	(1.36)	(5.44)	(1.32)
	[0.0075]	[0.0221]	[-0.0082]	[-0.0246]
D2001	-0.2059**	-0.1676*	-0.7862**	-0.3194**
	(3.01)	(2.04)	(7.29)	(5.85)
	[0.0064]	[0.0103]	[-0.0071]	[-0.0408]
D2002	-0.0667	-0.0649	-0.7024**	-0.2371**
	(1.02)	(0.77)	(6.35)	(3.95)
	[0.0180]	[0.0129]	[-0.0071]	[-0.0429]
Region 1	-0.0071	-0.0435	-0.0951	-0.0660+
	(0.10)	(0.69)	(0.90)	(1.66)
	[0.0076]	[-0.0001]	[-0.0008]	[-0.0118]
Region 3	0.0191	0.0052	-0.2796**	-0.0540
	(0.30)	(0.08)	(3.05)	(1.44)
	[0.0097]	[0.0047]	[-0.0037]	[-0.0139]
Region 4	0.0931	0.1190*	-0.0570	0.0557
	(1.47)	(2.01)	(0.70)	(1.28)
	[0.0060]	[0.0081]	[-0.0018]	[-0.0046]
Region 5	-0.0051	0.0165	-0.2675**	-0.0381
	(0.08)	(0.25)	(3.12)	(0.98)
	[0.0035]	[0.0057]	[-0.0036]	[-0.0083]
Region 6	-0.0625	-0.0770	-0.3677*	-0.0977**
	(0.97)	(1.03)	(2.43)	(2.60)
	[0.0035]	[0.0003]	[-0.0041]	[-0.0094]
INCNEIGH	0.0018**	0.0011*	0.0019*	-0.0009+
	(3.56)	(2.24)	(2.38)	(1.93)
	[0.0004]	[0.0001]	[0.0000]	[-0.0005]
%SF-UCOV (neigh)	0.0208**	0.0068	0.0148	0.0039
	(3.24)	(1.06)	(1.30)	(0.78)
	[0.0028]	[-0.0001]	[0.0001]	[-0.0019]
%F-UCOV (neigh)	-0.0031	0.0053	0.0081	-0.0041
	(0.40)	(0.66)	(0.61)	(0.71)
	[-0.0003]	[0.0011]	[0.0002]	[-0.0012]
%SF-COV (neigh)	0.0153	0.0170	0.0327+	0.0075
	(1.55)	(1.60)	(1.80)	(0.86)
	[0.0011]	[0.0010]	[0.0004]	[-0.0013]
%F-COV (neigh)	0.0014	0.0014	0.0106	0.0064
	(0.21)	(0.24)	(0.95)	(1.20)
	[0.0004]	[0.0001]	[0.0000]	[-0.0005]

Cont. Table 15

	Wald Test for Combining Labor Force States (Chi-Square value)			
H0: $\beta_{Unemp} = \beta_{column}$	9152**	1611**	6866**	9974**
H0: $\beta_{sf\text{-}ucov} = \beta_{column}$		6927**	2563**	8322**
H0: $\beta_{F\text{-}ucov} = \beta_{column}$			5032**	7525**
H0: $\beta_{sf\text{-}cov} = \beta_{column}$				3428**

Robust t statistics in parentheses + significant at 10%; * significant at 5%; ** significant at 1%.
Marginal evaluated at population means are given in square brackets.
Unemployed is the base category.

Table 16 - Multinomial Logit Coefficients and Marginal Effects – Self-Employed as Small Firm Workers - Females (Model 2)

Variables	Small Firm - Uncovered	Firm - Uncovered	Small Firm – Covered	Firm - Covered
age	0.2885**	-0.0024	0.3547**	0.1764**
	(40.12)	(0.32)	(21.79)	(23.46)
	[0.0305]	[-0.0198]	[0.0030]	[0.0095]
age2	-0.0024**	0.0005**	-0.0027**	-0.0015**
	(23.84)	(4.47)	(13.50)	(13.25)
	[-0.0003]	[0.0002]	[0.0000]	[-0.0001]
Primary Educ.	-0.2322**	0.0425	0.3634**	0.2935**
	(7.80)	(1.14)	(4.97)	(8.87)
	[-0.0719]	[-0.0084]	[0.0041]	[0.0921]
Second. Educ.	-0.5451**	0.1690**	0.4036**	0.8781**
	(13.67)	(3.79)	(6.59)	(23.14)
	[-0.1835]	[-0.0235]	[0.0005]	[0.2568]
College Educ.	-1.3444**	0.6948**	0.0806	1.5735**
	(15.47)	(7.92)	(0.60)	(22.27)
	[-0.2304]	[-0.0308]	[-0.0092]	[0.3657]
# Children under 5	0.0274*	-0.1190**	-0.1595**	-0.1288**
	(1.98)	(5.94)	(4.43)	(9.50)
	[0.0232]	[-0.0057]	[-0.0013]	[-0.0269]
Household head	0.1771**	0.0949*	0.3011**	0.0378
	(4.53)	(2.34)	(5.19)	(1.10)
	[0.0239]	[0.0022]	[0.0036]	[-0.0186]
Migrantes Up to 2 years	-0.1174	-0.1583*	-0.5725*	-0.4406**
	(1.59)	(2.28)	(2.43)	(7.44)
	[0.0331]	[0.0121]	[-0.0042]	[-0.0832]
Migrantes 2 to 5 years	0.2858*	-0.0389	-0.2627	0.0371
	(2.16)	(0.33)	(0.65)	(0.37)
	[0.0519]	[-0.0143]	[-0.0044]	[-0.0215]
Migrantes > 5 years	0.3876**	-0.0178	0.4646**	0.0298
	(11.27)	(0.50)	(7.37)	(1.14)
	[0.0672]	[-0.0160]	[0.0058]	[-0.0403]
Migrants outside state	0.0772+	-0.2649**	0.1817*	-0.4795**
	(1.88)	(4.94)	(2.26)	(11.63)
	[0.0769]	[-0.0059]	[0.0072]	[-0.1133]
Per capita family inc. (others)	-0.0001**	0.0000	0.0000*	-0.0000
	(2.75)	(0.00)	(1.99)	(1.32)
	[0.0000]	[0.0000]	[0.0000]	[0.0000]
D1994	0.1340+	0.1938*	0.0650	0.0852
	(1.83)	(2.39)	(0.44)	(1.25)
	[0.0084]	[0.0124]	[-0.0005]	[-0.0063]
D1995	0.2645**	0.3341**	0.4592**	0.2294**
	(3.62)	(3.91)	(2.83)	(3.26)
	[0.0095]	[0.0143]	[0.0039]	[0.0026]
D1996	0.0817	0.1871*	-0.2081	0.1293*
	(1.22)	(2.45)	(1.58)	(2.15)
	[-0.0049]	[0.0108]	[-0.0042]	[0.0129]
D1997	-0.0744	0.0191	-0.6092**	-0.0017
	(0.96)	(0.20)	(4.15)	(0.02)
	[-0.0117]	[0.0053]	[-0.0072]	[0.0102]

Cont. Table 16

D1998	-0.2688**	-0.1290	-0.8104**	-0.1834*
	(3.50)	(1.42)	(5.44)	(2.50)
	[-0.0211]	[0.0057]	[-0.0076]	[-0.0046]
D1999	-0.3458**	-0.0443	-0.9334**	-0.2872**
	(4.51)	(0.62)	(6.27)	(5.03)
	[-0.0259]	[0.0246]	[-0.0082]	[-0.0276]
D2000	-0.3858**	0.0975	-0.7718**	-0.1818**
	(5.30)	(1.10)	(5.05)	(3.06)
	[-0.0469]	[0.0362]	[-0.0073]	[-0.0081]
D2001	-0.6678**	-0.2827**	-0.9134**	-0.5930**
	(8.76)	(3.31)	(6.57)	(8.59)
	[-0.0432]	[0.0243]	[-0.0057]	[-0.0583]
D2002	-0.6198**	-0.2338*	-0.7640**	-0.5375**
	(7.95)	(2.52)	(6.07)	(7.05)
	[-0.0422]	[0.0252]	[-0.0045]	[-0.0528]
Region 1	-0.2948**	-0.1846**	-0.1030	-0.0083
	(4.86)	(2.98)	(0.95)	(0.15)
	[-0.0440]	[-0.0112]	[-0.0002]	[0.0413]
Region 3	-0.0525	-0.0037	-0.2047	0.0548
	(0.77)	(0.05)	(1.57)	(0.84)
	[-0.0145]	[-0.0019]	[-0.0030]	[0.0211]
Region 4	0.0417	0.1002	0.0019	0.1539**
	(0.69)	(1.61)	(0.02)	(2.59)
	[-0.0126]	[0.0004]	[-0.0014]	[0.0277]
Region 5	0.0381	0.1504*	-0.2710*	0.0219
	(0.59)	(2.53)	(2.15)	(0.36)
	[0.0007]	[0.0149]	[-0.0042]	[-0.0063]
Region 6	-0.0626	-0.0756	-0.1797	-0.0037
	(1.03)	(1.33)	(1.32)	(0.06)
	[-0.0079]	[-0.0058]	[-0.0022]	[0.0117]
INCNEIGH	0.0010*	-0.0009	0.0012	-0.0006
	(2.16)	(1.61)	(1.48)	(1.19)
	[0.0003]	[-0.0001]	[0.0000]	[-0.0002]
%SF-	0.0261**	0.0193**	0.0354**	0.0069
UCOV	(3.38)	(2.69)	(3.08)	(1.07)
(neigh)	[0.0032]	[0.0009]	[0.0003]	[-0.0026]
%F-UCOV	0.0070	0.0107	0.0243	0.0004
(neigh)	(0.76)	(1.05)	(1.59)	(0.04)
	[0.0008]	[0.0009]	[0.0003]	[-0.0015]
%SF-COV	0.0253*	0.0322**	0.0742**	0.0152
(neigh)	(2.35)	(2.61)	(4.14)	(1.53)
	[0.0016]	[0.0017]	[0.0008]	[-0.0015]
%F-COV	0.0059	0.0137+	0.0208+	0.0117
(neigh)	(0.72)	(1.77)	(1.91)	(1.58)
	[-0.0007]	[0.0006]	[0.0002]	[0.0013]

Cont. Table 16

Wald Test for Combining Labor Force States (Chi-Square value)				
H0:β_{Unemp} = β_{column}	13985**	1300**	5735**	7356**
H0:$\beta_{sf-ucov}$ = β_{column}		9978**	1099**	13089**
H0:β_{F-ucov} = β_{column}			2625**	3277**
H0:β_{sf-cov} = β_{column}				2009**

Robust t statistics in parentheses + significant at 10%; * significant at 5%; ** significant at 1%.
Marginal evaluated at population means are given in square brackets.
Unemployed is the base category.

Figure 17 – Logit Coefficients and Odds Ratio for Educational Variables – Self-Employed as Small Firm Worker - Model 2 – Males

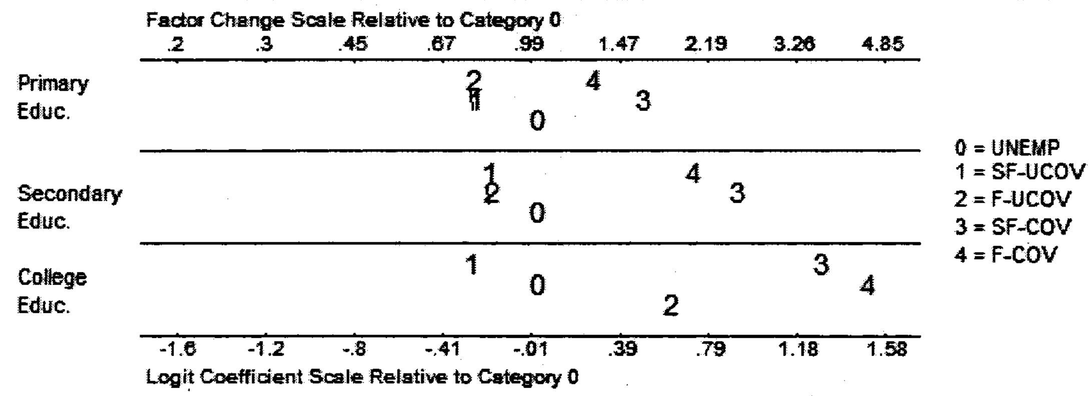


Figure 18 – Logit Coefficients and Odds Ratio for Educational Variables – Self-Employed as Small Firm Worker - Model 2 – Females

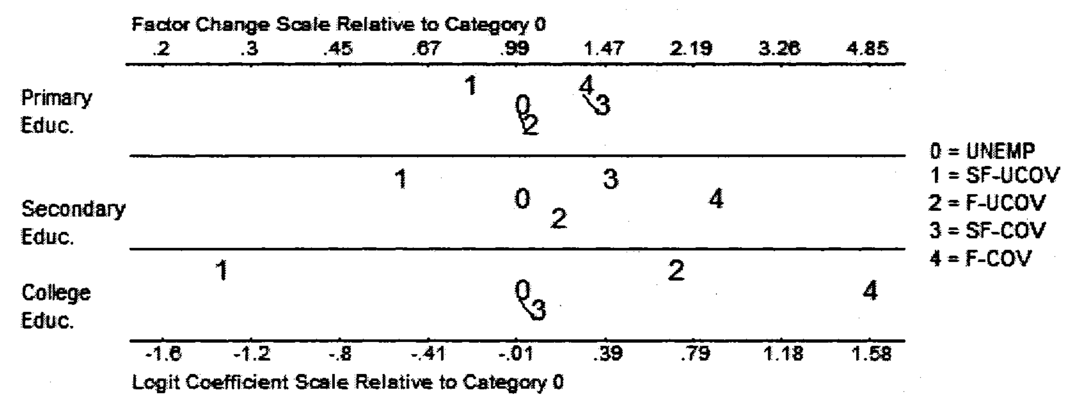


Figure 19 – Marginal Effects for Social Interaction Variables – Self-Employed as Small Firm Worker - Model 2 – Males

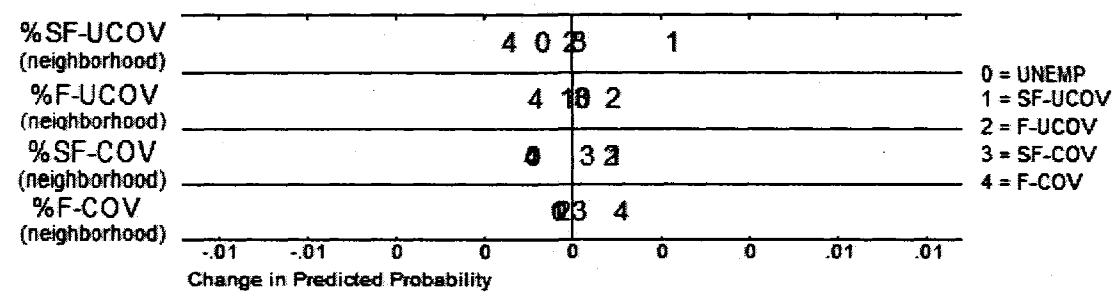


Figure 20 – Marginal Effects for Social Interaction Variables – Self-Employed as Small Firm Worker - Model 2 – Females

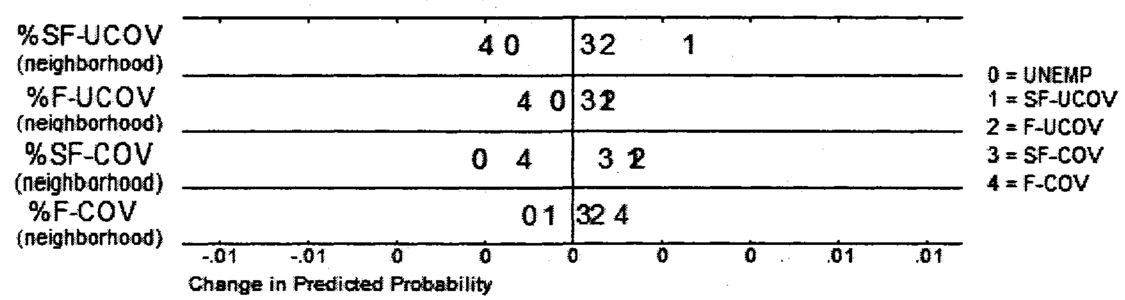


Figure 21 – Logit Coefficients and Odds Ratio for Educational Variables – Models With (Panel A) and Without (Panel B) Unemployed Workers – Males

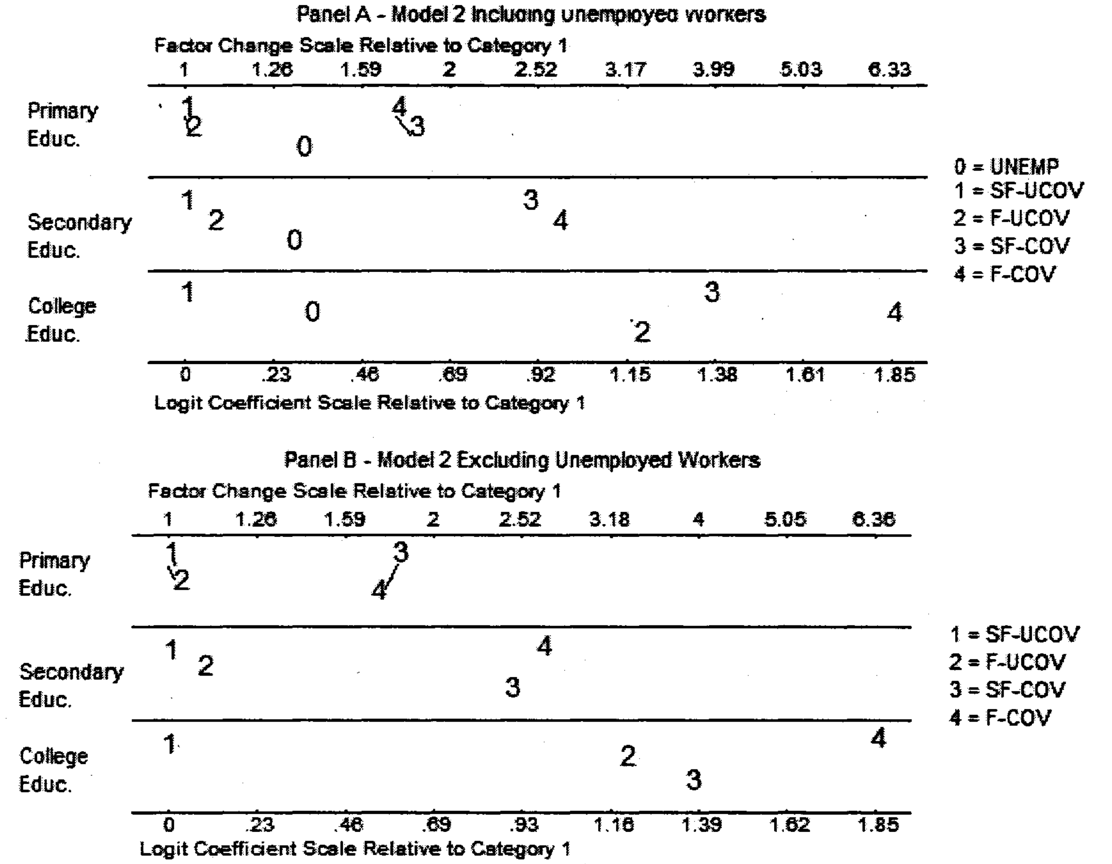


Figure 22 – Logit Coefficients and Odds Ratio for Educational Variables – Models With (Panel A) and Without (Panel B) Unemployed Workers – Females

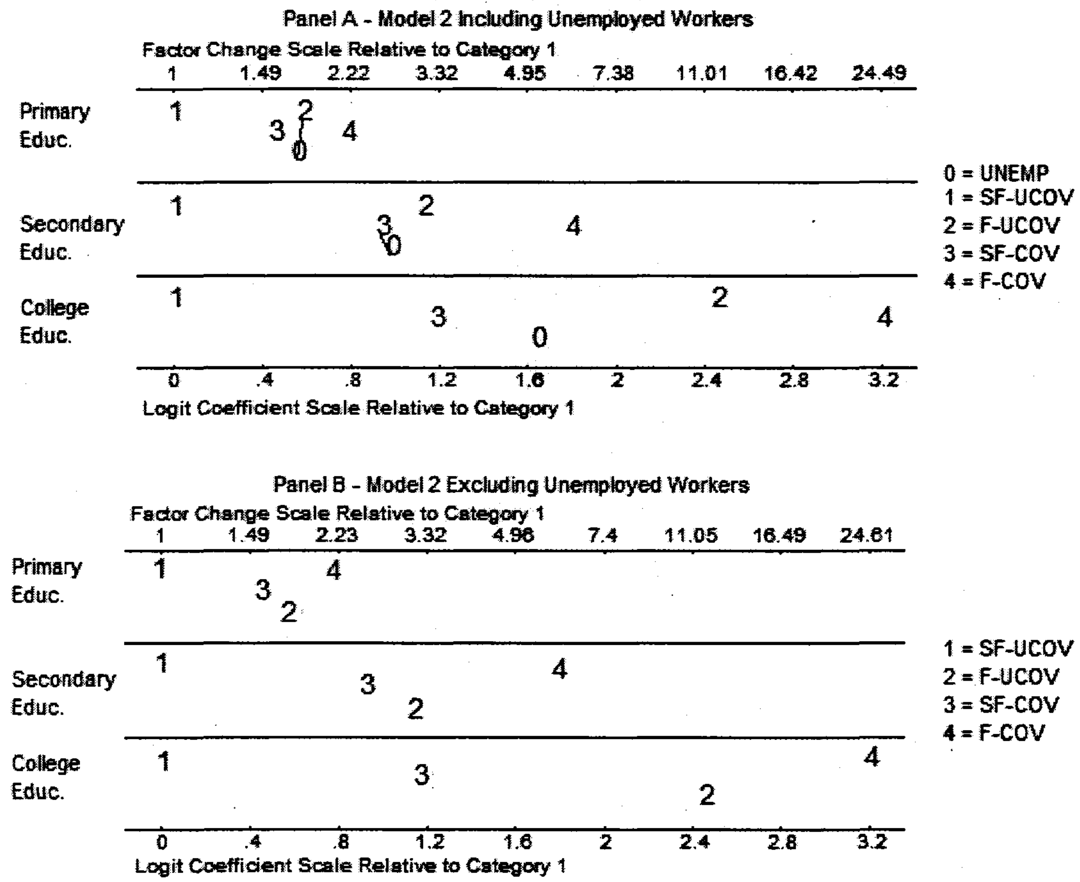


Table 17 - Multinomial Logit Coefficients – Excluding Unemployed Workers – Males (Model 2)

	Firm - Uncovered	Small Firm – Covered	Firm - Covered
age	-0.1703**	0.0143+	0.0088
	(28.75)	(1.76)	(1.50)
age2	0.0017**	0.0001	-0.0004**
	(23.44)	(0.58)	(6.07)
Primary Educ.	0.0223	0.5917**	0.5490**
	(1.07)	(10.99)	(23.97)
Second. Educ.	0.0854**	0.8918**	0.9704**
	(2.94)	(19.41)	(42.34)
College Educ.	1.1967**	1.3716**	1.8420**
	(14.26)	(14.90)	(28.61)
# Children under 5	0.0221+	-0.1325**	-0.0180+
	(1.77)	(5.66)	(1.67)
Household head	-0.2138**	0.2092**	0.0733**
	(7.92)	(6.00)	(3.28)
Migrantes Up to 2 years	0.1458**	-0.0797	0.0580
	(2.95)	(0.75)	(1.26)
Migrantes 2 to 5 years	0.0293	-0.0404	0.0600
	(0.36)	(0.26)	(0.95)
Migrantes > 5 years	-0.1100**	0.1020*	-0.0236
	(4.38)	(2.51)	(1.03)
Migrants outside state	-0.0778*	0.0729	-0.2558**
	(2.10)	(1.21)	(6.94)
Per capita family inc. (others)	-0.0000	0.0001**	-0.0000*
	(0.39)	(2.85)	(2.08)
D1994	-0.1424**	-0.3719**	-0.2275**
	(2.66)	(4.64)	(4.41)
D1995	-0.1361*	-0.1688+	-0.1844**
	(2.49)	(1.65)	(3.78)
D1996	-0.0276	-0.3772**	-0.0116
	(0.51)	(4.45)	(0.25)
D1997	-0.0692	-0.5865**	-0.0871+
	(1.17)	(7.51)	(1.77)
D1998	-0.1041+	-0.7306**	-0.1029+
	(1.72)	(8.77)	(1.96)
D1999	0.0719	-0.6982**	-0.1032+
	(1.18)	(8.16)	(1.85)
D2000	0.0847	-0.4919**	-0.1063+
	(1.31)	(4.83)	(1.91)
D2001	-0.0154	-0.3300**	-0.1604**
	(0.26)	(3.43)	(2.85)
D2002	-0.0046	-0.4441**	-0.1737**
	(0.07)	(5.00)	(3.55)
Region 1	-0.0301	-0.2354**	-0.0508
	(0.51)	(3.07)	(0.82)
Region 3	0.0015	-0.3716**	-0.0641
	(0.03)	(5.42)	(1.28)
Region 4	0.0163	-0.2062**	-0.0450
	(0.36)	(3.41)	(0.88)
Region 5	0.0238	-0.3105**	-0.0296
	(0.52)	(5.02)	(0.56)
Region 6	-0.0343	-0.3333**	-0.0564
	(0.66)	(3.10)	(1.03)

Cont. Table 17

INCNEIGH	-0.0010+	0.0001	-0.0029**
	(1.83)	(0.18)	(6.35)
%SF-UCOV (neigh)	-0.0134**	0.0002	-0.0159**
	(2.58)	(0.02)	(3.74)
%F-UCOV (neigh)	0.0120	0.0084	0.0017
	(1.64)	(0.81)	(0.31)
%SF-COV (neigh)	0.0049	0.0162	-0.0056
	(0.69)	(1.19)	(0.90)
%F-COV (neigh)	0.0027	0.0074	0.0071+
	(0.58)	(1.00)	(1.79)

Robust z statistics in parentheses + significant at 10%; * significant at 5%; ** significant at 1%.
Small-firm uncovered worker is the base category.

Table 18 - Multinomial Logit Coefficients – Excluding Unemployed Workers – Females (Model 2)

	Firm - Uncovered	Small Firm – Covered	Firm - Covered
age	-0.1797**	0.0009	-0.0074
	(24.24)	(0.09)	(1.17)
age2	0.0017**	0.0001	-0.0002*
	(17.49)	(0.92)	(2.27)
Primary Educ.	0.5899**	0.4479**	0.7815**
	(17.54)	(9.13)	(27.28)
Second. Educ.	1.1387**	0.9298**	1.7927**
	(27.97)	(18.38)	(47.34)
College Educ.	2.4685**	1.1801**	3.2033**
	(37.48)	(11.63)	(46.13)
# Children under 5	-0.1531**	-0.1693**	-0.1636**
	(8.56)	(6.46)	(12.08)
Household head	-0.0456	0.0482	-0.1110**
	(1.36)	(1.23)	(3.70)
Migrantes Up to 2 years	-0.4268**	-0.3379*	-0.6487**
	(5.92)	(2.43)	(12.38)
Migrantes 2 to 5 years	-0.3886**	0.1613	-0.2928**
	(3.09)	(0.79)	(3.15)
Migrantes > 5 years	-0.3766**	0.0295	-0.3246**
	(12.57)	(0.65)	(14.76)
Migrants outside state	-0.3008**	0.0170	-0.5241**
	(5.66)	(0.25)	(13.40)
Per capita family inc. (others)	0.0001**	0.0001**	0.0000*
	(2.68)	(4.61)	(1.98)
D1994	0.0484	-0.0513	-0.0540
	(0.78)	(0.48)	(1.07)
D1995	0.0574	0.1183	-0.0459
	(0.89)	(0.93)	(0.97)
D1996	0.0650	-0.1435	0.0209
	(0.95)	(1.45)	(0.48)
D1997	0.0120	-0.2235*	0.0139
	(0.16)	(2.11)	(0.28)
D1998	-0.0267	-0.2236*	-0.0367
	(0.41)	(2.18)	(0.83)
D1999	0.0890	-0.0836	-0.1095*
	(1.36)	(0.82)	(2.11)
D2000	0.2443**	0.1100	-0.0026
	(3.42)	(1.05)	(0.05)
D2001	0.1133+	0.2042+	-0.1415**
	(1.66)	(1.81)	(2.65)
D2002	0.0974	0.2576*	-0.1626**
	(1.34)	(2.46)	(3.10)
Region 1	0.1908**	-0.1541*	0.3510**
	(3.05)	(2.09)	(5.83)
Region 3	0.0881+	-0.3667**	0.1471**
	(1.70)	(4.75)	(2.99)
Region 4	0.1030+	-0.2788**	0.1521**
	(1.84)	(4.00)	(3.42)
Region 5	0.1561**	-0.3996**	0.0271
	(2.65)	(5.79)	(0.51)
Region 6	-0.0122	-0.3147**	0.0548
	(0.20)	(3.14)	(0.88)

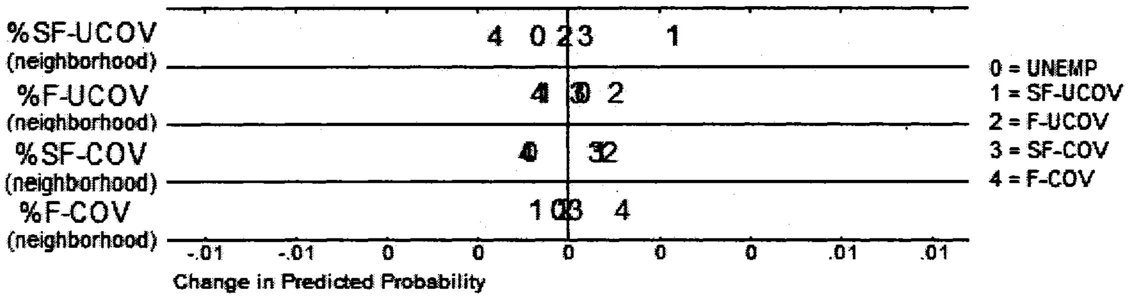
Cont. Table 18

INCNEIGH	-0.0016**	-0.0008	-0.0015**
	(2.68)	(1.24)	(3.32)
%SF-UCOV (neigh)	-0.0094	0.0070	-0.0216**
	(1.48)	(0.84)	(3.83)
%F-UCOV (neigh)	0.0049	0.0036	-0.0064
	(0.63)	(0.36)	(0.99)
%SF-COV (neigh)	-0.0008	0.0441**	-0.0160+
	(0.07)	(3.30)	(1.87)
%F-COV (neigh)	0.0080	0.0114	0.0062
	(1.31)	(1.40)	(1.29)

Robust z statistics in parentheses + significant at 10%; * significant at 5%; ** significant at 1%.
Small-firm uncovered worker is the base category.

Figure 23 – Marginal Effects for Social Interaction Variables – Models With (Panel A) and Without (Panel B) Unemployed Workers – Males (Model 2)

Panel A - Model 2 Including unemployed workers



Panel B - Model 2 Excluding Unemployed Workers

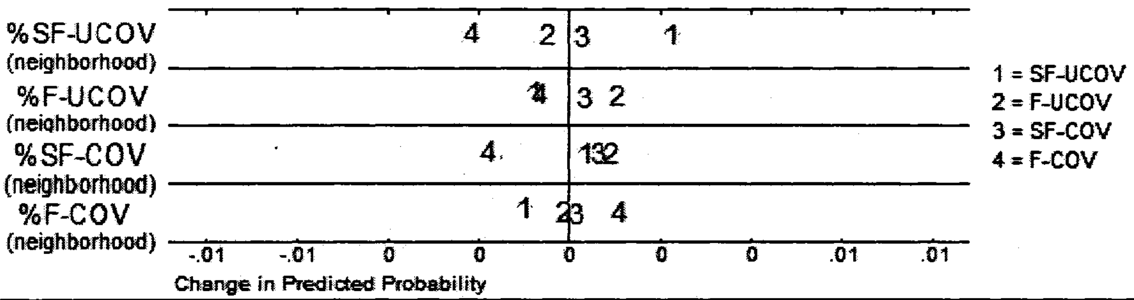


Figure 24 – Marginal Effects for Social Interaction Variables – Models With (Panel A) and Without (Panel B) Unemployed Workers – Females (Model 2)

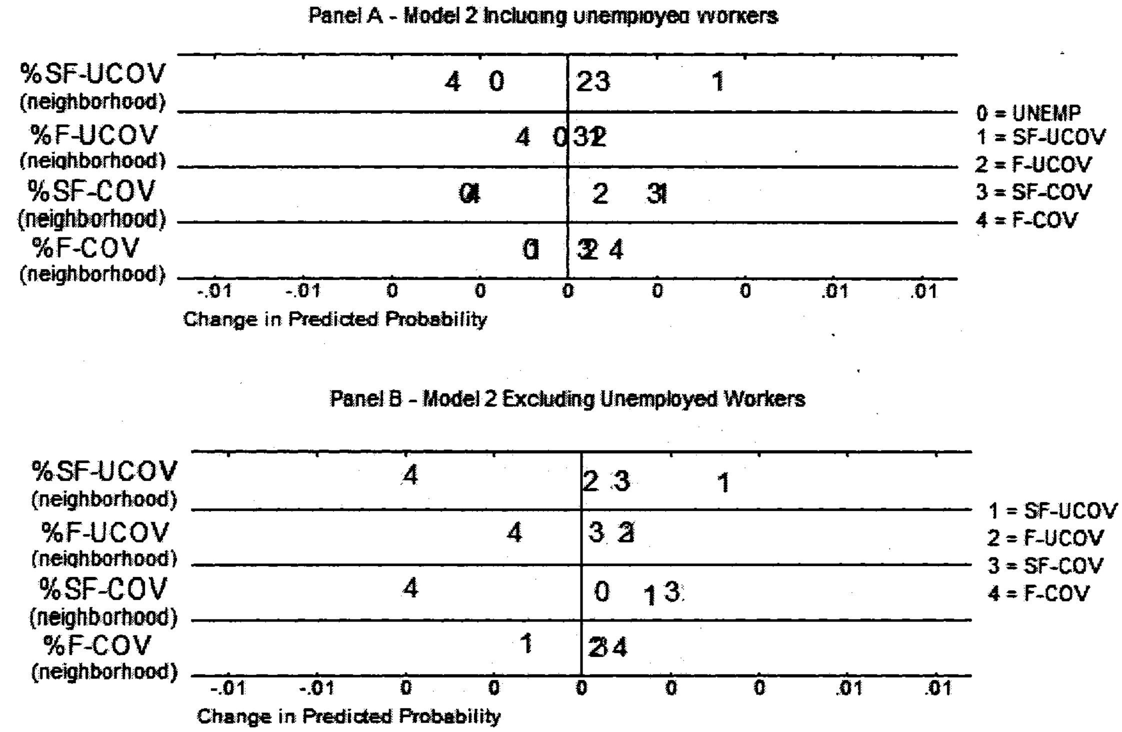


Table 19 – Bivariate Probit Coefficients
(Firm x Small Firm Workers) and (Contributors x Non-Contributors) - Males

	Firm workers	Social Security Contributors
age	-0.0147**	0.0473**
	(5.26)	(16.99)
age2	-0.0001	-0.0007**
	(1.60)	(18.89)
Primary Educ.	0.2194**	0.3412**
	(21.36)	(26.89)
Second. Educ.	0.4040**	0.5804**
	(34.27)	(49.85)
College Educ.	0.7918**	0.7877**
	(27.46)	(31.37)
# Children under 5	0.0058	-0.0173**
	(0.92)	(3.17)
Household head	-0.0142	0.1033**
	(1.16)	(9.51)
Migrantes Up to 2 years	0.0549*	-0.0096
	(2.26)	(0.42)
Migrantes 2 to 5 years	0.0361	0.0329
	(1.04)	(1.06)
Migrantes > 5 years	-0.0366**	0.0129
	(3.16)	(1.04)
Migrants outside state	-0.1412**	-0.1228**
	(7.30)	(6.42)
Per capita family inc. (others)	-0.0000**	-0.0000
	(3.64)	(1.61)
D1994	-0.0845**	-0.1162**
	(3.05)	(4.44)
D1995	-0.0810**	-0.0824**
	(3.05)	(3.01)
D1996	0.0395	-0.0222
	(1.45)	(0.92)
D1997	0.0185	-0.0650*
	(0.70)	(2.52)
D1998	0.0177	-0.0756**
	(0.62)	(2.88)
D1999	0.0352	-0.1121**
	(1.20)	(4.05)
D2000	0.0146	-0.1067**
	(0.49)	(3.63)
D2001	-0.0420	-0.1056**
	(1.59)	(3.20)
D2002	-0.0353	-0.1225**
	(1.32)	(4.75)
Region 1	-0.0009	-0.0359
	(0.03)	(1.17)
Region 3	0.0067	-0.0568*
	(0.25)	(2.15)
Region 4	0.0003	-0.0418
	(0.01)	(1.52)
Region 5	0.0160	-0.0404
	(0.59)	(1.45)
Region 6	0.0014	-0.0444
	(0.05)	(1.49)

Cont. Table 19

INCNEIGH	-0.0016**	-0.0014**
	(5.97)	(6.50)
%SF-UCOV (neigh)	-0.0095**	-0.0060**
	(4.12)	(2.71)
%F-UCOV (neigh)	0.0010	-0.0011
	(0.36)	(0.35)
%SF-COV (neigh)	-0.0047	-0.0033
	(1.36)	(0.91)
%F-COV (neigh)	0.0026	0.0036+
	(1.24)	(1.70)
	Wald Tests (Chi-Square Values)	
H0 : $\rho = 0$	8520.35**	
Primary Educ. H0 : $\beta_{\text{Firm workers}} = \beta_{\text{Social Sec. Contributers}}$	165.04**	
Secondary Educ. H0 : $\beta_{\text{Firm workers}} = \beta_{\text{Social Sec. Contributers}}$	345.41**	
College Educ. H0 : $\beta_{\text{Firm workers}} = \beta_{\text{Social Sec. Contributers}}$	0.02	

Robust z statistics in parentheses
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 20 – Bivariate Probit Coefficients
(Firm x Small Firm Workers) and (Contributors x Non-Contributors) - Females

	Firm workers	Social Security Contributors
age	-0.0263**	0.0300**
	(7.97)	(11.27)
age2	0.0001*	-0.0004**
	(2.54)	(11.45)
Primary Educ.	0.4255**	0.3962**
	(27.17)	(26.73)
Second. Educ.	0.9305**	0.8761**
	(43.33)	(47.39)
College Educ.	1.6353**	1.3091**
	(49.56)	(44.15)
# Children under 5	-0.0826**	-0.0732**
	(11.37)	(10.01)
Household head	-0.0649**	-0.0482**
	(4.08)	(3.00)
Migrantes Up to 2 years	-0.3155**	-0.2936**
	(11.38)	(9.72)
Migrantes 2 to 5 years	-0.1954**	-0.0638
	(3.92)	(1.25)
Migrantes > 5 years	-0.2023**	-0.1144**
	(17.28)	(9.32)
Migrants outside state	-0.2820**	-0.2186**
	(13.77)	(10.32)
Per capita family inc. (others)	0.0000	0.0000+
	(0.74)	(1.67)
D1994	-0.0193	-0.0445
	(0.71)	(1.54)
D1995	-0.0256	-0.0251
	(0.97)	(0.88)
D1996	0.0306	-0.0143
	(1.33)	(0.57)
D1997	0.0258	-0.0186
	(0.91)	(0.72)
D1998	-0.0038	-0.0385
	(0.15)	(1.62)
D1999	-0.0340	-0.0832**
	(1.19)	(3.00)
D2000	0.0138	-0.0393
	(0.52)	(1.33)
D2001	-0.0713*	-0.0799*
	(2.50)	(2.54)
D2002	-0.0909**	-0.0840**
	(3.28)	(2.82)
Region 1	0.2071**	0.1350**
	(7.04)	(4.42)
Region 3	0.1109**	0.0271
	(4.33)	(1.13)
Region 4	0.1107**	0.0330
	(4.66)	(1.49)
Region 5	0.0669*	-0.0525*
	(2.15)	(2.16)
Region 6	0.0531	-0.0081
	(1.54)	(0.28)

Cont. Table 20

INCNEIGH	-0.0008**	-0.0005*
	(3.24)	(2.40)
%SF-UCOV (neigh)	-0.0119**	-0.0093**
	(4.00)	(3.44)
%F-UCOV (neigh)	-0.0033	-0.0044
	(0.97)	(1.25)
%SF-COV (neigh)	-0.0116*	-0.0044
	(2.51)	(1.01)
%F-COV (neigh)	0.0024	0.0026
	(0.95)	(1.07)
	Wald Tests (Chi-Square Values)	
H0 : $\rho = 0$	6527.37**	
Primary Educ.		
H0 : $\beta_{\text{Firm workers}} = \beta_{\text{Social Sec. Contributers}}$	8.58**	
Secondary Educ.		
H0 : $\beta_{\text{Firm workers}} = \beta_{\text{Social Sec. Contributers}}$	19.17**	
College Educ.		
H0 : $\beta_{\text{Firm workers}} = \beta_{\text{Social Sec. Contributers}}$	198.69**	

Robust z statistics in parentheses
+ significant at 10%; * significant at 5%; ** significant at 1%

BIBLIOGRAPHY

Aizer, A. and Currie, J.(2002) – Networks or neighborhoods? Correlations in the use of publicly-funded maternity care in California – *National Bureau of Economic Research* Working Paper 9209.

Bartus, T. (2005) – Estimation of marginal effects using margeff - Manuscript submitted to Stata Journal.

Berger, M. and Buvinic, M. (1989) – Women's Ventures: Assistance to the informal sector in Latin America – *Hartford: Kumarian Press.*

Borjas, G.J. (1995) – Ethinicity, neighborhoods, and human-capital externalities – *The American Economic Review* 85 (3) june.

Brock, W. A. and Durlauf, S. (2002) – A multinomial-choice model of neighborhood effects – *American Economic Review* 92 (2) pp. 298-303.

Carman, K. G. (2004) – Social influences and the private provision of public goods: evidence from charitable contributions in the workplace – Working Paper Harvard University, October.

Carneiro, F. (1997) – The changing informal labour market in Brazil: cyclicality versus excessive intervention – *Labour* 11(1) pp.3-22.

Case, A. C. and Katz, L. F.(1991) – The company you keep: the effects of family and neighborhood on disadvantaged youths – *NBER,* may.

Chen, M. A., Jhabvala, R. and Lund, F. (2002) – Supporting Workers in the Informal Economy: A Policy Framework – *Working Paper on Informal Economy ILO*

Corcoran, M., Datcher, L. and Duncan, G. (1980) – Information and Influence networks in labor markets – in Duncan, G and Morgan, J. *Five thousand american families: patterns of economic progress* Ann Arbor, MI Institute for Social Research.

Crane, J. (1991) – The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing – *AJS* 96 (5) march.

Dietz, R. D. (2002) – The Estimation of Neighborhood Effects in the Social Sciences: An Interdisciplinary Approach – *Social Science Research* 31 (4) pp. 539-575.

Dart, S. T. et al (2002) – Reforma Previdenciária: Em Busca de Incentivos Para Atrair o Trabalhador Autônomo – *EPGE Working Paper 460.*

Devaney, S. A. and Chien, YW. (2000) – Participation in Retirement Plans: A Comparison of the Self-Employed and Wage and Salary Workers – *Compensation and Working Conditions* Winter pp. 31-36.

Doeringer, P. B. and Piori, M. J. (1971) – International labor markets and manpower analysis – Lexinton, Mass Health Lexington Books.

Fields, G. S. (1975) – Rural-urban migration, urban unemployment and underemployment, and job-search activity in LDCs – *Journal of Development Economics* 2 (2).

Fiess et al. (2002) – Labor Markets and Economic Adjustment: the Brazilian Case – in *Brazil Jobs Report* World Bank report 24408-br Chapter 2.

Fankhouser, E. (1996) – The urban informal sector in Central America: household survey evidence – *World Development* 24 (11) pp. 1737-1751.

Fankhouser, E. (1997a) – Mobility and labor market segmentation: the urban labor market in El Salvador – *Economic Development and Cultural Change* pp. 123-153.

Fankhouser, E. (1997b) – Demand-side and supply-side explanations for barriers to labor market mobility in developing countries: the case of Guatemala - *Economic Development and Cultural Change* pp. 341-366.

Gallaway, J.H. and Bernasek, A. (2002) – Gender and informal sector employment in Indonesia – *Journal of Economic Issues* June (2).

Gindling, T. H. (1991) – Labor market segmentation and the determination of wages in the public, private-formal, and informal sectors in San Jose, Costa Rica – *Economic Development and Cultural Change* pp. 585-605.

Gong, X. and Van Soest, A. (2002) – Wage differentials and mobility in the urban labour market: a panel data analysis for Mexico (2002) – *Labour Economics* 9 pp.513-529.

_________(2000) – Family structure and female labor supply in Mexico City – *The Journal of Human Resources* (1) pp. 163-191

Granovetter, M.S. (1995) – Getting a job: a study of contacts and career - Chicago: University Press of Chicado.

Grodner, A. (2002) – An empirical model of labor supply with social interactions – Mimeo Syracuse University (November).

Hart, K. (1973) – Informal Income Opportunities and Urban Employment in Ghana – *Journal of Modern African Studies* 11 (1) pp. 61-89.

Hill, M.A. (1988) – Female labor supply in Japan. Implications of the informal sector for labor force participation and hours of work – *Journal of Human Resources* (1) pp. 143-161.

International Labour Organization (1972) – Employment, Incomes, and Equality: a strategy for increasing productive employment in Kenya – *ILO,* Geneva.

Lewis, W. A. (1954) – Economic Development with Unlimited Supplies of Labour – *Manchester School,* 22 (May).

Loayaza, N. V. (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).

Long, J. S. (1997) – Regression models for categorical and limited dependent variables - SAGE

Magnac, Th. (1991) – Segmented or competitive labor market? – *Econometrica* 59 (1) pp. 165-187.

Malony, W. F. (1999) – Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico – *The World Bank Economic Review* 13 (2) pp. 275-302.

Manski, C. (1993) – Identification of Endogenous Social Effects: The Reflection Problem – *Review of Economic Studies,* 60 pp. 531-542.

Marcouiller, D. Castilla, V. R. and Woodruff, C. (1997) – *Economic Development and Cultural Change* – pp. 367-392.

McFadden, D. (1984) – *Econometric Analysis of Qualitative Choices* – in. Handbook of Econometrics, Amsterdam, North Holland vol. 2.

Meng, Xin (2001) – The informal sector and rural-urban migration – a Chinese case study – *Asian Economic Journal* 15 (1).

Neri, M. C. (2002) – Decent work and the informal sector in Brazil – *EPGE* Ensaios Econômicos 461.

Portes, A. Castells, M and Benton, L.A. (1989) – The Informal Economy Studies in Advanced and Developed Countries – Ed. John Hopkins Baltimore.

Portes, A. and Schauffler (1993) – Competing Perspectives on the Latin American Informal Sector – *Population and Development* 19 (1).

Powell, L. M.; Tauras, J. A. and Ross, H. (2002) – Peer effects, tobacco control policies, and youth smoking behavior – Mimeo – Queen's University. November.

Pradhan, M. and Van Soest, A. (1995) – Formal and informal sector employment in urban areas of Bolivia – *Labour Economics* 2 pp. 275-297

________(1997) – Household labor supply in urban areas of Bolivia – *The Review of Economics and Statistics* pp. 300-310

PREALC (1998) – Panorama laboral.- Lima, Peru: International Labour Organization.

Scheiner, F. and Enste, D. H. (2000) – Shadow economies: size, causes and consequences – *Journal of Economic Literature* 38(1).

Shaw, A. and Pandit, K. (2001) – The Geography of Segmentation of Informal Labor Markets: The Case of Motor Vehicle Repair in Calcutta – *Economic Geography* pp. 180-196.

Small, K. A. and Hsiao, C. (1985) – Multinomial Logit Specification Tests – *International Economic Review* – (26) pp. 619-627.

Soares, F.V. (2004) – Do informal workers queue for formal jobs in Brazil? – *IPEA* Discussion Paper 1021.

Tannnuri-Pianto, M. and Pianto, D. M. (2002) – Informal Employment in Brazil – A Choice at the Top and Segmentation at the Bottom: A Quantile Regression Approach – Working Paper University of Brasilia.

Tannen, M. B. (1991) – Labor Markets in Northeast Brazil: Does the Dual Market Model Apply? – *Economic Development and Cultural Change* pp.567-583

Telles, E. E.(1992) – Who Gets Formal Sector Jobs? – *Work and Occupations* 19 (2) pp. 108-127

Telles, E. E. (1993) – Urban Labor Market Segmentation and Income in Brazil – *Economic Development and Cultural Change* pp. 231-249.

Topa, G. (2001) – Social Interactions, Local Spillovers and Unemployment – *Review of Economic Studies* 68 pp. 261-295.

Ulyssea, G. (2005) – Informalidade no Mercado de Trabalho Brasileiro: Uma Resenha da Literatura – *Texto para Discussão do Ipea n. 1070.*

CONCLUSION

Two important indicators of development for less developed countries are the infant mortality rate and the level of informality in the labor market. During the 1990s these two indicators presented different trends in Ceara State, Brazil. Whereas infant mortality rate (IMR) presented a downward trend during this period, the share of informal workers in the labor market increased. This thesis contributed to explain these results by investigating particular aspects of these two indicators.

Both essays of this thesis improved upon traditional modeling of infant mortality rate and informality decisions by including spatial considerations into the analyses. Essay I included spillover effects of health policies in reduced form models of infant mortality for municipalities, and essay II included neighborhood effects in worker's allocation models for urban areas.

Spillover effects occur when improving a health care policy in one municipality affects not only the IMR of this municipality but also the IMR of neighboring municipalities. To capture this spillover effect of policies essay I suggested the adaptation of spatial analysis to traditional reduced form models of infant mortality rate (Corman and Grossman, 1985). By comparing models with and without spillover effects for a policy program known as Community Health Worker Program (CHWP) essay I evidenced that spillover effects were significant, and their exclusion in traditional models would compromise the full effectiveness of this program. Under certain conditions and based on simulations, for example, essay I showed a difference of up to 39% in the effectiveness of the program when spillover effects were included in traditional reduced form models. This result has very important policy implications. If the benefits of the program are underestimated in traditional analyses (models without spillovers), it is possible that governments will under-invest in improving this program. The possible side effects of under-investment in this case are fewer infant survivors, which is a tremendous burden borne by the suffering families and by the society as a whole.

The second essay investigated two hypotheses usually neglected by labor market segmentation theories: i) formal versus informal dichotomies may not be adequate to classify workers in the labor market, and ii) workers' positions in the labor market may also be influenced by their interactions in urban neighborhoods. The empirical analysis from an expanded sector allocation model provided evidence in favor of both hypotheses for the urban labor market of Fortaleza city, Brazil. Thus, as segmentation models are the reference guide for diagnostic inefficiencies in the labor market, it is important to take into consideration in future studies these two pieces of evidence in order to avoid misspecification problems.

The interaction between individuals and their social context has been promulgated in many study cases such as smoking, negligent behavior of youths, public insurance take-up, etc. Social interaction among workers is mostly understood as informational channels (networking) which improve the chances of unemployed individuals finding a job (Topa, 2001). Although interactions may occur for formal or informal workers, this essay found stronger interaction effects among informal workers (small-firm uncovered workers), which is consistent with Granovetter (1995) and may also have contributed to explaining the rise of the informal sector during the 1990s. Although the identification of interaction effects in the form of networking is difficult to do when the interacting agents (workers) are not randomly formed, essay II provided a modeling strategy to identify general neighborhood effects (net effect of interactions, sorting and local fixed effects) which indicated possibility of spatial segmentation between formal (firm workers covered by social security) and informal (small firm workers uncovered by social security) workers. That is, at the neighborhood levels, the formal and informal workers tend to cluster themselves where the other type is less prevalent. The effect of this spatial segmentation on the possibility of labor market segmentation is left for future studies involving natural (social) experiments and/or panel data.