

**FUNDAÇÃO GETULIO VARGAS
ESCOLA DE PÓS-GRADUAÇÃO EM
ECONOMIA**

Letícia Faria de Carvalho Nunes

**Practice Location of Physicians: A Discrete Choice
Model Approach**

Rio de Janeiro

07 de Abril de 2015

Letícia Faria de Carvalho Nunes

**Practice Location of Physicians: A Discrete Choice
Model Approach**

Dissertação submetida a Escola de Pós-Graduação em Economia como requisito parcial para a obtenção do grau de Mestre em Economia.

Área de Concentração: Microeconomia Empírica

Orientador: Francisco Costa

Rio de Janeiro

07 de Abril de 2015

Nunes, Leticia Faria de Carvalho

Practice location of physicians: a discrete choice model approach / Leticia Faria de Carvalho Nunes. – 2015.

31 f.

Dissertação (mestrado) - Fundação Getulio Vargas, Escola de Pós-Graduação em Economia.

Orientador: Francisco Costa.

Inclui bibliografia.

1. Modelos econométricos. 2. Médicos. 3. Processo decisório. 4. Oferta e procura. 5. Probabilidades. I. Costa, Francisco Junqueira Moreira da. II. Fundação Getulio Vargas. Escola de Pós- Graduação em Economia. III. Título.

CDD – 330.015195



LETÍCIA FARIA DE CARVALHO NUNES

**PRACTICE LOCATION OF PHYSICIANS: A DISCRETE CHOICE MODEL
APPROACH**

Dissertação apresentada ao Curso de Mestrado em Economia da Escola de Pós-Graduação em Economia para obtenção do grau de Mestre em Economia.

Data da defesa: 07/04/2015

ASSINATURA DOS MEMBROS DA BANCA EXAMINADORA

A handwritten signature in blue ink, appearing to read 'F. J. M. da Costa', is positioned above a horizontal line.

Francisco Junqueira Moreira da Costa
Orientador (a)

A handwritten signature in blue ink, appearing to read 'L. H. B. Braidó', is positioned above a horizontal line.

Luis Henrique Bertolino Braidó

A handwritten signature in blue ink, appearing to read 'G. Ulyssea', is positioned above a horizontal line.

Gabriel Ulyssea

Abstract

Economists and policymakers have long been concerned with increasing the supply of health professionals in rural and remote areas. This work seeks to understand which factors influence physicians' choice of practice location right after completing residency. Differently from previous papers we analyse the Brazilian missallocation and assess the particularities of developing countries. We use a discrete choice model approach with a multinomial logit specification. Two rich databases are employed containing the location and wage of formally employed physicians as well as details from their post-graduation. Our main findings are that amenities matter, physicians have a strong tendency to remain in the region they completed residency and salaries are significant in the choice of urban, but not rural, communities. We conjecture this is due to attachments built during training and infrastructure concerns.

KEYWORDS: *Physician labor, underserved areas, discrete choice model, multinomial logit.*

Contents

1	Introduction	8
2	Background and Data	13
2.1	Brazilian Health System	13
2.2	Data and Summary Statistics	15
3	Model and Empirical Strategy	20
3.1	Model of Practice Location	20
3.2	Multinomial Logit and Identification	21
4	Results	23
5	Conclusion	29
6	References	31

List of Figures

1	Physicians per 1000 people: State's Capital x State's Countryside	9
2	Number of Residency Programs per Region	17
3	Medical Residency Region x Practice location Choice	17

List of Tables

1	Physicians Area of Residency Specialization	16
2	Summary Statistics, Amenities, Brazil 2002-2010	19
3	Multinomial Logit Results - Alternative Constants and Individual Attributes	25
4	Multinomial Logit Results - Alternative-specific Variables	27
5	Multinomial Logit Results - Residency Region Dummies	28
6	Actual and Predicted Frequencies	29

1 Introduction

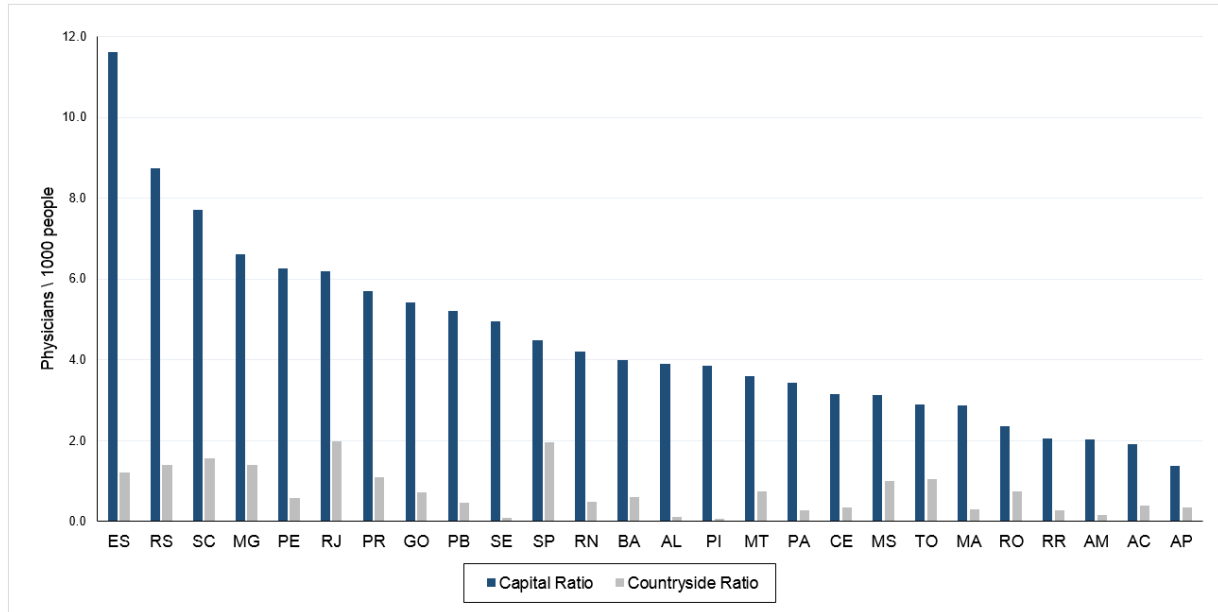
In many developed countries, despite the substantial growth in the total number of physicians, their maldistribution across regions persists. Rural areas and outskirts of big metropolis still suffer from the shortage of these professionals, entailing unequal access to health care. For this reason, physician supply in medically underserved regions became a main concern for policymakers and economists.

Some governments have implemented policies to address these distributional imbalances. Their aim is to act upon the physicians' choice of practice location. For instance, in 1970, the United States created the National Health Services Corps (NHSC), a program that tries to encourage newly trained physicians to work in medically underserved communities. It relies in loan forgiveness options and scholarships programs with a return-in-service provision. For every year that the student receives a scholarship, he has a contractual obligation to serve one year at a location defined by the Corps. The program has also the goal to encourage professionals to remain in the placement location after the funding is complete.

In Canada, since 1982, the provincial government of Québec has developed a whole set of incentive measures to influence its geographical distribution of physicians. The country health system is based on the Universal health Insurance Plan and has uniform medicare fees. This tend to intensify the inefficient spacial allocation of health professionals, since prices cannot adjust to equilibrate demand and supply. The measures implemented try to bypass this restriction through scholarships programs, settlement grants, relocation reimbursement, continuing education opportunities and a program of regionally differentiated remuneration for recently graduated physicians.

This is an even bigger challenge for developing countries, where rural and remote communities usually lack the basic health infrastructure and do not offer a reasonable quality of life for it's citizens. Physicians claim it is difficult and many times unsafe to help diseased people in such areas, because essential equipments and medications are in short of supply. This makes these regions much more undesirable, discouraging even professionals who have preferences for working in needy places. As can be seen in Figure 1, Brazil is a developing country with a sizeable misallocation of physicians between States' capital and countryside.

In the face of this impressive supply gap between rural and urban regions, Brazil



Source: Federal Council of Medicine (CFM), Brazilian Institute of Geography and Statistics (IBGE); Year 2012

Figure 1: Physicians per 1000 people: State's Capital x State's Countryside

has recently emerged as a developing country trying to address the maldistribution of physicians in its large geographical area. In 2011, the government created the Primary Care Professional Valorization Program (Provab). It intends to attract newly trained physicians to needy regions through two incentives: a sizeable remuneration and a 10% increase in the scoring of medical residency exams. In 2013, the country also implemented the More Physicians program. It's main goal is to create ten thousand primary care jobs in underserved communities for Brazilian and foreign physicians. The program is also increasing the number of medical schools and medical residency programs in areas suffering from shortages, such as the North and North-east of the country.

The analysis of such incentive policies requires a clear understanding of the factors that influence physicians practice location decisions. There is an extensive literature covering this issue, but most articles use a simple methodology, without an structural model to guide its conclusions. Moreover, and as will be latter discussed, the ones that use a discrete choice model approach could improve their econometric estimates through a mixed logit specification. Finally, these articles focus mainly in the supply of physicians in North America, neglecting the positive effects this analysis might bring to health care in developing countries.

Our project is to analyse the practice location decision of physicians in Brazil that recently completed their residency through a utility-theoretic framework with a mixed

logit specification. We intend to understand which factors are crucial for their choices, considering the rich diversity in social and regional characteristics that a developing country such as Brazil has.

The contribution to the literature will come in two fronts. First, we will estimate physician's preference parameters using a random coefficients multinomial logit, which is a much more flexible model and can approximate any random utility model (McFadden and Train, 2002). It, unlike standard logit, allows for random taste variation and unrestricted substitution patterns. Mixed logit also exceeds probit since it is not restricted to normal distributions. Therefore our approach will provide better estimates for the physicians preferences. Second, it will expand the study of physician's practice location choices to developing countries by using data on Brazilian professionals who recently completed their residency. This way we will analyse more thoroughly the importance of a good healthcare infrastructure and of amenities in their choice, since these are more preeminent issues in these countries.

The present dissertation shows the first steps taken to achieve the goals described above. We use comprehensive data of all physicians that completed medical residency between 2002 and 2010 linked to the official records of formal employees from the Ministry of Labor and Employment in Brazil. With it, we estimate a standard multinomial logit model of physician's demand for practice location right after the conclusion of their post-graduation. The results obtained, which we consider preliminary, suggest there is an empirical evidence that rural residency training enhances retention of rural physicians. Also, professionals finishing obstetrics and gynecology, pediatrics and family medicine residencies seem to have an increased probability of choosing to work in rural areas. Amenities such as region's education and health appear to be as important in their decision. An intriguing result is that wages are significant only for metropolitan regions, but not for the countryside. Finally, the living costs also seems to affects negatively the decision to work in a certain area.

The more recent articles in conformity with the methodology used in this work are Hurley(1991), Bolduc et al. (1996), Homes(2004) and Agarwal (2014). Hurley (1991) is the first paper to model the physician supply using a utility maximization framework. In its study, physician's choice concerning the specialty, community size and mode of their first practices in the USA are analysed in a simultaneous decision model with a

nested logit specification. The utility of a specific alternative depends on, among other things, the future expected discounted gains associated with this alternative. It is through changes in this variable that a series of income-based policies simulations are performed. The author finds that income is statistically significant and plays an important role in physicians' decisions.

Although Hurley's paper has a great merit for introducing the approach of discrete choice models in the study of physician's decision of practice location, it has some problems that might bias the estimation. The main one concerns to the utilization of a multinomial nested logit, in which, after choosing his specialty and practice mode, the physician chooses his work location. As pointed out by Bolduc et al. (1996), by implementing this method, the author assumes there is no correlation among the various location alternatives. This is known as independence of irrelevant alternatives (IIA) and, since it is not very reasonable in this context, it might produce biased estimates through a misspecification of the error correlation structure.

Bolduc et al. (1996) tries to overcome this and other problems in Hurley (1991). Their article estimates a model in which a general practitioner chooses his first practice location in Québec, Canada. The model is used to assess the effect of various incentive measures introduced in the province in order to influence the geographical distribution of these professionals. The estimation strategy used is a spatially autoregressive multinomial probit. This allows for interdependencies among specific errors of alternative location choices to be modeled using a first-order spatial autoregressive (SAR[1]) process based on a distance-decaying relationship (for more details see Bolduc and Ben-Akiva, 1991). The authors conclude that this specification is superior to a usual multinomial logit, a nested multinomial logit and an independent multinomial probit. Their results provide evidence that the measures implemented by Québec government influence the physicians' choice of location.

Homes (2004) employs a rich database, containing the location of physicians in the USA at 5-year intervals, in order to compare the types of regions chosen by enrollees and nonenrollees in the NHSC program. He uses a sequential multinomial logit with discrete unobserved heterogeneity to account for endogeneity of enrollment in the program. This specification allows non-observable factors that influence the decision to participate or not in the Corps to also affect the decision to work in needy areas. The unobserved

heterogeneity is assumed to be additive in the utility and its distribution is approximated using a discrete factor model (Heckman and Singer, 1984). According to Homes, the advantage of using this approach instead of a multinomial probit is the better computational performance added to the fact that it doesn't place an a priori distributional assumptions on the unobserved factors, which could result in inconsistent parameter estimates.

However, an important aspect of this specification is that, conditional on the unobserved factor, independence of irrelevant alternatives still holds. In other words, individuals with a given value of the unobserved factor face independent shocks. This problem is well documented in the article, but it could be overcome with the use of a sequential mixed logit with discrete unobserved heterogeneity instead of a sequential multinomial logit. This way, the model would allow for more flexibility in the correlation between location alternatives and still account for endogeneity of enrollment in the program.

At last, it is important to discuss Agarwal (2014). When data on stated preferences are available, discrete choice methods provide a straightforward tool for analysis. However, researchers usually face the constraint of only having data on employer-employee final matches rather than stated preferences. Agarwal (2014) makes an important contribution to overcome this obstacle by developing a framework that estimates preferences in a many-to-one matching market using only observed matches. He uses pairwise stability together with a vertical preference restriction on one side of the market to estimate preferences parameters using the method of simulated moments.

With this procedure, in the NBER version of the paper, Agarwal estimates the preference distribution in the market for family medicine residents and analyses wage and supply policies aimed at increasing the number of residents training in rural areas. He shows that price incentives and quantity regulations are not equivalent policy instruments. Supply regulations are more effective at increasing the number of residents in rural areas while financial incentives combined with capacity constraints prevents an increase in the numbers but allow an increase in the quality of residents matched at subsidized programs.

At first sight, it seems we could use a similar approach to investigate the physician's labor market. However, we don't have data that could be related to the preferences of hospitals, ambulatories and clinics while hiring physicians. And, even if we had, Agarwal's method doesn't allow a complex model for both segments of the market. It could, as documented in the article, raise identification issues and other methodological

difficulties due to multiple equilibria. Because of it, Agarwal had to restrict attention to a model in which the program’s preferences for residents are homogeneous. Yet, modelling the demand side of the whole physician’s labor market requires much more flexibility. For these reasons, we follow the literature and model physician’s choices for practice regions, and not their preferences for jobs in health establishments, which would raise demand restrictions issues. In other words, hospital and clinics preferences would have to be taken into consideration to understand final matches. In our approach, the choices observed can be considered the favourite region of each physician.

This dissertation is organized as follows: in Section 2, we provide an overview of the brazilian health system, a description of the data used and summary statistics regarding it. We then present the theoretical model of physicians’ initial practice location and our empirical specification in Section 3. The estimation results are discussed in section 4. The final section concludes with an analysis of which steps should be taken next, together with some thoughts on the implications these preliminary results have for public policies in developing countries.

2 Background and Data

In this section we first describe the brazilian health system, explain how physicians degrees and resident specialist degrees are awarded and discuss the supply of these professionals over the country. Then, we characterize the data used and present it’s summary statistics.

2.1 Brazilian Health System

The Brazilian health system is based on the Federal Constitution of 1988, which defined healthcare as a citizen’s right and a state’s duty. In order to translate this basic right into practice, the Unified Health System, known as SUS, was created, which universally supplies preventive and curative care through a decentralized management. It is now the largest public health system in the world, providing complete coverage for 75% of the Brazilian population. The remaining 25% of the population, covered by the private sector, also has the right to access services provided by the SUS.

The constitution delineates the basic structure of SUS, prioritizing the preventive care and emphasizing the regionalization of health services by transferring responsibilities to

states and cities. It is believed that this decentralization makes local health systems more attuned to the health needs of the population and enables the exercise of public control over health policies. So, in the Brazilian health system, the Federal government is responsible for developing national public policies and financially supporting states and municipalities, which in turn are responsible for the execution of regulatory functions in their respective spheres with considerable autonomy.

The private sector also plays an important role in the Brazilian health system. The country has the second-largest private insurance market by population in the world, behind only the United States. A significant share of Brazilian health plans are either small or medium in size and operate via contractual arrangements, providing medical care in private doctors' offices and hospitals. As in the United States, most private health plans in the country are connected to employment. The private healthcare market is regulated by the National Health Agency (ANS), whose responsibility is to promote public interest in private health insurance and regulate the operators in the industry.

Now we turn to how the degrees of physician and resident specialist are awarded in Brazil. The Brazilian medical schools follow the European model of a six-year curriculum, divided into three cycles of two years each (basic cycle, clinic cycle and internship cycle). After six years of training, students graduate and are awarded the title of physician, allowing them to register with the Regional Council of Medicine (CRM). The recent graduate will be able to exercise the medical profession as a general practitioner and may apply to undertake postgraduate training.

Physicians who want to join a residency program must undergo a new selection examination considered as competitive as that required to join a medical school. These specialization programs are divided into two categories: direct access and prerequisite. The specialties with direct access are those in which the doctor can enroll without having any prior expertise. To apply to proprietary pre-requisite, the doctor should have already completed a specialty prior, usually clinical or surgery residency. The programs may need from 2 to 6 years of training. In Brazil, 53 residency programs are currently recognized by the Federal Council of Medicine (CFM), the Brazilian Medical Association (AMA) and the National Commission of Medical Residency (CNRM). The completion of these programs gives the title of resident physician specialist. Around 60% of medical school graduates does not have a residency specialization in Brazil.

Finally, it is important to discuss physician’s supply in Brazil. According to registers from the Federal Council of Medicine (CFM), in 2012, Brazil had 388.015 active physicians, reaching the mark of two professionals per 1000 inhabitants. This considerate escalation in the number of physicians, from 1.15 in 1980 to 2.0 in 2012, followed the inauguration of several medical schools from 1965 onwards. Now the country is considerably above the global ratio of 1.4 and near the 2.4 ratio of the United States, as stated by the 2014 World Health Statistics. Together with this increase came a change in the focus of public health policies in Brazil. Programs such as Protab and More Physicians began to prioritize the maldistribution of these professionals across regions instead of their shortage in general. As can be seen in Figure 1, Brazilian physicians are mainly concentrated in metropolitan regions, specially in South and Southeast of the country, instead of its outskirts and rural areas. In our study we investigate exactly what entails such practice location decisions by physicians in a developing country.

2.2 Data and Summary Statistics

Next we describe the two main data sets, RAIS and CNRM, together with other sources of information used in this study.

The primary data source is the Annual Report of Social Information (RAIS) from 2002 to 2010, maintained by the Ministry of Labor and Employment in Brazil. Currently, it covers 95% of the formal sector in the country. The observation units are formal establishments with detailed information about its employees. This database made it possible to see in which city physicians with a formal link are working, as well as their average wages, working hours, age, gender and race. Except for race, which had a lot of missing values, these information were all used in this research from 2002 until 2010. It is important to highlight that, although we know the cities physicians decided to work, for computational and analytical reasons we chose to aggregate their decision unit into metropolitan region and countryside of the five main Brazilian regions (South, North, Northeast, Southeast and Midwest), amounting to ten possible choices. If a physician worked in more than one region, we considered only the one with higher working hours.

The measure constructed for wages deserves a little more explanation. With information provided by RAIS, we calculated the average wage per hour physicians earn in each of the ten decision units. We used the salary per hour of the whole sample of physicians,

and not only of the ones with a residency specialization. So, in our model, while making their decisions, individuals look at the average hourly wage physicians formally employed earn in each region. It is important to state that this measure does not account for the full income of physicians. There are many off-book earnings associated with being self-employed, like owning their own office or clinic.

Afterwards, we linked these registers with data from the National Commission of Medical Residency (CNRM). This database provides information about the residency programs every physician in the country completed. Through it, we have access to the physician name, his social security number, the specialty chosen together with the beginning and conclusion date of training. Since many physicians completed more than one residency (the first one usually being clinical medicine or surgery, both prerequisite for other programs), we kept only the most recent one for analysis. Also, for analytical reasons, we decided to aggregate the specialties into five major fields: surgery, clinic, pediatrics, family medicine and obstetrics and gynaecology. Table 1 shows how physicians are distributed among these areas. Finally, while merging the two databases, we lost around 26,7% physicians from CNRM. We believe this loss is mainly due to mistakes in the social security numbers and to self-employed physicians. As stated previously, RAIS includes only formally employed workers.

Table 1: Physicians Area of Residency Specialization		
	Observations	Percent (%)
Surgery	8,116	24.36
Clinic	14,382	43.16
Pediatrics	4,318	12.96
Family Medicine	941	2.82
Obstetrics and Gynaecology	3,915	11.75
Others	1,651	4.95
Total	33,323	100

NOTES - Sources: RAIS, CNRM. This table displays how physicians, post-graduated between 2002 and 2010, are distributed among the five major areas of residency programs in Brazil.

From this joint dataset we can extract relevant information such as the high concentration of residency programs in the Southeast (Figure 2). Moreover, comparing the choices of practice location done by recently graduated physicians with the region they completed residency we see a strong positive relation (Figure 3). This indicates a tendency for physicians to remain in the same region they completed residency.

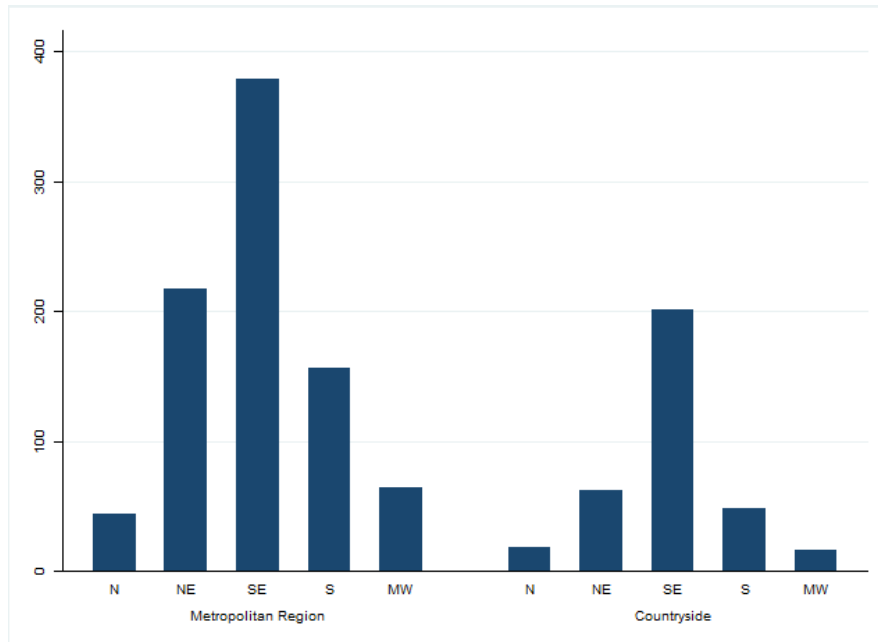


Figure 2: Number of Residency Programs per Region
Source: CNRM, Year: 2014.

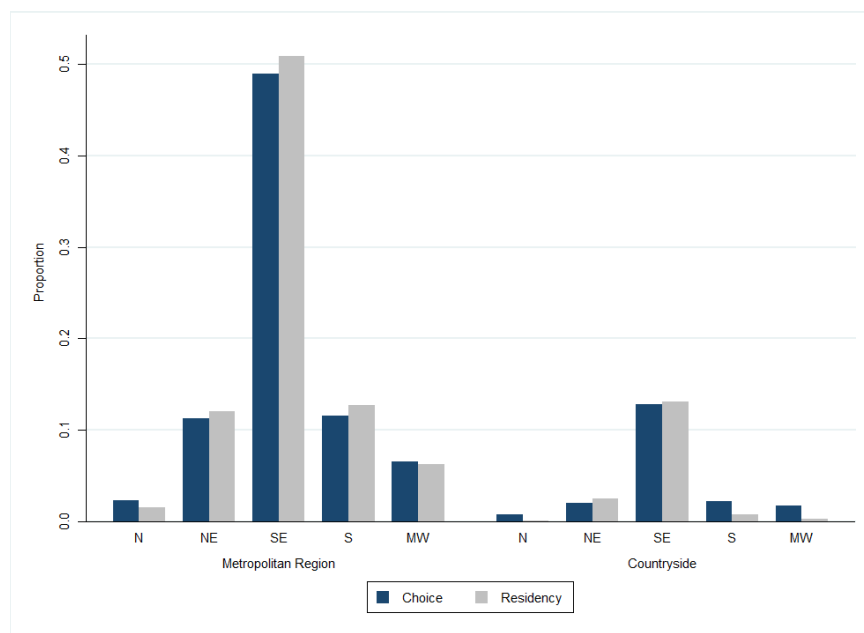


Figure 3: Medical Residency Region x Practice location Choice
Source: CNRM, RAIS, Years: 2002-2010

Measures of education were obtained from the National Treasury, which provides data of the expenditures made by each city in this sector. This variable was modified to expenditure per capita. We also used the rates of infant mortality as a proxy for the cities' health system. This data is provided by the Brazilian Institute of Geography and Statistics (IBGE). The regions' GDP per capita measure were also obtained from IBGE. As a violence proxy, we used homicide rates provided by the SUS Database (DATASUS). The proxy for living costs derived from data in RAIS and was constructed based on salary earned by bus drivers in each region for 44 weekly working hours. We used bus drivers, because they provide in every city a non-traded service. The summary statistics of these variables together with salary per hour are displayed in Table 2.

Our study will be latter improved with health infrastructure data for each region. This information is provided by the National Register of Health Establishments (CNES), whose purpose is to register every health establishment, private or public, in the country. The records have data on the installed capacity of health establishments since 2003 and are published monthly. Since it is very detailed, we intend to create seven summary indices following Katz, Kling and Liebman (2007) methodology.

Table 2: Summary Statistics, Amenities, Brazil 2002-2010

	Education expenditure [1]	Infant mortality rate [2]	Homicide rate [3]	GDP per capita [4]	Living Costs [5]	Salary per hour [6]
Southeast						
Metropolitan Region	343.44 (97.38)	0.209 (0.026)	0.317 (0.084)	12,038.78 (859.70)	1,339.32 (198.48)	158.34 (41.74)
Countryside	307.00 (88.77)	0.204 (0.029)	0.152 (0.004)	7,994.19 (671.64)	1053.11 (170.24)	152.60 (38.36)
South						
Metropolitan Region	292.75 (80.11)	0.174 (0.024)	0.245 (0.025)	10,185.45 (760.81)	1209.55 (175.96)	159.69 (38.55)
Countryside	281.50 (68.10)	0.205 (0.040)	0.185 (0.007)	7,350.85 (511.51)	1065.66 (149.75)	155.35 (41.00)
North						
Metropolitan Region	208.35 (67.91)	0.366 (0.052)	0.336 (0.082)	6,519.36 (458.78)	1162.94 (239.09)	128.95 (30.71)
Countryside	265.69 (84.65)	0.423 (0.045)	0.257 (0.035)	3,913.49 (444.92)	928.89 (163.84)	175.30 (45.42)
Northeast						
Metropolitan Region	216.80 (71.35)	0.298 (0.058)	0.410 (0.062)	5,294.39 (431.69)	1077.11 (133.76)	113.25 (33.68)
Countryside	276.32 (96.34)	0.347 (0.066)	0.203 (0.029)	2,562.92 (252.81)	848.19 (112.13)	117.29 (32.94)
Midwest						
Metropolitan Region	121.90 (29.16)	0.242 (0.032)	0.320 (0.018)	14,050.20 (1,046.33)	1042.19 (156.03)	168.21 (43.56)
Countryside	305.70 (90.64)	0.271 (0.038)	0.281 (0.013)	7,168.74 (645.61)	924.37 (141.55)	168.79 (48.15)
Brazil	261.95 (100.15)	0.274 (0.089)	0.270 (0.087)	7,707.74 (3,433.92)	1065.13 (217.36)	149.78 (44.95)

NOTES - Sources: IBGE, DATASUS, National Treasury, RAIS, CNRM; years: 2002-2010; standard deviations in parenthesis. This table displays the means from the region's amenities. Education expenditure per capita made by cities are from the National Treasury. Infant mortality rates by 1,000 population are from the Brazilian Institute of Geography and Statistics (IBGE). The regions' gdp per capita in constant 2000 reais were also obtained from IBGE. Homicides rate by 1,000 population are provided by the Computer Department of SUS (DATASUS). The proxy for living costs is based on the bus drivers' wage of 44 weekly working hours. This information was derived from data in RAIS. Physicians' salary per hour were also obtained from RAIS.

3 Model and Empirical Strategy

This section describes the physician’s decision model, the general setup of the econometric model and the identification strategy.

3.1 Model of Practice Location

We assume physicians, right after finishing their residency program, maximize utility over initial practice locations. The utility is a function of the physician’s socioeconomic and training characteristics as well as of the alternative’s attributes. The utility physician i receives from region j is:

$$U_{ij} = V_{ij}(w_j, x_j, a_i) + \epsilon_{ij} \quad (1)$$

where

V_{ij} is the representative utility,

w_j is the expected hourly salary of physicians formally employed in region j ,

x_j is the vector of amenities from region j ,

a_i is the vector of individual characteristics and

ϵ_{ij} is the stochastic term.

The vector of amenities includes the region’s gdp, living costs, health system, violence index and investment in education. We also added alternative-specific constants that accounts for unobserved heterogeneity in regions that remain stable over time. In the vector of individual characteristics we have age, gender, region where the physician completed the residency program and the speciality chosen. To capture different effects wages may have in rural and urban areas, we added an interaction between salary per hour and a dummy for metropolitan region. As for the random component of utility, ϵ_{ij} , it takes into account factors such as unobserved heterogeneity, unobserved attributes of regions that vary over time and measurement errors. The assumption placed in its distribution will determine the econometric model.

3.2 Multinomial Logit and Identification

Now we describe the multinomial logit model used to estimate the parameters of the utility function in (1) and the location-choice probabilities for each physician in the population. We assume the utility function is linear in parameters, where β is a vector containing them. This simplification is reasonable since, under fairly general conditions, any function can be approximated arbitrarily closely by one that is linear in parameters. We also suppose errors follow a type 1 extreme value distribution, which will give some desirable properties to the probabilities. They will necessarily be between zero and one and the choice probabilities for all alternatives will sum to one.

$$U_{ij} = X_{ij}\beta + \epsilon_{ij} \quad (2)$$

$$\epsilon_{ij} \stackrel{iid}{\sim} \text{extreme value} \quad (3)$$

The physician's choice is such that:

$$y_{ij} = \begin{cases} 1 & \text{if } U_{ij} \geq U_{ik}, \forall k \in \{1, \dots, J\} \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

where y_{ij} indicates the observed choice made by the physician. The probability that he chooses alternative j is given by:

$$P_{ij} = \text{Prob}(U_{ij} > U_{ik}, \forall k \neq j) \quad (5)$$

$$= \text{Prob}(V_{ij} + \epsilon_{ij} > V_{ik} + \epsilon_{ik}, \forall k \neq j) \quad (6)$$

$$= \text{Prob}(\epsilon_{ik} < \epsilon_{ij} + V_{ij} - V_{ik}, \forall k \neq j) \quad (7)$$

$$= \int \left(\prod_{k \neq j} e^{-e^{\epsilon_{ij} + V_{ij} - V_{ik}}} \right) e^{-\epsilon_{ij}} e^{-e^{-\epsilon_{ij}}} d\epsilon_{ij} \quad (8)$$

After some algebraic manipulation of the integral, we get a succinct, closed-form expression:

$$P_{ij} = \frac{e^{V_{ij}}}{\sum_k e^{V_{ik}}} = \frac{e^{X_{ij}\beta}}{\sum_k e^{X_{ik}\beta}}$$

which is the multinomial logit choice probability.

It is important to recall that only differences in utility matter for the specification and estimation of any discrete choice model. If a constant is added to the utility of all alternatives, the physician's decision will not change. Also, the choice probability depends only on the difference in utility, not its absolute level. Thus, only parameters that capture differences across regions can be identified. Since region's attributes vary over alternatives, their identification is straightforward.

However, the physicians' characteristic also influence their decision and they do not vary over the options. To account for this fact, we needed to normalize the social-demographic parameters of one alternative. The alternative chosen for comparison was the Southeast's metropolitan region and we set its social-demographic parameters to zero. The other parameters are interpreted as being relative to it. The same issue affects alternative-specific constants, included to capture the average effect on utility of all factors that are not included in the model. Therefore we normalized the absolute levels of these constants and chose the same region for comparison.

Since the logit probabilities take a closed form, the traditional maximum-likelihood methods can be used. Assuming that each physician decision is independent, the probability of each person in the sample choosing the alternative that he was observed actually to choose is

$$L(\beta) = \prod_{i=1}^I \prod_j (P_{ij})^{y_{ij}}$$

where $y_{ij} = 1$ if the physician i chose region j and zero otherwise. The log-likelihood function is then:

$$LL(\beta) = \sum_{i=1}^I \sum_j y_{ij} \ln P_{ij}$$

and the estimator is the value of β that maximizes this function. At the maximum of the loglikelihood function, its derivative with respect to each of the parameters is zero.

It is important to comment on the limitations of the multinomial logit specification. First, logit models can capture tastes that vary only with respect to observed variables. Neither random taste variation nor preferences that vary with unobserved variables are taken into account. Since the observable variables in our model do not capture the whole

spectrum of physician’s preferences, logit is not the ideal specification. Also, this model implies proportional substitution patterns across alternatives, a characteristic known as independence of irrelevant variables. This assumption is not reasonable in our context because it imposes zero correlation across regions. Physicians that choose a certain region are more likely to prefer alternatives similar to it than completely different ones. Thus, IIA is not a reasonable assumption in our context and might lead to inconsistent estimates.

To incorporate random taste variation appropriately and capture more flexible substitution patterns, the ideal specification is the multinomial mixed logit. This will be exactly our empirical approach and contribution to the literature in latter developments of this paper. But, before refining our method, we intend to see in which direction a discrete choice model approach is leading us with this simpler specification. This will give us a better understanding into how to take the next steps.

4 Results

In this section we discuss the results of our multinomial logit specification of equation 2, presented in Tables 3, 4 and 5. The results and observations made in this section are considered preliminary given our intention to later improve the model and data presented here.

First, we will examine how personal characteristics affect physicians decisions of practice location. Table 3 provides the coefficients and standard deviation for region-specific constants and individual attributes. Since these coefficients do not vary over alternatives, we normalized to zero the parameters corresponding to the Southeast metropolitan region. The other parameters should be interpreted relative to it. For example, being a male physician raises de probability of choosing the Southeast rural area instead of its metropolitan region, since its coefficient is positive and significant at a 1% level (0.334). We chose the Southeast metropolitan region for comparison, because it is the most developed one in Brazil and most physicians are currently concentrated in it. Also, São Paulo and Rio de Janeiro States belong to the this region.

Analysing the gender parameter we see that they are mainly positive and significant (at 1% and 5% level) for countryside regions, which means that being a male general

practitioner increases the probability of practising in these areas compared to the Southeast metropolitan region. Moreover, the large coefficient associated with the Northeast countryside, which is one of the poorest regions in the country, indicates that this effect is stronger for impoverished areas. Thus, female physicians are less likely to work outside large urban center than male physicians. Concerning the effect of age, all the coefficients are positive and highly significant, which means that being older raise the probability of working outside the Southeast metropolitan region. This effect is stronger with the countryside regions, indicating that older physicians are more likely to choose a rural or smaller community. Regional dummies have been used to control for the effect of regional amenities varying across regions but remaining stable over time. These constants are all negative and highly significant.

As for the residency areas physicians completed, we can see that their coefficients follow an interesting, though not perfect, pattern. Physicians that completed pediatrics, family medicine or gynaecology and obstetrics programs have a higher probability of working in countryside regions than physicians that completed surgery programs. These coefficients tend to be positive and highly significant. Moreover, having completed residency programs in one of these three areas reduces the probability of working in urban centers other than the Southeast metropolitan region. These results are consistent with the fact that these specialties are related to primary care, which is more needed in remote and rural areas. It doesn't need as much a good infrastructure as surgery practices. Also, physicians that chose these programs have probably a higher preference for working in needy places.

We now analyse the results for alternative-specific variables displayed in Table 4. They exploit the effects of regional amenities in physician's choice. The coefficient for education expenditure of municipalities is positive and significant at a 10% level, signalling the importance of regional schooling in their choice. However, this parameter is small probably because part of its effect is being captured by the region-specific constants. The quality of the education system in Brazil has remained relatively stable between 2002-2010. As for the infant mortality rate, it's coefficient is highly negative and significant at a 1% level. This indicates that economic and social conditions for the health of mothers and newborns, as well as the effectiveness of health systems are taken into consideration by physicians. Also, our living costs proxy is negative and significant at

Table 3: Multinomial Logit Results - Alternative Constants and Individual Attributes

	Constant [1]	Gender [2]	Age [3]	Residency Areas				
				Clinic [4]	Pediatrics [5]	Family [6]	GO [7]	Others [8]
Metropolitan S	-2.610*** (0.367)	0.069 (0.064)	0.024** (0.011)	0.140* (0.079)	-0.033 (0.108)	0.036 (0.155)	-0.197* (0.109)	0.561*** (0.150)
Metropolitan N	-5.666*** (0.600)	0.068 (0.109)	0.139*** (0.013)	0.202 (0.131)	-0.351* (0.200)	-0.849** (0.418)	-0.138 (0.191)	0.156 (0.258)
Metropolitan NE	-2.784 (0.526)	-0.049 (0.061)	0.072*** (0.009)	-0.014 (0.073)	-0.485*** (0.105)	-0.033 (0.190)	-0.530*** (0.104)	0.589*** (0.134)
Metropolitan MW	-2.538*** (0.394)	0.191*** (0.073)	0.023* (0.012)	-0.102 (0.088)	-0.410*** (0.127)	-0.466 (0.293)	-0.375*** (0.123)	0.491*** (0.158)
Rural SE	-0.927** (0.366)	0.334 *** (0.047)	0.019** (0.009)	0.156*** (0.058)	0.212*** (0.080)	0.622*** (0.137)	0.271*** (0.080)	0.199* (0.115)
Rural S	-3.712*** (0.472)	0.479 *** (0.086)	0.071*** (0.012)	0.133 (0.111)	0.427*** (0.144)	1.388*** (0.185)	0.669*** (0.135)	0.042 (0.233)
Rural N	-6.429*** (0.743)	0.370*** (0.141)	0.156*** (0.013)	0.106 (0.177)	0.465** (0.227)	0.741** (0.333)	0.406* (0.227)	-0.176 (0.406)
Rural NE	-5.121** (0.715)	0.641 *** (0.089)	0.129*** (0.011)	0.204* (0.112)	0.345** (0.155)	0.000 (0.317)	1.120*** (0.128)	0.228 (0.230)
Rural MW	-2.503*** (0.581)	0.239** (0.094)	0.041*** (0.015)	-0.078 (0.112)	-0.100 (0.160)	-0.907** (0.427)	0.426*** (0.140)	-0.625** (0.305)

NOTES - Source: CNRM, RAIS; Years: 2002-2010; alternative-specific dummies were also included; Southeast metropolitan region is the region of reference; N = 33,323; Log likelihood:-30132.99; standard deviations in parenthesis; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table displays the multinomial logit estimates for region-specific constants and individual attributes. Since these coefficients do not vary over alternatives, we normalized to zero the parameters corresponding to the Southeast Metropolitan Region. The other parameters should be interpreted relative to it. Also, the dummies for residency area should be analysed comparative to surgery. As for the variable gender, one corresponds to male physician and zero to female physician.

10% level suggesting that this factor is taken into consideration by physicians. The reason for the small estimate might also be the inclusion of alternative-specific constants.

The GDP per capita is the core indicator of the regions' standard of living and economic performance. In our results this variable is positive but not significant. This might be due to high correlation with other variables, such as education, infant mortality and the regional-specific dummies. On the other hand, the homicides rate is significant at a 5% level and positive, which is counterintuitive. This probably indicates an endogeneity problem caused by omitted variables, which we intend to correct by introducing new amenities inputs, such as region's health infrastructure and entertainment quality.

As for the salary per hour variable and its interaction with a metropolitan region dummy, we obtain an intriguing result. Both have a positive, though small, coefficient. However, the former is not significant and the later is at a %1 level. This indicates that

salaries in metropolitan regions are important for physicians that recently completed residency, but not salaries in the countryside. We first thought this might be due to salaries in rural and remote areas coming mainly from public health establishments and being therefore more regulated and with a smaller variance. Yet, their variance is mostly greater than salaries in metropolitan regions, as can be seen in the summary statistics of Table 2. Another reasonable explanation could be that this result, together with the low coefficient values, is related to the usual endogeneity problem of supply and demand models. We will address this point in the next step of this work through an instrumental variable.

However, it could be that wages are not a key determinant in the decision of physicians who decide to work in rural areas. Other factors, such as health infrastructure and life quality, might be much more crucial in their decision that obfuscates the influence salaries usually have. Unlike developed countries, where most cities in the outskirts of large urban regions have the essential infrastructure to attend well its citizens, Brazil's remote areas are usually very poor and lack basic infrastructure and a good education and health system. This is the argument used by many physicians in the country, who claim that better salaries are not as crucial as better work and life conditions in these regions. Hence, our results, though preliminary, could be pointing in this direction. If this is so, public policies that offer better wages for physicians in remote areas are not very effective in developing countries. Our next step is exactly to incorporate better measures of infrastructure from CNES to directly assess these dimensions.

Finally, let's examine the dummies for residency region. Although they relate to individual attributes, we constructed these variables in a manner similar to the dummy that indicates which region a physician chose. This way they vary through alternatives and their coefficients can be identified without a normalization. That is why we included their results in Table 5, a separate one. We built a dummy for each region in order to capture if completing residency in different regions retain physicians differently.

The parameters associated with these dummies are all positive and highly significant (1% level) providing evidence that recently post-graduated physicians have strong incentives to remain in the same region they completed residency. This might be due to attachments developed during the training period and to connections made that facilitate the process of finding a job. This result is of high value for policy-makers, who

Table 4: Multinomial Logit Results - Alternative-specific Variables

	MNL Results
Education Expenditure	0.001* (0.000)
Homicides Rate	0.514** (0.249)
Infant Mortality Rate	-3.012*** (0.969)
GDP per capita	0.068 (0.051)
Salary per hour	0.002 (0.002)
Salary per hour x Metropolitan Region Dummy	0.003*** (0.001)
Living Costs	-0.001* (0.000)

NOTES - Sources: IBGE, DATASUS, National Treasury, RAIS, CNRM; Years: 2002-2010; alternative-specific dummies were also included; N = 33,323; Log likelihood:-30132.99; standard deviations in parenthesis; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table displays the multinomial logit estimates for region-specific variables. Standard deviations are in parenthesis. Education expenditure per capita made by cities are from the National Treasury. Infant mortality rates by 1,000 population are from the Brazilian Institute of Geography and Statistics (IBGE). The regions' gdp per capita in constant 2000 reais were also obtained from IBGE. Homicides rate by 1,000 population are provided by the Computer Department of SUS (DATASUS). The proxy for living costs is based on the bus drivers' wage of 44 weekly working hours. This information was derived from data in RAIS. Physicians' salary per hour were also obtained from RAIS.

could encourage the opening of more residency programs in the countryside to retain this workforce. Also, we can see that post-graduate programs in metropolitan regions usually have a larger power in retaining physicians than rural communities.

It is interesting to point the high value associated with the North rural and metropolitan regions. A possible explanation is the opening of many new residency programs in this region between 2006 and 2010. The North region is traditionally the one with least residency programs in the country. Roraima, for example, opened its first post-graduation in 2006 and Tocantins had its first residency class graduating in 2013. The North also has 1.1 physicians per 1000 people, the lowest ratio in the country. The combined effect of new residency programs in a region with few of them and of a large physician shortage might explain the strengthened effect the North dummies presented (5.255 and 5.902 with 1% significance).

Table 5: Multinomial Logit Results - Residency Region Dummies

	MNL Results		MNL Results	
Metropolitan SE	2.214*** (0.037)	Rural SE	2.000*** (0.044)	
Metropolitan S	4.084*** (0.060)	Rural S	2.630*** (0.138)	
Metropolitan N	5.255*** (0.147)	Rural N	5.902*** (0.748)	
Metropolitan NE	3.694*** (0.057)	Rural NE	1.358*** (0.102)	
Metropolitan MW	3.986*** (0.070)	Rural MW	1.288*** (0.276)	

NOTES - Sources: RAIS, CNRM; Years: 2002-2010; alternative-specific dummies were also included; N = 33,323; Log likelihood:-30132.99; standard deviations in parenthesis; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table displays the dummies for residency region where physician completed its post-graduation study. The way they were constructed created variation between alternatives, that is why we included this individual characteristic in this table. These dummies were created with information from the National Commission of Medical Residency (CNRM).

The model's predictive power can be assessed in Table 5, which compares the predictions with the observed outcome. The predictions are obtained by assuming each physician chooses the region that has the highest predicted probability. Overall, the fit is relatively good, with a proportion of total observations correctly predicted of 78.70%. The model slightly overpredicts the choices of Southeast, South and Northeast metropolitan regions and of the Southeast countryside, all high-frequency alternatives. Nonetheless, their relative high proportion of correctly predicted choices (column 4) are only marginally due to overprediction. As column 5 reveals for these four regions, 83.3%, 76.52%, 78.79% and 63.77% of the predictions are actually correct. The North and Midwest metropolitan regions also presents a good degree of agreement between the predicted and actually chosen regions. However, the South, North and Northeast countrysides correctly predicted rates are considerably lower and the choice of Midwest rural region by physicians couldn't be predicted at all by our model. The contrast in the predictive power of our model between rural and metropolitan regions indicates there are factors influencing physicians decision to work in remote areas that are not being captured. We believe this will be mainly improved when we incorporate health infrastructure indexes in our model.

Table 6: Actual and Predicted Frequencies

Region	Actual [1]	Predicted [2]	Correctly Predicted [3]	% Correctly Predicted (3/1) [4]	% Predictions Correct (3/2) [5]
Metropolitan SE	16,317	17,722	14,771	90.52	83.35
Metropolitan S	3,850	4,220	3,325	86.36	78.79
Metropolitan N	773	506	440	56.92	86.96
Metropolitan NE	3,742	4,000	3,061	81.80	76.52
Metropolitan MW	2,161	2087	1,606	74.32	76.95
Rural SE	4,254	4,367	2,785	65.47	63.77
Rural S	741	248	127	17.14	51.21
Rural N	257	24	21	8.17	87.50
Rural NE	675	144	56	8.30	38.89
Rural MW	553	5	0	0.00	0.00
1 Total	33,323	33,323	26,192	78.70	78.70

NOTES - Multinomial Logit Model; Log likelihood:-30132.99; alternative-specific dummies were included; N = 33,323; This table provides information on the predictive frequency distribution of outcomes based on predicted probability. The predictions are obtained by assuming each physician chooses the region that has the highest predicted probability. The multinomial logit imposes that the sample average predicted probabilities are equal to the actual sample average probabilities.

5 Conclusion

In this dissertation, we presented a discrete choice model based on a multinomial logit specification that assess the physician's first choice of practice location in developing countries. Our aim is to better understand which factors are central in their decision process in order to guide public policies used to influence their geographical distribution.

One of our main findings was how the place where physicians' finished their residency program seems to deeply influences their decision of practice location. The variables presented a high positive value with a 1% significancy level indicating that this is highly taken into consideration by them. This is probably related to attachments built during the training period and to professionals connections, essential in the beginning of their careers. Also, we found that physicians who completed a residency in pediatrics, family medicine and gynaecology and obstetrics appear to be more inclined to work in country-side regions, where primary care is the health system focus. This is most likely because people that choose this specialties have a higher preference for working in needy regions.

Therefore, these preliminary results indicate that opening more residency programs related to primary care in underserved areas might increase the physicians supply in these regions.

Second, there is the intriguing result concerning physicians wages. They appear to be relevant in metropolitan regions but not in the countryside. This might be due to endogeneity and specifications problems which we intend to assess in the next steps of our work through the use of a mixed logit specification and the introduction of an instrumental variable for salary. However, this could also be related to the complaint made by many health professionals in developing countries. They declare the main reason for not wanting to work in remote communities is not its wages, but the precariousness of work conditions and its low life quality. We also intend to further investigate this possibility by adding new explanatory variables, such as health infrastructure and an entertainment proxy. If the latter possibility proves itself right in our refined model, the government should prioritize the provision of essential infrastructure to its citizens instead of higher salaries in underserved communities.

Finally, we would like to comment on two other extensions we intend to implement next. First, we intend to not only model physicians' first practice location decisions, but also their dynamic over time. This way we can assess how long they stay in a region along with which factors, such as work experience, might be responsible for changes. Understanding this dynamic is important since public policies should not only aim at increasing contemporaneous supply of health professionals in needy regions, but also encourage them to stay.

Second, we will try to incorporate general practitioners in our approach. These physicians did not complete a residency program and are usually oriented to primary care, the focus of remote and rural regions health system. Underserved areas are in more need of basic health care than of a specialized one, which emphasizes the importance of including general practitioner in our study. We intend to do it through the Federal Council of Medicine (CFM) database. It has the name of all physicians in Brazil, their state of birth, the medicine school completed and date of conclusion. The main difficulty is that we do not have their social security number, the key variable to merge this database efficiently with data in RAIS.

6 References

1. Agarwal, N. *An empirical model of the medical match*, American Economic Review, forthcoming
2. Bolduc, D., Ben-Akiva, M. *A multinomial probit formulation for large choice sets*. Paper presented at the sixth International Conference on Travel Behaviour, Quebec, May 22-24, 1991
3. Bolduc, D., Fortin, B., Fournier, M. *The effect of incentive policies on the practice location of doctors: a multinomial probit analysis*, Journal of Labor Economics 14 (4)(Oct, 1996), 703-732.
4. Homes, G.M. *Increasing physician supply in medically underserved areas*, Labor Economics 12 (2005), 697-725
5. Hurley, J. E. *Physicians choices of specialty, location and mode: a reexamination within an interdependent decision framework*, Journal of Human Resources 26 (1989): 47-91
6. Kling, J.R., Liebman J. B., Katz L.F. *Experimental Analysis of Neighborhood Effects*, Econometrica 75 (2007), 83-m119
7. McFadden, D., Train, K. *Mixed MNL Models for Discrete Response*, Journal of Applied Econometrics 11 (2000), 447-470
8. Train, K. *Discrete Choice Methods with Simulation*. Massachusetts: Cambridge University Press (2009).