UNIVERSIDAD CARLOS III DE MADRID

working papers

A METHODOLOGY TO MEASURE HOSPITAL QUALITY USING PHYSICIANS' CHOICES OVER TRAINING VACANCIES*

Matilde P. Machado, Ricardo Mora y Antonio Romero- Medina [1]

Abstract

In this paper, we propose an alternative methodology to rank hospitals based on the choices of Medical Schools graduates over training vacancies. We argue that our measure of relative hospital quality has the following desirable properties: a) robustness to manipulation from the hospital's administrators; b) comprehensiveness in the scope of the services analyzed; c) inexpensive in terms of data requirements, and d) not subject to selection biases. Accurate measures of health provider quality are needed in order to establish incentive mechanisms, to assess the need for quality improvement, or simply to increase market transparency and competition. Public report cards in certain US states and the NHS ranking system in the UK are two attempts at constructing quality rankings of health care providers. Although the need for such rankings is widely recognized, the criticisms at these attempts reveal the difficulties involved in this task. Most criticisms alert to the inadequate risk-adjustment and the potential for perverse consequences such as patient selection. The recent literature, using sophisticated econometric models is capable of controlling for case-mix, hospital and patient selection, and measurement error. The detailed data needed for these evaluations is, however, often unavailable to researchers. In those countries, such as Spain, where there is neither public hospital rankings nor public data on hospital output measures such as mortality rates our methodology is a valid alternative. We develop this methodology for the Spanish case. In a follow-up paper we will present results using Spanish data. In Spain graduates choose hospital training vacancies in a sequential manner that depends on their average grade. Our framework relies on three assumptions. First, high quality hospitals provide high quality training. Second, graduates are well informed decision makers who are well qualified to assess hospital quality. Third, they prefer to choose a high quality vacancy rather than a low quality one ceteris paribus. If these assumptions hold, then the first physicians to choose are likely to grab the best vacancies while the ones who choose last are stuck with the worst available. Thus, it is possible to infer from physicans' choices quality differentials amongst hospitals. We model the physician's decision as a nested-logit a la McFadden. Unlike in standard applications of McFadden's model, in our application the choice set is not constant across physicians but it shrinks along the sequential hospital choice process

Keywords: Hospital Quality, Ranking, Nested Logit, Revealed Preference, Physicians labor market.

[1] Universidad Carlos III de Madrid. Departamento de Economía. C/ Madrid, 126 28903 Getafe (Madrid).
E-mails: mmachado@eco.uc3m.es, ricmora@eco.uc3m.es , aromero@eco.uc3m.es ,

1　Introduction

Assessing the quality of health care providers is a priority of many governments, state agencies, insurance companies and often also patients. Accurate measures of provider quality are needed in order to establish incentive mechanisms, to identify which ones need quality improvements, or simply to increase market transparency and competition. Public report cards in certain US states and the NHS ranking system in the UK are two attempts at constructing quality rankings of health care providers. Although the need for such rankings is widely recognized, the number of criticisms at these attempts reveal the difficulties involved in this task (see for example the review by Shahian et al., 2001 and the references therein). Most criticisms alert to the inadequate risk-adjustment and the potential for strategic behavior among providers, such as patient selection, induced by these rankings. Nonetheless, the available studies show that there is a wide variation in the performance of health care providers (e.g. Shahian et al., 2001 and Burgess et al., 2003) and that, in some cases, differences in providers have widened in the past decades (e.g. McLellan and Staiger, 1999a).

Hospital quality assessment is particularly complex due to several reasons: First, hospitals produce a wide range of heterogeneous services, which makes it impossible to define "hospital production" in a simple way (see, for instance, McClellan and Staiger, 1999b). Second, the randomness of hospital output, e.g. because of small number of patients or uncommon conditions, and the existence of confounding factors, such as location-specific patient health characteristics, may lead to noisy measures of hospital quality. Third, patient selection and other non-random sources of patients assignment (see, for instance, Gowrisankaran and Town, 1999) will bias estimates of hospital quality. Finally, due to data collection costs, there is a lack of follow-up measures of treatment results. In this paper we provide a methodology to rank hospitals which is immune to these difficulties.

In many countries there is neither public hospital rankings nor public data on hospital output measures such as mortality rates. Lacking this information, we propose to construct a relative hospital quality measure using publicly available data on physicians' choices over hospital vacancies at the beginning of their specialized training. We adapt our methodology to the case of Spain. In Spain, after graduation from medical school, physicians willing to start a career as a specialist must pass a national exam. Conditional on passing the exam, they choose hospital training vacancies in a sequential manner that depends on their grade.

Our methodology relies on three assumptions: first, high quality hospitals provide high quality training; second, physicians are well informed decision makers who are well qualified to assess hospital quality; third, all else equal, physicians prefer a higher quality vacancy. The latter assumption is reasonable especially in this very competitive and tight labor market and can be justified not only because physicians receive a better training but also because it provides a signal to the post-training market.[1] If all three assumptions hold, physicians who choose first are likely to grab the "best" vacancies

---

[1]Specialist certificates, obtained after completing the training period, specify the hospital where the training took place. The name of the training hospital is, therefore, used in the labor market as a signal of quality.

while the ones who choose last are stuck with the "worst" vacancies available. We use the traditional random utility framework to model the physician's choice among the vacancies available to her.

In short, we exploit the physicians' sequence of hospital choices to infer quality differentials among hospitals. Conditional on having enough observations, our methodology for measuring relative hospital quality has the following desirable properties: a) robust to manipulation from the hospital's administration; b) it can be extended to allow for differences in quality across services within a hospital; c) inexpensive in terms of data requirements, and d) not subject to selection bias from patients (Gowrisankaran and Town, 1999) nor hospital screening of patients. We expect property a) to hold because it should not be possible for hospitals to influence physicians' choices but through their performance.[2] Property b) follows because hospitals open vacancies in several specialties. Property c) is satisfied because physicians' choices are routinely collected. Finally, our relative quality measure is not subject to selection bias because it is neither based on the decisions of patients nor of hospitals.

When using physicians' decisions to evaluate hospital quality three issues should be addressed. First, physicians' choices depend on personal as well as career considerations. For example, when choosing a hospital physicians are simultaneously choosing place of residence. Second, the sequential structure of the problem implies dependence across observations. As physicians gradually fill in the vacancies, they decrease the choice sets of remaining physicians. We propose a model similar in spirit to the sequence of "exploded" conditional multinomial logit proposed by Bradlow and Fader (2001). Our model is then extended, following McFadden's nested logit, to account for different patterns of substitutability across hospitals in different locations. Finally, drawing from the informational cascades literature (see, for example, Banerjee, 1992, and Allsopp and Hey, 1999) dependence may arise because the decisions from previous physicians embody useful information regarding vacancies' attractiveness.

The remainder of this paper proceeds as follows. Section 2 describes the literature on hospital evaluation. Section 3 sketches the underlying institutional background in Spain. Section 4 describes the data set and specification strategies. Section 5 presents the econometric framework. Section 6 provides some concluding comments. Finally, the Appendix contains figures and tables.

2   Literature Review

Early evaluations of hospital production were based on accounting exercises of the inputs used during an average stay. As better data were made available, other hospital indicators, such as the length of the average stay or the average turnover, were gradually incorporated in the studies. Yet, it soon became clear that the use of such indicators could be easily influenced by hospital administrators so

---

[2]Besides hospital quality, one could also claim that physicians worry about tenure promotion. Physicians might tradeoff quality for less future job uncertainty. Hospitals administrations could give assurances of future promotions to attract good candidates. Since these promises would be unobserved, a way to control for the probability of getting a tenure position in a given hospital would be to condition for the number of tenured physicians per number of beds, since this would be a proxy for future openings.

that improvements in the indicators were often associated with situations for which hospital quality had not actually improved. This was considered a serious drawback for the practical usefulness of these indicators. Ideally, quality measures should not be easily manipulated by the hospital administrators (see for instance, Lu, 1999, for a description of this problem in the context of substance abuse treatment).

More recently, some studies have used treatment results, such as mortality rates, to assess hospital quality after controlling for observable characteristics. For instance, Normand *et al.* (1997) compare mortality rates for acute myocardial infarction after using hospital and patient characteristics as risk-adjusters. At least two difficulties may still arise. First, since the patient's true health status is not reflected in just one binary indicator, such as a imminent survival, the results are necessarily a partial analysis of hospital quality. Recent studies have tried to address this limitation by incorporating additional health measures over time (see, for instance, McClellan and Staiger, 1999a, 1999b, and Ackerberg, Machado and Riordan, 2006). Nonetheless, the potential for biased estimates of hospital quality due to insufficient gathering of relevant health information remains. The obvious solution to this problem, i.e. the exhaustive gathering of information regarding treatment results, does not seem practical as data collection costs could increase substantially.

The second difficulty with simple risk-adjustment is that it fails to adjust for non-observable factors. For example, Gowrisankaran and Town (1999) as well as McLellan and Staiger (1999a) argue that the endogeneity in the process of patients' assignment to hospitals is a source of bias when estimating hospital quality because the best hospitals will usually be assigned the most severe cases. Gowrisankaran and Town propose a simple linear instrumental variables approach to the mortality rate model using distances to non-selected hospitals as instruments of hospital choice to deal with the endogeneity but, they caution, their method is not appropriate to construct a ranking of hospitals due to the high standard deviations of the estimated hospital-dummy coefficients. McLellan and Staiger, on the other hand, decide to restrict their analyses to heart attack patients who have less of a possibility to choose hospitals due to the urgency of treatment. Geweke, Gowrisankaran and Town (2002) go one step further and jointly model the hospital choice and the mortality outcome for Medicare patients suffering from pneumonia. They compute the probability of mortality under the hypothetical experiment of random admission to hospitals using bayesian Markov Chain Monte Carlo simulation techniques. In a somewhat different vein, Ackerberg et. al. (2006) address the problem of endogeneity in their study on alcoholism treatment by distinguishing between the intrinsic quality of the health provider and the effect of the patient's unobservable characteristics in the assignment to health providers.

In contrast to the previous studies where patient level data are used, we use public data on physicians' choices over training vacancies at the beginning of their careers. Thus, our proposal relies on a revealed preference argument in the labor market for training vacancies. Avery et al. (2004) use a similar argument to rank colleges in the United States, however, our methodologies and applications differ in several ways. First, Avery et al. (2004) acknowledge a potential self-selection problem that would bias their ranking results. In their application, students' choice sets are not exogenous, instead they are a subset of colleges pre-selected by the students. In the Spanish case, choices sets are given for each physician. Second, their methodology hinges on the availability of exogenous and precisely recorded

3

students' place of residence. As we argue below, this is not the case for the Spanish data. Hence, we extend the basic framework to address the endogeneity issue.

3   The labor market for specialists

In Spain, it is mandatory for medical school graduates to complete training programs in hospitals during a number of years before they can practice as specialist physicians. Not all hospitals open training vacancies. Hospitals must fulfill certain requirements before they can open training vacancies. These requirements are established by the Ministry of Education and the Ministry of Health. The number of vacancies in each hospital is determined by the government after consultation with hospital administrators and depends on the hospital training capacity and the needs of the population in the surrounding area. In conversations with government officials involved in the process we were told hospitals seek to obtain as many vacancies as they can manage as an inexpensive way of filling their personnel needs. Wages for training positions are the same across all hospitals. The data suggest that given hospital size, the number of vacancies is unrelated to hospital quality.[3]

Overall, according to data from the National Catalogue of Hospitals, only around 22% of health care providers are part of the training programs for specialists. These providers share a number of features. All of them are hospitals and, on average, larger than those institutions which do not train specialists. Training hospitals have, on average, 540 beds as opposed to an average of 117 beds for those hospitals without training vacancies. In addition, training hospitals have special health care equipment, such as extracorporeal shock wave lithotripsy, cobalt treatment equipment, and particle accelerators. Nearly 75% of training hospitals report at least one type of special equipment while the figure for non-training hospitals is only 36%. The percentage of private and public hospitals that open training vacancies are 4.65% and 43%, respectively. These statistics suggest that the sample of training hospitals includes a large proportion of the most important hospitals in the Spanish health care system.

Hospitals open vacancies in different specialties. Across all hospitals, the total number of vacancies per specialty-year is very unequal.

The match between physicians and vacancies follows a serial dictator procedure. First, physicians are ranked according to an average score, which is obtained as a weighted average between a standardized national exam score[4] —with a weight of 75%— and the medical school grade-point-average — with a weight of 25%. Every year the number of physicians who pass the exam is higher than the total

---

[3]The number of training vacancies is highly correlated (0.83) with the number of beds, a proxy for hospital size. On the other hand, in a regression of the number of training vacancies per hospital on its lagged value, the number of beds, and the lagged proxy for hospital quality (the rank of the first candidate who chooses that hospital) shows that the latter is not significant. This suggests that the order in which physicians choose each hospital does not affect the number of training vacancies per hospital during our sample period.

[4]The standardized national exam is called the MIR exam after the Spanish acronym for "in-hospital resident physician."

number of vacancies available.[5] Physicians may choose a vacancy following the sequence established by the rank. Some physicians, however, decide to drop out from the process. Most of these have a low rank which indicates that they prefer an outside option to the vacancies still available to them (e.g. Gonzalez, 2004). Conditional on choosing a vacancy, physicians' dominant strategy is to choose their best option from those still available.[6]

4   Data

In the follow-up paper we plan to use several data sets. The main dataset contains the sequence of hospital-specialty choices made by physicians candidates from 1995-2000. These data are publicly available from the Spanish Ministry of Health and contain information on physician characteristics such as her position in the queue, gender, self-reported province of residence, college where she graduated, and her hospital-specialty choice.

As mentioned in the previous Section, physicians may drop out from the process and choose their outside option. We can identify the rank of those individuals who drop out from the process, however there is neither information on any of their other characteristics nor on their outside option. Therefore, we model the physicians' choice as a two-part model reflecting both the decision of choosing either the outside option or one of the available vacancies.

During our sample period either military or social services is compulsory for males. Individuals may comply with the compulsory service during their student years or delay it up to a limit. Physicians who decide to join the service just before their specialist training have the right to reserve a vacancy and fill it after service completion. Because these physicians do not fill the vacancy immediately, physicians who follow in the queue will fill it. This means that the set of vacancies available is not affected by reservations. We are unable to deal properly with male physicians headed to the compulsory service. From 1995-1997 the data do not include the reservations during these years which implies a loss of information. The absence of these reservations will not result in a bias as long as there is enough heterogeneity in the hospitals chosen by reservists. We do not expect the bias to be significant given that the number of reservations is always less than $6, 7\%$ a year. During 1998-2000 the data includes the reservations but we are not able to identify them, causing two potential problems. First, it could be argued that because the number of vacancies in hospital-specialty combinations chosen by reservists is artificially increased, i.e. is endogenous, it could bias results. This reasoning is wrong because what is relevant for physicians' choices is the set of available hospital-specialty combinations. Second, as

---

[5] According to the Curso Intensivo Mir http://www.curso-mir.com/nuestros_result/06.htm the excess of physicians over vacancies during the last years fluctuated between 19% and 78% of the vacancies.

[6] González (2004) claims that before 1996 some physicians chose from the available set any specialty in the hospital they wished to be located with the aim of requesting a change of specialty once the training period started. She also argues that regulation put in place in 1995 stopped these short-cuts by making it very costly to change specialties. We believe this strategic effect which, if at all, is only present in the 1995 data could only bias the hospital coefficients for those hospitals where switching specialties during training is perceived by the physician candidates to be easier.

reservists cannot choose an outside option, the set of choices for them is different than the one for other physicians. We do not expect this problem to be severe as the number of reservations has most likely decreased in these years due to the announcement of the abolishment of the compulsory service by 2001.

The main dataset has around $2,845$ vacancies each year totalling $16,977$ observations from $1995-2000$, after applying several filters to control for misspelling in the coding of the hospital or the specialty.[7] The total number of hospitals in the data set is around 175. This number varies by year because some hospitals are merged and others are split into several ones during the period under analysis. We treat hospitals resulting from either merges or divisions as new hospitals.[8]

Typically, for a particular hospital-specialty combination a very small number of vacancies is offered. Of all hospital-specialty combinations available, roughly 84% correspond to a single vacancy being offered and in 95% of the cases the number of vacancies is smaller than three, as the next figure illustrates:
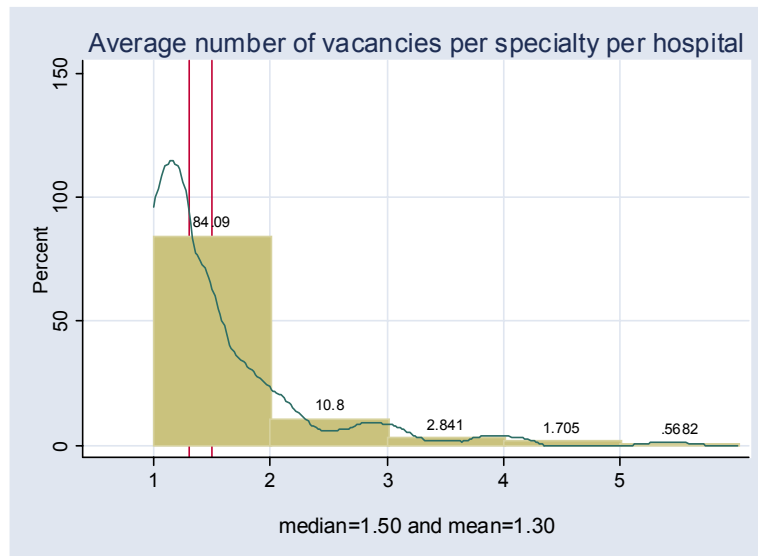
The average number of vacancies per hospital (over all specialties) per year is 18.14 but this distribution is very disperse and skewed as the next figure shows:

---

[7]After 1996 physicians who wished to become General Practitioners participated in a separate process which included a different exam and a different selection mechanism. This mechanism run parallel to the other specialties. For this reason and because GP is a non-hospital based specialty we decided not to consider these cases.

[8]The year with the lowest number of hospitals is 1995 with 144 while the year with the highest number is 2000 with 158.

Average number of vacancies per hospital

median=7.58 and mean=18.14
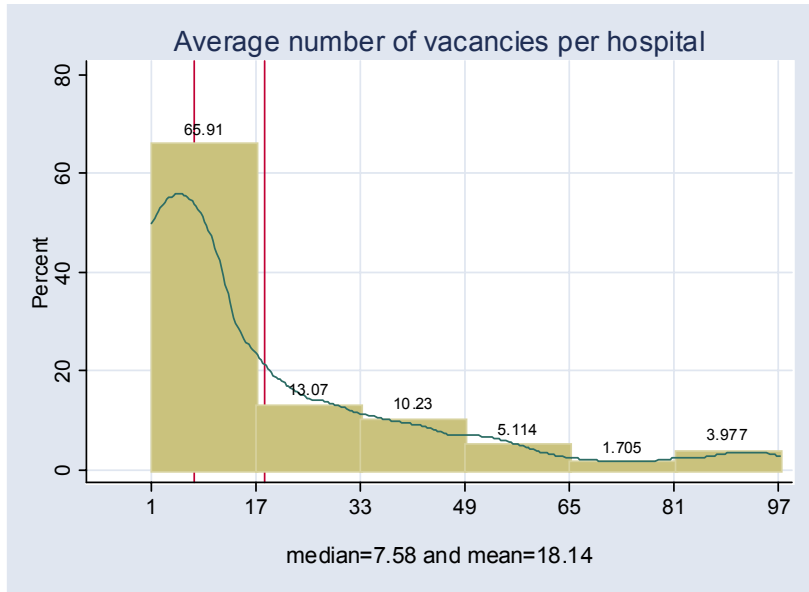
However, the average number of different specialties offered at each hospital-year is on average relatively large (10.68).

Average number of specialties per hospital
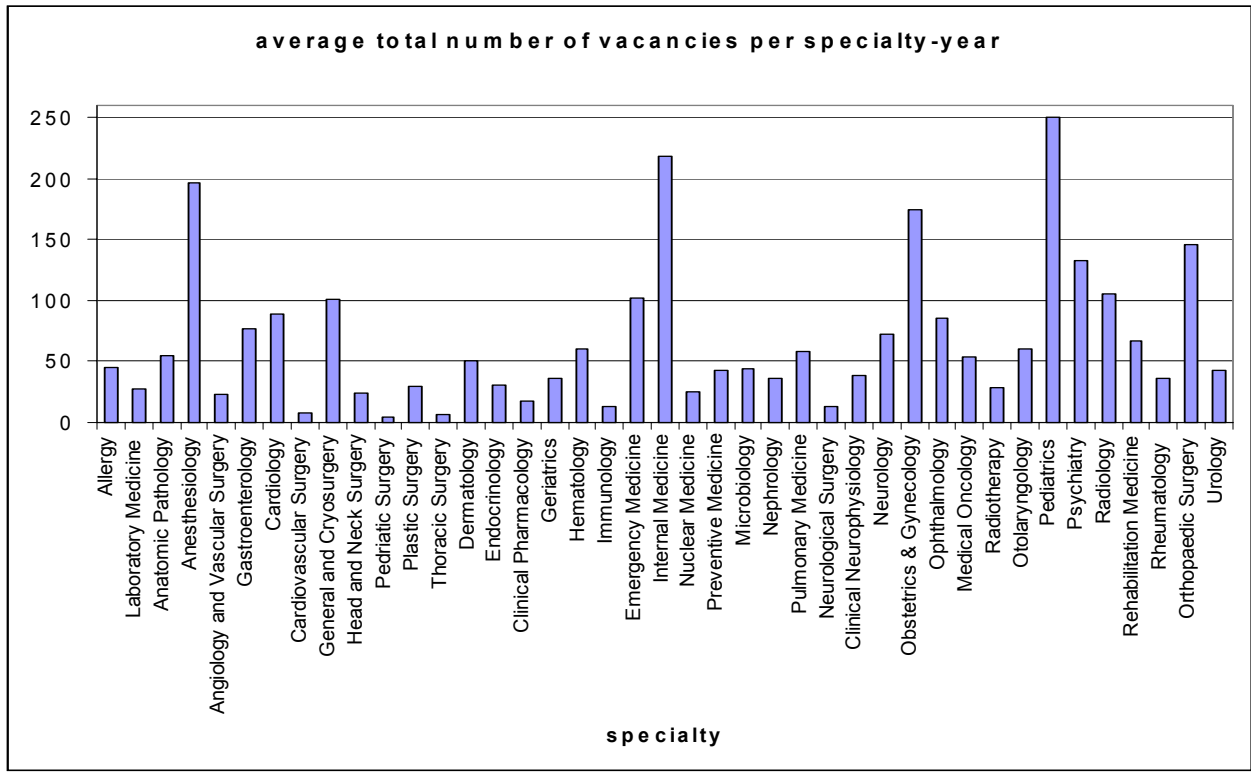
median=6.83 and mean=10.68
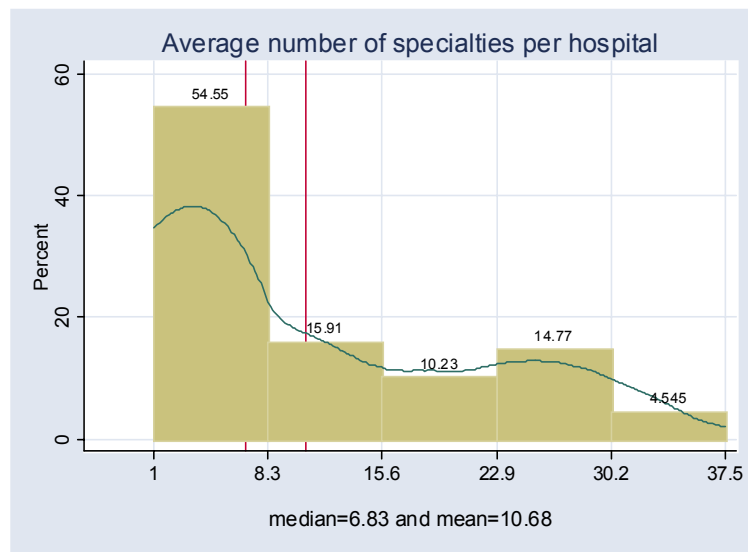
The specialties in our sample are unevenly represented as the following figure shows:

average total number of vacancies per specialty-year

specialty

In addition, there is a marked contrast between the distribution of male and female physician across specialties. Figure 1 shows the deviation of the female share by specialty to the overall female share in percentage terms. This index will be close to zero whenever the female share in a given specialty is close to the overall female share (58.6 percent) as is the case for Emergency Medicine. The most striking feature of Figure 1 is the large dispersion of female shares by specialty. For 27 out of 41 specialties, female shares deviate from the mean by more than 10 percent. At one extreme of the distribution are Pediatrics, Allergist, and Obstetrics & Gynecology (O&G) where female shares exceed by more than 30 percent the overall average. At the other extreme, we find Urology, Neurological Surgery, Cardiovascular Surgery, and Orthopaedic Surgery where the percentage of female is lower than the average by more than 50 percent.

From the information available in the physicians' data set it is possible to construct a variable which reflects a notion of geographical proximity between the physicians' residence and each of the hospitals. Whereas the hospitals' exact address is available, there are, however, two problems with the physicians' declared residence. First, physicians only declare the province of residence, not the city. Second, we believe this variable is not measured in a consistent way. Physicians who study where their parents live,
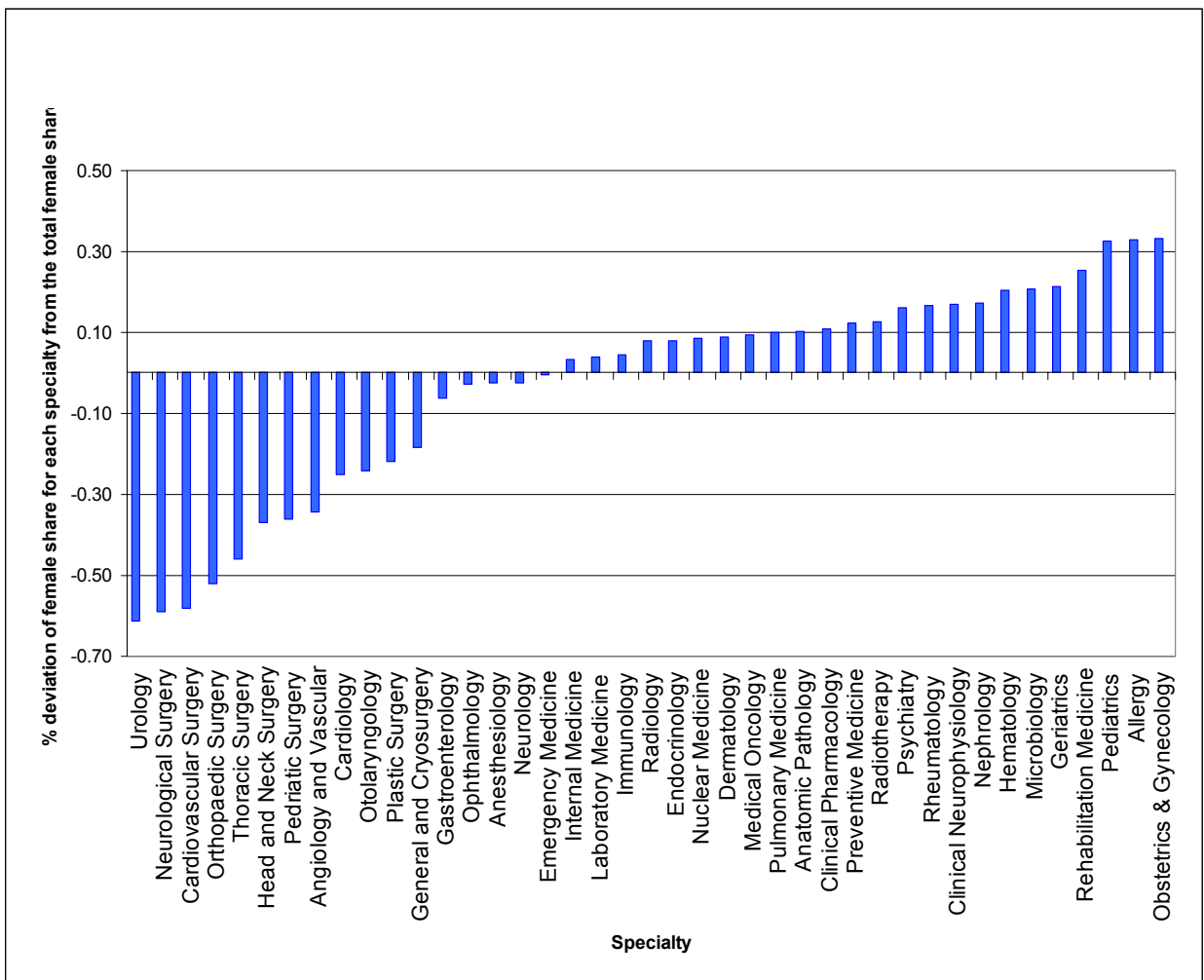
Figure 1: Gender segregation by specialty

declare their residence in a consistent way.[9] In contrast, among the physicians who study in a different province from their parents', some declare their parents' residence as their own (these represent 14% of physicians in our sample) while, as we argue below, others declare the province where their college is located. We think that this inconsistency is not random. Studying in a good college enhances the likelihood of attaining a high score in the MIR exam. Consequently, good students, who aim to be trained at a top hospital, are more likely to move from their parents' home to study at a good college. Hence, the physician's province of residence is not only measured inconsistently but this inconsistency is likely correlated with the student's ability and her ranking. Alternatively, it may be argued that there is no such inconsistency because the declared province of residence reflects intended location and, therefore, would be correct to use it as the physician's location.

As an illustration of these issues consider the cases of Madrid and Barcelona, the two most populated provinces in Spain. Both Madrid and Barcelona act as focal points for medical students from other parts of the country. As can be seen from the statistics in Table 1, the percentage of the overall college student population in Madrid over the total national (22.13%) is considerably higher than the percentage of citizens under 18 years of age over the total national (12.99%), which shows that Madrid's universities attract students from the rest of the country. The same phenomenon is partly observed in our data where 17.32 percent of physicians graduated in Madrid. This number is very close to the percentage of physicians in the sample who declare Madrid as their province of residence (18.10%). Therefore, the variable "province of residence" in our data shows Madrid as a focal point. In contrast, although the percentage of college population in Barcelona is also higher than the percentage of minors (14.18% versus 10.94%), the percentage who declare Barcelona as their province of residence in our sample (10.75%) is not higher than the percentage of minors living in Barcelona (10.94%). This suggests that while almost all students who went to study in Madrid declare Madrid as their province of residence, those who went to study in Barcelona still consider their parents' as their own residence.

| | % Province of Residence[a] | % College Population | | % Population under 18 Census 2001[b] |
		Sample[a] (medical schools)	Census 2001[b] (all colleges)	
Barcelona	10.75	11.54	14.18	10.94
Madrid	18.10	17.32	22.13	12.99
[a]Data from the physicians' data set.				
[b]Data from the Spanish National Statistics Office (INE).				

Table 1: Percentage of populations in Barcelona and Madrid over total national

To circumvent the problem of inconsistency and the potential biases associated with the residence location variable we also use the location of the college where the physician studied. This variable has

---

[9]Using data from the European Community Household Panel for the years of study, in Spain around 80% of college students live in their parents' residence.

the advantage of being precisely measured. Nonetheless, since declared residence may be interpreted as intended location we will show results based on both college as well as residence location.

Tables 3 and 4 in the Data Appendix show that physicians mostly choose hospitals in their college location (Table 3) or in their declared place of residence (Table 4). The preference to stay in the same region will be introduced in our model in the form of "nests" a la McFadden. The simplest specification has two nests. A hospital is in nest 1 when it is located in the same region as the college (residence) and is in nest 2 when it is located in any other region. We may use two definitions of region: the administrative division of the country (Comunidades Autónomas)[10], and travel-time-to-work region, which allows a more flexible definition of nests.[11] Using the data on travel time, a hospital is classified into nest 1 whenever it is located within reasonable commuting time, say 45 minutes, from the physician's college (residence) location. With this new definition, the classification of hospitals into nest 1 and nest 2 remains the same for 89 percent of the observations with respect to the administrative division of the country; for 9 percent of the observations, the classification changes from nest 1 to nest 2, reflecting the fact that in several administrative regions some hospitals are located beyond the 45-minute commuting time limit.

Nests based on geographical proximity variables play an important role in the identification of hospital quality because part of the identification comes precisely from those physicians who choose a hospital which is not in their province of residence. Physicians who are highly ranked will have more options and, therefore, a movement to another province reveals a stronger preference for the chosen hospital. On the contrary, physicians who are lower ranked may have to move to another province because there are fewer options to choose from. Therefore, those movements made by highly ranked physicians are central to identify hospital quality.

The second data source is the Spanish National Catalogue of Hospitals (*Catálogo Nacional de Hospitales*), which is published annually by the Ministry of Health. These data contains hospital characteristics such as location (city), number of beds (usually used as a proxy for hospital size), type of hospital (e.g. in our sample: general purpose (81%), and psychiatric (7%)), ownership (e.g. in our sample: public (86%), private non-for-profit (8%), or private for-profit (6%)), the corresponding regional authority, and some information about the available technical equipment (e.g. in our sample: emission tomography equipment—75% of hospitals have at least one—, and hemodynamic monitoring equipment—22% of hospitals have at least one).

Third, we also gathered annual average housing prices at provincial level from the Spanish National Statistics Office (INE) to control for regional differences in living costs.

---

[10]Spain is composed of 52 provinces. These are grouped into 17 larger administrative regions denominated Autonomous Communities.

[11]We use a similar dataset to the one used in Holl (2004) on travel time between college and hospital locations at the city level to construct a measure of geographical proximity.

5 The Econometric Framework

5.1 The Basic Model

In this section we propose the framework to construct a ranking of hospitals based on individual choices made by recently graduated physicians. In this subsection we simplify the decision model by assuming that geographical proximity to hospital plays no role in the physicians decision and that physicians must choose one of the available vacancies. Subsection 5.2 will extend the model to incorporate location, and in subsection 5.3 we will allow for outside options.

In our data set, physicians are ordered according to their average score. Let $i = 1, ...N$ simultaneously identify a physician as well as her position in the data. For example, $i = 1$ ($i = N$) denotes the first (last) observation in the data which corresponds to the physician who obtained the highest (lowest) score. Each physician i chooses one hospital-specialty combination from those available to her. Hospitals may open more than one vacancy for a given specialty. Denote by C_i the set of all hospital-specialty combinations available to physician i. If i chooses a hospital-specialty combination for which there are at least two remaining vacancies then $i + 1$'s choice set is identical to C_i. Eventually, however, as hospital-specialty vacancies are exhausted, choice sets must shrink: C_1 contains all possible hospital-specialty vacancies and C_N has a single vacancy available. Hence, choice sets satisfy the following:

$$C_i \supseteq C_{i'} \text{ for } i' > i. \tag{1}$$

We model physician preferences according to the stochastic random utility model where U_{ij} represents the utility to physician i from selecting hospital-specialty combination j. U_{ij} is decomposed into a deterministic component V_{ij} and a iid stochastic term ε_{ij} :

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{2}$$

where ε_{ij} is assumed to follow an extreme value distribution, i.e. $\Pr(\varepsilon_{ij} \leq x) = \exp(-\exp(-x))$. McFadden (1974) shows that the probability that physician i chooses the hospital-specialty combination $j^* \in C_i$ is given by:

$$P_{ij^*} = \frac{\exp(V_{ij^*})}{\sum\limits_{j \in C_i} \exp(V_{ij})}, \text{ for } j^* = 1, ..., J_i \tag{3}$$

where J_i denotes the number of elements in C_i.

Let $\pi = (\pi_1, ...\pi_N)$ be the observed sequence of physician choices where π_i indexes the hospital-specialty combination chosen by physician i. The assumption of iid error terms implies that the probability of observing a given sequence of hospital-specialty choices, $\Pr(\pi)$, is a sequence of independent multinomial logit models:

$$\begin{aligned} \Pr(\pi) &= \Pr(U_{1\pi_1} \geq U_{1j} | \forall j \in C_1) \times \Pr(U_{2,\pi_2} \geq U_{2j} | \forall j \in C_2) \\ \times... \times \Pr(U_{N-1,\pi_{N-1}} &\geq U_{N-1j} | \forall j \in C_{N-1}) \end{aligned} \tag{4}$$

12

Equation (4) is analytically equivalent to the "exploding conditional multinomial logit" of Chapman and Staelin (1982) and Bradlow and Fader (2001).[12] Since physicians choose their preferred hospital-specialty combination from their choice set C_i, the sequence of observed hospital-specialties choices π provides useful information to construct a ranking of hospital-specialties combinations.

Physician preferences over hospital-specialty combinations depend on interactions of hospital-specialty characteristics and individual characteristics. We assume that V_{ij} depends linearly on a vector x_h of hospital characteristics, a vector z_s of specialty characteristics, and finally on a vector y_{ihs} of interactive variables relating physician i to hospital-specialty j:

$$U_{ij} = x_h\beta_h + z_s\beta_s + y_{ihs}\beta_{hs} + \varepsilon_{ij}. \tag{5}$$

In a simple specification of the basic model, the vector x_h in (5) contains only hospital dummy variables, z_s contains specialty dummies, and y_{ihs} is an interaction term between physician's gender and specialty chosen. The motivation for the latter is taken from the pattern of segregation between female and male specialty choices described in the Data Section.

In order to construct an appropriate hospital ranking we need to infer from the physicians' choices a measure of the hospitals' intrinsic quality. The estimates of the hospital dummy coefficients, β_h, provide a reasonable approximation for a measure of the latter. Nevertheless, two shortcomings should be acknowledged: first, hospital quality may vary across specialties. This issue could, in theory, be addressed by including hospital-specialty dummy variables. For some specialties and hospitals, however, there are not enough observations to obtain precise estimates. The assumption that hospital quality is constant across specialties can be checked by estimating the model for a subset of specialties for which there are enough observations. Second, there are alternative interpretations of the β_h's. By intrinsic quality we understand the unobserved component of hospital quality that stems from the proficiency of its management and staff.[13,14] Most likely the estimated β_h's capture not only intrinsic quality but also other hospital characteristics with little time variation in our sample. For example, physicians may prefer to be placed in a larger hospital because they may learn more. Hence, if we do not control for size, the estimated β_h's for larger hospitals would be an upward-biased estimate of quality. In order to minimize this potential bias we should control for hospital observable characteristics such as the availability of sophisticated equipment.

---

[12]Our model, however, is intrinsically different. In our case, the ranking π is obtained from the choice of different individuals. Each physician chooses one option only and does not provide a complete ranking of hospital-specialties. In contrast, Chapman and Staelin's (1982) model assumes that each individual in the data set provides a ranking of choices. Bradlow and Fader (2001) data, on the other hand, consists of weekly observations of the pop songs hit list Billboard "Hot 100". Their ranking is provided by the level of record sales and, therefore, implicitly aggregate the choice of many different individuals. They do not model or mention the aggregation aspect of their data but instead work as if a single entity, *society,* chooses the ranking.

[13]McLellan and Staiger (1999a) find that hospital characteristics such as teaching institutions, for-profit organization, number of beds, and volume cannot explain all the differences in standardized mortality rates. This suggests that unobservables such as intrinsic quality may explain them.

[14]Ideally, one should control for unobserved heterogeneity. However, given the large number of available vacancies in our data set, we decided for computational reasons not to extend the model to incorporate unobserved heterogeneity.

The Likelihood function for the basic model takes the form:

$$L(\pi|\beta) = \Pr(\pi) = \prod_{i=1}^{N} \frac{\exp(V_{ij^*}(X_{ij};\beta))}{\sum_{j \in C_i} \exp(V_{ij}(X_{ij};\beta))} \tag{6}$$

where X_{ij} is the set of explanatory variables, and β the vector of parameters. An important drawback from this model is that it requires the *Independence of Irrelevant Alternatives* (IIA) to hold.

5.2 Introducing Location

When physicians choose a particular hospital they are implicitly choosing their residence for the following four to five years. Physicians have a preference to stay in the region where they graduated or reside. If physicians decided upon location randomly then, for example, the percentage of physicians from Madrid, the region with the highest number of vacancies, who choose to stay in Madrid should be on average equal to the percentage number of vacancies in Madrid (24%). Yet, Tables 3 and 4 in the Data Appendix show that the figure for Madrid is 83 percent well above the prediction of the random location. For other regions the percentages are between 24.82 and 89.37 percent. These statistics suggest that hospital-in-same-region as college (residence) is an important hospital characteristic affecting the physicians' choice. Interestingly, Table 2 shows that the higher the rank of the physician, the more likely she is to stay in the same region.

At first, Table 2's statistics are counter-intuitive since one expects the highest ranked physicians to give a higher weight to hospital quality and, therefore, to be more willing to move to a different region in order to be trained at a better hospital. There are two possible explanations for the numbers in Table 2. The first one lies on a clustering argument whereby neither college nor hospital quality are evenly distributed across the country but are instead concentrated in the same regions. In the presence of clustering, top ranked students want to move to study medicine in the best colleges in order to increase their position in the MIR ranking. Since the best colleges are located in the same regions as the best hospitals then physicians do not need to move again for specialized training. Under this explanation the hospital-in-same-region-as-college variable is endogenous since unobserved factors that affect hospital choice also affect college location. For this reason, the introduction of hospital-in-same-region as college (residence) in the physicians' utility, equation (5), would lead to biased estimates. The nested-logit framework presented below allows us to introduce this preference for proximity without causing a bias in the estimated coefficients.

The second explanation for the statistics in Table 2 lies in the existence of high moving costs. In this case, only the top ranked physicians would find enough vacancies in their region to choose from while the lowest ranked must take whatever is left.

	Rank quantiles			
	25%	25% − 50%	50% − 75%	75% − 100%
hospital in same region as college	74.11	73.38	67.51	55.73
hospital in same region as residence	76.00	76.90	71.12	59.26

Table 2: Percentage of Physicians in each quantile that choose a hospital in the same region as college or residence

Figure 2 shows the percentage of students from each region that are ranked amongst the top 25 percent.[15] The students from the best performing region fare roughly 50 percent better than those students coming from the worse performing region. If the same analysis is performed for the students in the top 50 percent, the differences between the best performing region and the worst are still large (30 percent).

On the one hand, data from figure 2 suggest that college quality is not evenly distributed across the country, on the other hand, the high moving costs explanation goes against the evidence that Madrid and Barcelona attract a large number of students from other regions (Table 1). To summarize, the clustering hypothesis seems a more plausible explanation for the statistics in Table 2.

We introduce location by extending the multinomial logit framework to a nested logit (McFadden, 1978) where the physician's hospital-specialty choice implicitly involves a location choice. Although the choice of location and hospital-specialty is simultaneous, the nested logit can be interpreted as a sequential choice model where one first chooses a nest and then a hospital-specialty within that nest.[16] In contrast to the multinomial logit, in the nested logit *IIA* only holds within nests. The ratio of choice probabilities of any two hospitals-specialty combinations in the same nest is independent of all other alternatives within that nest but the ratio between combinations from different nests depends on the alternatives on those nests.

We partition the set of all hospital-specialty combinations available for each physician i into two nests, B_{ik}, $k = 1, 2$. Without loss of generality, nest 1 includes hospitals located close to the physician's location (college or residence) while nest 2 includes hospitals located elsewhere. Thus, nests are defined as an interaction between hospital and physician locations.

As in the previous Section, choice sets, C_i, are individual specific and shrink as hospital-specialty combinations are exhausted along the process. The physician's utility from choosing alternative j belonging to nest k, $j \in B_{ik} \subset C_i$, is:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{7}$$

[15] We dropped from Figure 2 the regions with only one college to comply with a confidentiality agreement. The range for the excluded regions is similar to the range shown in Figure 2.

[16] Conceivably we could extend this framework to the hospital-specialty choice with a three-level nested logit where locations and specialties define nests. This would be a more general model than the one we propose here. However, given the high number of specialties, the number of parameters would become computationally burdensome.
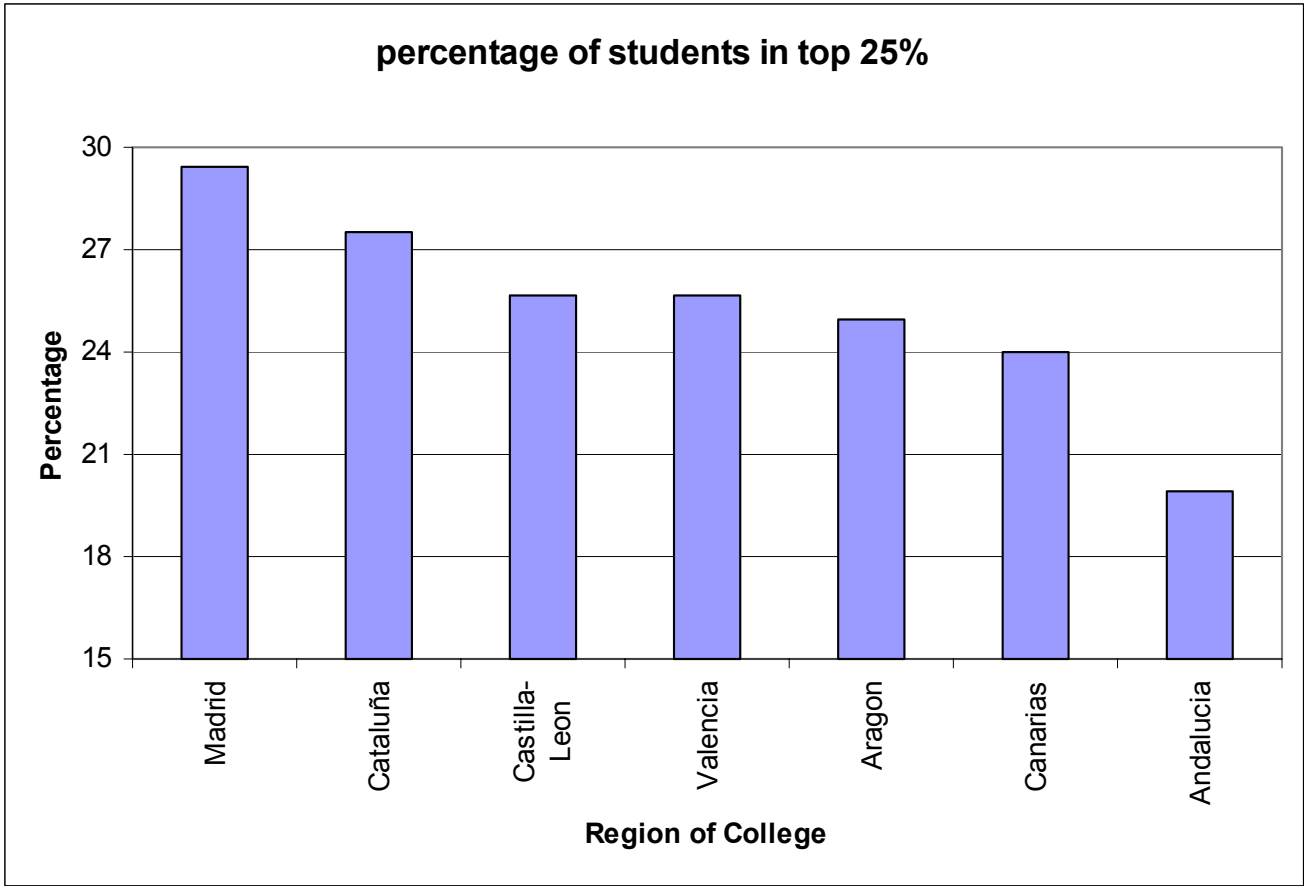
Figure 2:

where V_{ij} depends linearly on a vector x_h of hospital characteristics, a vector z_s of specialty characteristics, and finally on a vector y_{hsi} of interactive variables relating physician i to hospital-specialty hs just as in equation (5). The difference with the multinomial model lies in the distribution of the error term, which now depends on the definition of nests. To be precise, the vector $\varepsilon_i = (\varepsilon_{i1}, ... \varepsilon_{ij}, ... \varepsilon_{iJ})$ has cumulative distribution:

$$\Pr(\varepsilon_i \leq x) = \exp\left(-\sum_{k=1}^{2}\left(\sum_{j \in B_{ik}} e^{-\frac{x_j}{\lambda_k}}\right)^{\lambda_k}\right). \tag{8}$$

If $\lambda_k = 1$ for all k then the ε_i's are independently distributed and the model coincides with the multinomial logit. Values of $\lambda_k < 0$ are not consistent with utility maximization.[17] The parameter λ_k is a measure of the degree of independence among the ε_i's in nest k (see for example Train, 2003, pp 83). Less independence between options in the same nests (lower value λ_k) increases differences in choice probabilities across vacancies within nest k. Intuitively, the higher the correlation between the unobserved components of utility for different options the more often the same hospitals are "winners" or "losers" for different physicians.

It can be shown (McFadden, 1974) that the probability that physician i chooses the hospital-specialty combination $j^* \in B_{ik} \subset C_i$ is given by:

$$P_{ij^*} = \frac{e^{\frac{V_{ij^*}}{\lambda_k}}\left(\sum_{j \in B_k} e^{\frac{V_{ij}}{\lambda_k}}\right)^{\lambda_k - 1}}{\sum_{l=1}^{2}\left(\sum_{j \in B_l} e^{\frac{V_{ij}}{\lambda_l}}\right)^{\lambda_l}}, \text{ for } j^* \in B_{ik} \tag{9}$$

The difference $\lambda_1 - \lambda_2$ measures the relative preference for nest 1. This can be seen in a simple example. Suppose both nests have the same number of vacancies, N, and $V_{ij} = V$ for all i and j. Then the probability of choosing any vacancy in nest 1 relative to the probability of choosing any vacancy in nest 2 equals $N^{(\lambda_1 - \lambda_2)}$. In our data, we expect λ_1 to be close to 1 and $\lambda_2 < 1$ implying that the valuation of a vacancy in nest 1 relative to one in nest 2 is independent of alternatives in nest 1 but may depend on alternatives in nest 2.[18] In other words, for $\lambda_1 = 1$, a lower value of λ_2 increases the probability of choosing any hospital in nest 1 vis-a-vis any hospital in nest 2 and simultaneously widens

---

[17]Train (2003) explains that values of $\lambda_k \in (0, 1]$ are always consistent with utility maximizing behavior while $\lambda_k > 1$ is consistent only for a range of values of the independent variables X's.

[18]For any candidate i, if $\lambda_1 = 1$ and $\lambda_2 < 1$ the odds ratio of an alternative in nest 1 relative to an alternative in nest 2 is:

$$\frac{e^{V_{ij^*}}}{e^{\frac{V_{ij}}{\lambda_2}}\left(\sum_{j \in B_2} e^{\frac{V_{ij}}{\lambda_2}}\right)^{\lambda_2 - 1}}$$

17

the differences in probability ratios between hospitals in nest 2, i.e. hospitals in nest 2 are perceived as less substitutable. Analogously, a high value of λ_1 implies that alternatives in nest 1 are seen as more substitutable, which is intuitive since proximity to hospitals is a crucial factor in the physicians' choice.

Importantly, the nested-logit model allows us to introduce preference for proximity without causing a bias in the estimated coefficients. To see this, note that the proximity variable, which defines individual-specific nests, does not affect the deterministic component of utility directly, i.e. is not included in V_{ij}, and only affects the distribution of the error term ε_i.[19]

The existence of clustering, irrespectively of moving costs, can be tested in our framework by rees-timating the model for a subsample of top ranked physicians. If college quality were evenly distributed across all regions then college region would be regarded as predetermined and uncorrelated with the MIR rank. In other words, the student population of any two regions would be ranked similarly in the MIR process. This would imply that $\lambda_1 - \lambda_2$ would only reflect the relative preference for nest 1 but should not vary with the restriction of the sample to the highly ranked physicians. On the other hand if clustering exists, i.e. the best hospitals are located in the same region as the best colleges, we would expect higher values for $\lambda_1 - \lambda_2$ when we restrict the sample to the top ranked physicians. On the contrary, if the best hospitals are located in different regions from the best colleges then, restricting the sample to the top ranked should lead to a lower $\lambda_1 - \lambda_2$ i.e. a preference to move to a different region.[20]

5.3   Introducing the Outside Option

As explained in Sections 3 and 4 physicians have the option of dropping out from the process, i.e. not choosing any of the vacancies in their choice set. We can identify the rank of dropouts from the process, however there is neither information on any of their other characteristics nor on their outside option.

In the previous Sections, the econometric model does not include an outside option. As long as the odds ratio between any two hospital-specialty combinations does not depend on the individuals' outside option, maximizing the likelihood conditional on choosing a vacancy should give consistent estimates of the parameters in P_{ij*}. This result holds if the probability of choosing an outside option does not depend on the quality of the hospitals available in the choice set. However we find that the probability of choosing an outside option decreases with the physicians order in the queue. As the next figure shows for the year 2000, for example, the percentage of dropouts among the top 3000 physicians is only 8.5 percent while this percentage jumps to 38 percent for the top 5000.

The previous graph suggests that the probability of choosing an outside option may depend on the

---

[19]Avery et al. (2004) introduce the same-region-as-residence variable as a component of V_{ij} in their model of college choice. This specification causes them no bias since, in their data, the location of residence is precisely measured and exogenous.

[20]We also expect both λ_1, λ_2 to increase in a restricted sample because there are more options and correlation between vacancies should decrease.
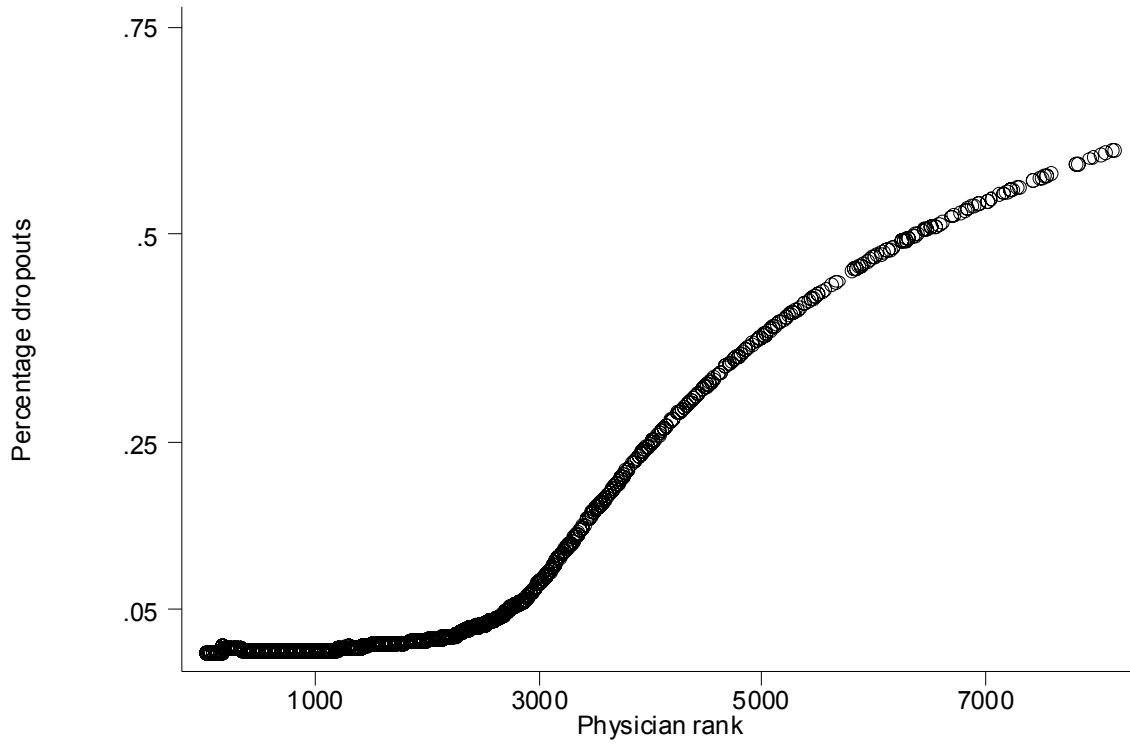
Figure 3:

19

quality of the hospitals available. Hence a physician chooses her outside option iff:

$$U_{i0} > \max_{j \in C_i} \{U_{ij}\} \tag{10}$$

where U_{i0} denotes physician i's utility derived from her outside option. One possibility would be to view the outside option just as another vacancy. The difficulty with this approach is the lack of information about the location of each physicians outside option. Given that physicians prefer to stay close to their college or residence, we assume that physicians' best outside option is always located in nest 1.

6   Conclusions

In this paper, we propose an alternative methodology to rank hospitals based on the choices of Medical Schools graduates over training vacancies. We model the physician's decision as a nested-logit a la McFadden that takes into account preferences for geographical location. Unlike in standard applications of McFadden's model, in our application the choice set is not constant across physicians but it shrinks along the sequential hospital choice process.

The dataset used in this paper and in the follow-up paper has a number of specificities such as physicians who dropout from the process and decide not to take any vacancy from those available as well as potential measurement error in the residence location information. Although important for our particular application, none of these problem affects the validity of our methodology and its practical usefulness.

References

[1] Ackerberg, Daniel A., Matilde P. Machado and Michael H. Riordan (2006): "Benchmarking for Productivity Improvement: a Health Care Application." International Economic Review, vol 47, $n^o 1$.

[2] Allsopp, A. and J. D. Hey, (2000): "Two Experiments to Test a Model of Herd Behaviour", Experimental Economics, 3(2):121-136.

[3] Avery, Christopher, Mark Glickman, Caroline Hoxby, Andrew Metrick (2004): "A Revealed Preference Ranking of U.S. Colleges and Universities," NBER 10803.

[4] Banerjee, A. V. (1992): "A simple model or herd behaviour", Quartely Journal of Economics, 107, 797-817.

[5] Bradlow, Eric T. and Peter S. Fader (2001): "A Bayesian Lifetime Model for the Hot 100 Billboard Songs," Journal of the American Statistical Association, Vol. 96, No. 454, Applications and Case Studies, pp. 368:81.

[6] Burgess, Simon, Gossage, Denise, and Propper, Carol (2003): "Explaining Differences in Hospital Performance: Does the Answer Lie in the Labor Market," CEPR DP 4118.

[7] Chapman, Randall G. and Staelin, Richard (1982): "Exploitingh Rank Ordered Choice Set Data Within the Stochastic Utility Model," Journal of Marketing Research, vol. XIX (August 1982), pp. 288-301.

[8] Commons, Margareth, McGuire, Thomas G and Michael Riordan: "Performance Contracting for Substance Abuse Treatment" Health Services Research, 32(5) December 1997 pp 630-650

[9] González, Beatriz, Patricia Barber, Mª Isabel Tocino, Elena Rodriguez (1997): "El mercado laboral sanitario y sus consecuencias en la formación. *Numerus Clausulus*". Mimeo Universidad de las Palmas de Gran Canaria.

[10] González, Beatriz (2000): "Formación y empleo de profesionales sanitarios en España. Un análisis de desequilibrios," Gaceta Sanitaria, 14(3), pp 237-246.

[11] González, Beatriz, Patricia Barber (2004): "El programa MIR como innovación y como mecanismo de asignación de recursos humanos." Mimeo

[12] Gowrisankaran, Gautam and Town, Robert (1999): "Estimating the Quality of Care in Hospitals Using Instrumental Variables" Journal of Health Economics, 18: 747-67.

[13] Geweke, John, Gowrisankaran, Gautam and Town, Robert (2002): "Bayesian Inference for Hospital Quality in a Selection Model." Mimeo.

[14] Herriges, J. and C. Kling (1996), 'Testing the consistency of nested logit models with utility maximization', Economic Letters 50, 33–39.

[15] Holl, Adelheid (2004): "Manufacturing location and impacts of road transport infrastructure: empirical evidence from Spain", Regional Science and Urban Economics 34 (3), 341-363.

[16] Kling, C. and J. Herriges (1995), 'An empirical investigation of the consistency of nested logit models with utility maximization', American Journal of Agricultural Economics 77, 875–884.

[17] Lu, Mingshan (1999)': "Separating the "True Effect" from "Gaming" in Incentive Based Contracts in Health Care," Journal of Economics and Management Strategy, 8, issue 3, 383-431.

[18] McLellan, Mark and Staiger, Douglas (1999a): "Comparing Hospital Quality at for-profit and not-for-profit hospitals," NBER WP 7324.

[19] McLellan, Mark and Staiger, Douglas (1999b): "The Quality of Health Care Providers" NBER WP 7327.

[20] McFadden, Daniel (1974):

[21] McFadden, Daniel (1978): "Modelling the choice of residential location" in A. Karlquist, et al. (eds.), Spatial Interaction theory and planning models, Amsterdam, North-Holland.

[22] Normand, Sharon-Lise T., Glickman, Mark E., and Gatsonis, Constantine A., (1997) "Statistical Methods for Profiling Providers of Medical Care: Issues and Applications", Journal of American Statistical Association, Vol. 92, No. 439, Applications and Case Studies, 803-814.

[23] Romeu, Andrés and Vera-Hernandez, Marcos: "Counts with an Endogenous Binary Regressor: a series expansion approach," Econometrics Journal, forthcoming.

[24] Roth, Alvin E. and Sotomayor, Marilda, (1990): "Two-Sided Matchings Markets: A Study in Game-Theoretic Modeling and Analysis". Econometric Society Monograph. Cambridge, MA: Cambridge University Press.

[25] Roth, Alvin E. (1984), "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory," Journal of Political Economy, 92, pp991-1016.

[26] Roth, Alvin E. (1989), "A Natural Experiment in the Organization of Entry Level Labor Markets: Regional Markets for New Physicians and Surgeons in the U.K.," American Economic Review, vol. 81, June 1991, pp415-440

[27] Shahian, David M., Normand, Sharon-Lise, Torchiana, David F., Lewis, Stanley M., Pastore, John O., Kuntz, Richard E. and Dreyer, Paul I. (2001): "Cardiac Surgery Report Cards: Comprehensive Review and Statistical Critique," The Annals of Thoracic Surgery, Vol. 72, pp 2155-2168.

[28] Shen, Yujing (1999), "Selection and its impacts on Health Payment Systems," PhD dissertation, Boston University.

[29] ter Hofstede, Frenkel and Michel Wedel (1998): "A Monte Carlo Study of time-aggregation in continuous-time and discrete time parametric hazard models," Economics Letters, 58, pp149-156.

[30] Train, Kenneth, (2003): Discrete Choice Methods with Simulation, Cambridge University Press.

[31] Train, K., D. McFadden, and M. Ben-Akiva (1987), "The demand for local telephone service: A fully discrete model of residential calling patterns and service choice", Rand Journal of Economics 18, 109123.

7 Data Appendix

In this appendix we give some of the statistical evidence that guided some of the specification choices made. In particular the next table shows that most people choose a vacancy in the same region as their college. Each column represents a college location, so for example 79.05 percent of the students that graduated in Andalucian colleges choose a hospital located in Andalucia. The numbers in diagonal represent the percentage of physicians who choose a hospital in the same region as their college.

Table 3: Distribution of hospital location by region of college

College → / Hospital ↓	And	Ara	Can	CL	Cat	Val	Mad
And	**79.05**	1.52	1.92	1.94	0.79	1.34	1.47
Ara	0.14	**56.53**	0.24	1.37	1.48	0.45	0.65
Ast	0.10	0.38	0.48	4.68	0.23	0.26	0.34
Bal	0.20	2.15	1.20	0.65	2.22	1.21	0.38
Can	1.43	0.38	**77.94**	1.13	0.42	0.90	0.51
Cant	0.24	1.27	0.72	3.63	0.09	0.32	0.41
CM	1.50	1.77	0.24	1.61	0.09	2.75	5.16
CL	1.16	1.65	0.00	**46.77**	0.69	0.58	2.43
Cat	1.70	12.67	6.95	3.63	**89.37**	6.91	2.05
Val	1.36	3.17	2.16	2.18	1.39	**75.06**	1.06
Ext	1.50	0.13	0.00	1.21	0.18	0.26	0.34
Gal	0.34	0.38	0.24	2.18	0.51	0.19	0.82
Mad	9.33	11.15	6.71	26.21	2.03	5.56	**83.01**
Mur	1.36	0.38	0.96	0.40	0.23	3.58	0.48
Nav	0.48	4.56	0.00	1.05	0.14	0.38	0.65
PV	0.07	0.25	0.24	1.13	0.14	0.19	0.17
R	0.03	1.65	0.00	0.24	0.00	0.06	0.07
Total	100	100	100	100	100	100	100

Table 4: Distribution of hospital location by region of residence

Residence → / Hospital ↓	And	Ara	Can	CL	Cat	Val	Mad
And	**79.82**	0.761	2.37	1.58	0.60	1.17	1.93
Ara	0.10	**66.21**	0.22	1.97	1.49	0.49	0.67
Ast	0.14	0.00	0.43	5.13	0.09	0.00	0.27
Bal	0.27	0.30	1.08	0.16	0.88	0.80	0.30
Can	1.39	0.46	**76.34**	0.87	0.46	1.05	0.40
Cant	0.31	0.91	0.43	3.71	0.05	0.31	0.43
CM	1.36	1.37	0.22	1.03	0.05	1.73	4.10
CL	1.12	1.37	0.43	**48.26**	0.56	0.56	2.20
Cat	1.80	12.79	5.81	3.63	**92.11**	7.53	2.33
Val	1.19	2.28	1.72	2.13	0.70	**76.30**	0.97
Ext	1.46	0.15	0.00	0.87	0.14	0.25	0.30
Gal	0.37	0.00	0.00	2.76	0.32	0.25	0.67
Mad	8.61	9.74	8.17	24.80	1.86	4.75	**83.78**
Mur	1.29	0.15	1.08	0.24	0.23	3.70	0.37
Nav	0.71	2.89	0.86	1.74	0.37	0.86	1.00
PV	0.07	0.15	0.86	0.87	0.09	0.19	0.23
R	0.00	0.46	0.00	0.24	0.00	0.06	0.07
Total	100	100	100	100	100	100	100