HEDG Working Paper 10/09

The effect of waiting time and distance on hospital choice for English cataract patients

Peter Sivey

June 2010

york.ac.uk/res/herc/hedgwp

The effect of waiting time and distance on hospital choice for English cataract patients*

Peter Sivey
University of Melbourne

June 22, 2010

Abstract

This paper applies latent-class and multinomial logit models to the choice of hospital for cataract operations in the UK NHS. We concentrate on the effects of travel time and waiting time and especially on the waiting time elasticity of demand. Models including hospital fixed effects rely on changes over time in waiting time to indentify coefficients. We show how using latent-class multinomial logit models characterises the unobserved heterogeneity in GP practices' choice behaviour and affects the estimated waiting time elasticities of demand. The models estimate waiting time elasticities of demand of approximately -0.1, comparable with previous waiting time-demand models. For the average waiting time elasticity, the simple multinomial logit models are good approximations of the latent-class logit results.

Key words: hospital choice, waiting time, latent class model.

JEL Classification: I11

*Email address: psivey@unimelb.edu.au. This research was funded by a PhD scholarship from the Centre for Health Economics, University of York. I would like to acknowledge advice and support from my thesis supervisor Hugh Gravelle. I have also benefited from helpful comments from Arne Hole, Michael Damiani, Mark Dusheiko, Andrew Jones, Luigi Siciliani and Steve Morris, and from seminar participants in the Health, Econometrics and Data Group, University of York, and Microeconometrics Unit, University of Melbourne. Any errors remain my own.

1 Introduction

Recent policy developments in the UK NHS emphasise patient choice of hospital for elective procedures (Coombes, 2006). Recent literature has sought to answer whether choice and competition reduces waiting times both theoretically (Brekke et al., 2008) and empirically (Dawson et al., 2007). Policymakers may be interested in how changes in waiting time will affect demand at a hospital due to substitution from alternative hospitals. In this paper we answer the question: how much do waiting times affect choice of hospital? We examine the trade-off between travel time and waiting time and estimate the waiting time elasticity of demand. To achieve this we apply latent-class multinomial logit models to the choice of hospital for cataract operation in the UK NHS.

Our approach continues recent developments in extending hospital choice models to using mixed logit methodology (Tay, 2003; Howard, 2005; Goldman and Romley, 2008). It contrasts with existing literature set in the US system where rationing is on price. We examine hospital choice in the UK National Health Service where elective care is rationed by waiting time. Our identification strategy uses hospital fixed effects and changes over time in hospital waiting time observed in the UK in the early 2000's. Propper et al. (2008) has evaluated the policy of 'targets and terror' which coupled incrementally reducing waiting time targets with punishments for hospitals who failed to meet them. Waiting time targets for elective admissions fell from 18 months in 2000 to 6 months by 2004 giving our data substantial variation over time in individual hospital's waiting time.

We estimate waiting time elasticities to compare our approach to the hospital waiting time demand literature (Martin and Smith, 1999; Windmeijer et al., 2005; Martin et al., 2007). We find comparable waiting time elasticities, around -0.1 for inpatient and outpatient waiting time. Our approach emphasises substitution between NHS hospitals whereas the previous literature emphasises substitution between NHS treatment and private treatment or no treatment. We show how using latent-class logit models characterises the unobserved heterogeneity in choice behaviour.

1.1 Background

This paper is related to the new emphasis on 'patient choice' in UK healthcare policy over the last five years. A new computer system for referrals for elective operations from general practice was put into place in 2006 to explicitly offer NHS patients the choice of at least 4 different hospitals (Coombes, 2006). This policy appears to have two motivations. Firstly to combat high and variable waiting times for many elective operations in the NHS (Dawson et al., 2007), and secondly to encourage hospital competition on quality (Propper et al., 2004).

Our data includes elective admissions for cataract surgery for financial years 2001/2 to 2003/4 at the General Practitioner (GP) level. Prior to 2006 GPs were solely responsible for choosing hospitals on patient's behalf. Our results give some policy relevant insights about GP choice behaviour before the recent policy changes.

We analyse choice for cataract operations because they are one of the most common elective procedures and have significant waiting times. In the dataset, hospitals perform an average of nearly 1000 cataract operations per year and the average waiting time is around six months.

We assume for the analysis that GPs had free choice of hospital when referring patients for cataract operations. There is evidence from qualitative surveys that GPs in the UK have exercised their choice to refer patients to hospitals with lower waiting time and higher quality prior to the 2006 reform (Mahon et al., 1993; Whynes and Reed, 1994; Earwicker and Whynes, 1998). This paper provides quantitative evidence based on observed GP choices.

Our approach is related to previous papers which estimate the relationship between waiting times and demand for hospital care in the UK (Martin and Smith, 1999; Gravelle et al., 2002; Windmeijer et al., 2005; Martin et al., 2007). They estimate a more aggregate model of overall demand than in this paper, using data at the hospital or small area level. They examine the choice between having an NHS operation and going private or delaying/foregoing an operation. They do not consider explicitly choice between NHS hospitals. This paper contributes to the literature by taking a different approach: substitution *between* NHS hospitals based on trade-offs between distance, waiting times and other attributes.

A literature developed in the early 1990's using multinomial logit models to study hospital choice (Luft et al., 1990; Burns and Wholey, 1992; Hodgkin, 1996). These models focused on hospital quality characteristics such as mortality and readmission rates. Hodgkin (1996) is particularly relevant for this paper as this study is the first to use hospital fixed effects and uses changes over time in quality to identify coefficients. We also use fixed effects in this way but uses changes over time in *waiting time.* Recent literature has used mixed logit models to analyse hospital choice for Acute Myocardial Infarction patients (Tay, 2003), for kidney transplants (Howard, 2005), and pneumonia patients (Goldman and Romley, 2008) in the US, for a variety of outpatient services in rural India (Borah, 2006) and China (Qian et al., 2009). These papers have focused on hospital quality and/or price as the determinants of choice. Mixed logit models relax some restrictive assumptions of the multinomial logit models and can capture unobserved heterogeneity in the coefficients of the hospital attributes. This paper contributes by using a latent-class logit methodology in a different setting, where waiting times are the focus of the analysis.

Papers in the transport and environmental economics literature have compared the use of latent class logit models to mixed logit models with continuous distributions and found that each approach has its own merits (Greene and Hensher, 2003) and can produce very similar results (Hynes et al., 2008). Latent class models have the benefit of not imposing specific distributional assumptions on coefficients. They have been widely used in health economics in studies of healthcare utilisation (Deb and Trivedi, 2002; Bago d'Uva, 2006) and have also been applied in analysing discrete choice experiments (Hole, 2008).

A literature is also developing on waiting time as a determinant of hospital choice (Bessho, 2003; Burge et al., 2004; Kjerstad, 2006; Monstad et al., 2006). A UK study evaluating the London Patient Choice Project (LPCP) used a discrete choice experiment alongside some revealed preference data from the London area to analyse the trade off between distance, waiting time and quality (Burge et al., 2004). A Norwegian study (Monstad et al., 2006) uses revealed preference data on hip replacements in Norway. It takes a similar approach to this paper in using hospital fixed effects to control for unobserved quality at each hospital (Hodgkin, 1996), and emphasises the trade-off between waiting times and distance. Patient heterogeneity is taken into account using observed patient characteristics (age, gender education) as interactions in a simple MNL model rather than exploring unobserved heterogeneity in a mixed or latent-class logit model.

Another attempt to study waiting times and hospital choice has used a 'hospital bypassing' model (Varkevisser and van der Geest, 2007). This study shows that lower-than-average waiting time reduces the probability of bypassing a hospital by between 2% and 10%. However this study is limited by only using a binary choice model. The same authors use a conditional logit model to measure hospital competition in a hospital market with no prices (Varkevisser et al., 2010). Although their model controls for the effects of waiting time, they concentrate on travel time elasticities of demand. In this paper we focus on waiting time elasticities.

2 Data

The main data source used is NHS Hospital Episode Statistics (HES). This is an administrative dataset containing records of all inpatient admissions to English hospitals. The analysis in this paper uses HES data for three financial years (2001/2 to 2003/4) for cataract operations (HRG B02). Only elective spells are considered, 'booked' and 'waiting list' spells are included and ‘planned’ spells are excluded. We only include elective patients because we know elective patients are referred by their GP and may plausibly choose their hospital, whereas non-elective patients may choose their hospital very differently. Also, electives are the main group of patients who have to wait for care and we estimate the effect of waiting time on choice. We exclude planned elective spells because for these spells the waiting time before an operation is for clinical reasons not due to lack of capacity (HES Online, 2007).

Patients recorded as being treated privately (although in NHS hospitals) and episodes that are not the first of the spell are also excluded. To make the analysis more manageable we restrict the geographical area covered to include referrals from GP practices located within the North West region of England. This area is currently covered by one strategic health authority (NHS North West) but previously was three separate strategic health authorities. This area has a population of nearly 7 million people, and the estimation sample includes 1237 GP practices in the area which make 87161 referrals for cataract operations in the time period specified.

Two types of waiting time affect cataract patients. The *outpatient* waiting time is the wait between the referral from the GP and the outpatient appointment with the specialist. The *inpatient* waiting time is the wait from the outpatient appointment (and decision to admit the patient) and the actual date of the operation.

Although HES includes it, we do not use the *actual* inpatient waiting time for each patient at the hospital they visit. The reason is that we do not know the time that the patient may have waited if they visited a different hospital in the choice set, and we must include a waiting time for each hospital in the choice set. Instead, we create a measure of inpatient waiting time for each hospital site, for each year which is the same for every GP practice in the dataset. We calculate the median of the inpatient waiting time for patients discharged during each year at each hospital. The median is used as opposed to the mean to reduce the effect of outliers which represent measurement or coding error.

Outpatient waiting time is not observed in Hospital Episode Statistics for the time period of these data. Instead we use Department of Health waiting time statistics (Department of Health, 2008) which are available quarterly at hospital *trust* level. These data provide information on the number of patients waiting certain time intervals for an outpatient appointment. The time intervals are 0-4 weeks, 4-16 weeks, 16-25 weeks and over 25 weeks.

We create a measure of outpatient waiting time for each hospital trust, for each quarter by using the mid-points of each interval. This is the approach used in Martin and Smith (2007). We aggregate this to yearly level by taking an average over the four quarters in each year at each trust. The average inpatient waiting time is approximately 6 months and so the average inpatient appointment will be 6 months after the average outpatient appointment. For this reason we use the outpatient waiting time data lagged by two quarters to match it approximately with the inpatient waiting time data.

The overall hospital set for each year is selected by including every hospital receiving at least 30 referrals for cataract operations from GP practices within the North West region. This set of hospitals varies slightly from year-to-year as some hospitals cross the 30-referral threshold. There are 30 hospitals in total which appear in the overall hospital set.

The choice set for each GP practice in each year is assumed to be the nearest 10 hospitals in the overall hospital set. Howard (2005) also uses the nearest 10 sites in his analysis. Any referrals from practices to hospitals not within the nearest 10 are disregarded. We conduct sensitivity analysis of this assumption.

The HES data is linked to a number of other sources to provide additional information. Data from the NHS National Administrative Codes Service (NACS) is used to provide names and addresses (postcodes) for the hospital sites and trusts. NHS trusts are the managing organizations of NHS hospital sites.

The use of data on site of treatment in this paper is an improvement on previous studies using trust-level data (Damiani et al., 2005; Martin et al., 2007). Some site codes changed during the time period studied, meaning the site codes had to be checked with archives of NACS data. In addition, some site codes were missing, or failed to identify a site, identifying only the NHS trust. In this case, some further investigation was done using the NHS website and Binleys directory to check which sites in the trust carried out cataract operations. We also contacted the trust directly in some cases to verify information about hospital sites.

The GP practice codes in HES are linked to postcodes in the October 2003 General Medical Services (GMS) census of GP practices. A number of other GP practice characteristics are also obtained from this GMS census data and included in the models to capture observable heterogeneity in the choice behaviour of practices. These include: the proportion of practices' patients who live in rural areas, the number of full-time equivalent (FTE) GPs in the practice, whether the practice is involved in training new GPs, the average number of patients assigned to each GP (list size) in the practice. The variable measuring the proportion of practices' patients living in rural areas shows a significant proportion of zeros and when positive, is often quite small. For use in the estimations, a dichotomous variable was created measuring if the 'rural proportion' of each practice was positive (=1) or not (=0). GP practice socioeconomic information is obtained by linking the practice Low Income Scheme Index (LISI) score for 2004/5. This is a measure of deprivation based on exemptions from prescription charges on the grounds of low income.

Travel times were calculated between the postcodes of GP practices and hospitals. The postcodes were first converted to Ordnance Survey Eastings and Northings using Royal Mail Postzon data which defines the centre-point of each postcode. Eastings and Northings were then converted into Latitudes and Longitudes using an Ordnance survey spreadsheet (Ordnance Survey, 2006). Travel times were then calculated for every pairwise combination of practice and hospital coordinates using Microsoft MapPoint. This calculation was automated using a Visual Basic program within Microsoft Excel[1]. MapPoint calculated travel times between practice and hospital coordinates assuming European-averaged speed limits for different types of road (motorways: 121kph, major roads: 64kph, minor roads: 55kph, streets: 31kph).

[1] I used Microsoft Excel spreadsheets with macros in Visual Basic to automate the process of converting the easting/northing coodinates to latitude/longitude and the process of calculating the travel times in Map Point. Both spreadsheets were kindly provided by Michael Damiani.

The number of observations in the original HES dataset is 93,365, the estimation sample contains 87,128 observations. Observations are lost when GP practice codes are not found in the GMS data, GP practice postcodes from the GMS data are not found in MapPoint or in Royal Mail Postzon data and when GPs refer patients to hospitals that receive less than 10 referrals per year from practices in the north-west region.

3 Econometric model

3.1 Multinomial logit (MNL) framework

A random utility choice model (McFadden, 1974) underpins the empirical analysis. There are n GP practices denoted by i. The subscript k denotes patients, and K_i patients are admitted to hospitals from practice i. GPs and patients choose hospitals, denoted by j, from the choice set of the practice H_i. We assume, in this paper, the observable portion of utility u varies only by GP practice and hospital, and not by patient k. The additive unobserved component of utility ε_{kij} varies by patient as well as by GP practice:

$$U_{kij} = u_{ij} + \varepsilon_{kij}; k = 1, .., K_i;\ \ i = 1, ..., n;\ \ j \in H_i; \tag{1}$$

GPs choose a hospital for each patient to maximise U_{kij}.We label the hospital chosen by practice i for patient k as Y_{ki} :

$$Y_{ki} = \arg\max_j (U_{kij}; j \in H_i) \tag{2}$$

We do not include an 'outside option', the choice to go private or forego treatment. We assume U_{kij} is sufficiently high at one of the hospitals in the choice set H_i for the GP to choose a hospital and not deny the patient treatment or recommend private treatment. We assume therefore, a fixed overall demand for NHS treatment, with the model coefficients determining the choice between NHS hospitals. This can be thought of as a model for the second stage in a two-stage decision process, where the patient and GP first decide whether the patient should be treated in the NHS and secondly where they should be treated.

Where the ε_{kij} are independently and identically distributed (i.i.d.) according to the type 1 extreme value distribution, the choice probability for patient k choosing hospital j is:

$$\Pr(Y_{ki} = j) = \Pr(U_{kij} > U_{kil}) = \frac{\exp(u_{ij})}{\sum\limits_{s \in H_i} \exp(u_{is})}, \ \forall l \in H \tag{3}$$

We assume the observable portion of utility is a linear function of variables and coefficients, $u_{ij} = x_{ij}\beta$. The term x_{ij} represents a vector of characteristics of hospital j some of which may also vary by GP practice, i (eg. travel time). The vector may include interactions between GP practice characteristics and characteristics of the alternative hospitals.

We sum the log of choice probabilities over practices, hospitals and patients to form the log-likelihood:

$$\ln L(\beta) = \sum_{i=1}^{n} \sum_{j \in H} \sum_{k=1}^{K_i} d_{ijk} \ln \left[\frac{\exp(x_{ij}\beta)}{\sum\limits_{s \in H} \exp(x_{is}\beta)} \right] \tag{4}$$

where $d_{ijk}$is an indicator variable equal to one if hospital j is chosen for patient k and equal to zero otherwise. The logit formula, $\frac{\exp(x_{ij}\beta)}{\sum\limits_{s\in H}\exp(x_{is}\beta)}$, in equation (4) does not contain any k subscript, as we assume u only varies by i and j, so we can aggregate the formula up to the practice level. We define the admissions from practice i to hospital j as $\sum_{k=1}^{K_i} d_{ijk} = adm_{ij}$ where there are K_i admissions from practice i.

$$\ln L(\beta) = \sum_{i=1}^{n}\sum_{j\in H} adm_{ij} \ln\left[\frac{\exp(x_{ij}\beta)}{\sum\limits_{s\in H}\exp(x_{is}\beta)}\right] \tag{5}$$

In (5), the natural log of the choice probability is weighted by the number of admissions from practice i to hospital j (adm_{ij}). The log-likelihood can be maximised numerically given data on attributes of hospitals, x_{ij}, and admissions at each hospital from each practice, adm_{ij}, to give estimates of the coefficients of interest β.

3.2 Latent-Class Multinomial Logit (LCMNL)

The latent-class multinomial logit model extends the MNL model to allow for heterogeneity in the vector of coefficients β. Each GP practice is assumed to have a coefficient vector β_c falling into one of C classes. Some elements of β may be specified to be common across classes. The LCMNL model nests the simple MNL model.

The log-likelihood for the Latent Class model (LCM) with C classes is:

$$\ln L(\beta_1, \pi_1, ..,\beta_C, \pi_C) = \sum_{i=1}^{n} \sum_{c=1}^{C} \pi_c \sum_{j \in H} adm_{ij} \ln \left[\frac{\exp(x_{ij}\beta_c)}{\sum_{s \in H} \exp(x_{is}\beta_c)} \right]; \qquad (6)$$

where π_C is the probability that patients in practice i are in class C.

The IIA restriction mentioned in section 3.1 does not apply to LCMNL models in general. For a given GP practice we still have a fixed coefficient vector β_c, implying IIA. However, across GP practices IIA does not apply because there is correlation of choice probabilities which is not due only to the level of the attributes x_{ij} but also caused by the probability distribution of the coefficients across the C classes. Latent class models can approximate the true distribution of coefficients without imposing specific distributional assumptions on coefficients (Greene and Hensher, 2003).

We aggregate data to the GP practice level rather than using the individual patient-level data in order to better identify the latent-class logit models. An intuitive reason to aggregate data to the practice level is that preferences about the choice of hospitals may be driven by GPs rather than by patients, as in the time period of our data GPs actually make the choice. Where patients' preferences do influence choice, the GP coefficient vector β_c represents the average preferences of GPs in class c's patients.

3.3 Assumptions of the empirical model

It is standard in the hospital choice literature to assume that the data available represent the only choices available to the patient or doctor making the choice. There are two main 'outside options' that we are omitting from the analysis: (1) choosing to delay or forego the operation and (2) choosing to have the operation outside the NHS. Our data does not allow us to include these two choices.

In this analysis we know only that the GP 'officially' makes the choice of hospital. However, we assume the patients preferences will be reflected in the choice of hospital even if the GP has a role in making the choice. We can reasonably assume the GP works as an agent to the patient (McGuire, 2000). The GP may not act as a perfect agent but during the time period of the data we study (2001/2 to 2003/4) the GP has no incentive to contradict the preferences of the patient. For example, GPs in this time period were not budget-constrained (as in the 'fundholding' system), where a budget constraint for the GP may affect choice of hospital if the prices varied between hospitals.

3.4 Empirical Specification

We now specify the deterministic portion of the utility function ($u_{ij} = x_{ij}\beta$) in equation (1). Allow the characteristics x_{ij} and outcome Y_i to vary also by time period so they become x_{ijt} and Y_{it}. Suppose that the vector of characteristics x_{ijt} are composed of two sub-vectors (x_{ijt}^1, x_j^2), where x_j^2 are unobserved characteristics of hospital j that are time-invariant. For example, we consider hospital quality will be largely time invariant in the time period of our dataset.

Assuming all the elements of x_{ijt} vary either by practice (i) or over time

(t) as well as between hospitals (j), it is possible to specify and identify a hospital-specific fixed effect ($\lambda_j = x_j\beta_2$) for each hospital in addition to the identification of β_1.

Now (1) becomes:

$$
\begin{aligned}
U_{ijt} &= x_{ijt}^1\beta^1 + x_j^2\beta^2 + \varepsilon_{ijt}; && (7) \\
&= x_{ijt}^1\beta^1 + \lambda_j + \varepsilon_{ijt} && (8)
\end{aligned}
$$

The vector x_{ijt}^1 is specified as follows: $x_{ijt}^1 = (ln(traveltime_{ij}), wait_{jt})\prime$. $ln(traveltime_{ij})$ is the natural log of the measure of travel time from practice i to hospital j and $wait_{jt}$ is the waiting time at hospital j in period t. We use the natural log of travel time to allow a nonlinear relationship between utility and travel time. The effect on utility of absolute changes in travel time falls with travel time whereas the effect of proportional changes remains constant.

The waiting time coefficient is identified differently in model (7), without hospital fixed effects compared with model (8), with hospital fixed effects. In model (7) the coefficient is estimated from cross-sectional differences between hospitals' waiting times *as well as* changes over time in hospitals' waiting times. In model (8), the coefficient is *only* estimated using differences in changes over time in waiting times between hospitals. This is because the hospital fixed effects λ_j, captures all time-invariant, cross sectional hospital attributes, including time-invariant, cross-sectional differences in waiting time. In this way, model (8) is comparable to a fixed-effects panel data estimator allowing robust estimation of time-varying variables but leaving time-invariant effects to be picked up by a fixed effect.

The models estimated with hospital fixed effects (8) therefore crucially rely on variation over time in waiting time between hospitals. In this case hospital waiting times are generally falling over time due to the 'targets and terror' regime of progressively reducing waiting time targets (Propper et al., 2008). This create changes over time in hospital waiting times that differ by hospital because some hospitals will not need to reduce waiting times to meet the target, some hospitals may not be able to reduce waiting times sufficiently, and others may 'overshoot' the target and reduce waiting times more than necessary.

In order to estimate the model we assume the x^1_{ijt} are exogenous to every decision-maker, here the GP practice. This seems a plausible assumption given that GP practices are relatively small and numerous compared to hospitals. For example, it seems unlikely that an individual practice's choice of hospital would have an impact on a given hospital's waiting time.

Other possible determinants of the choice of hospital, such as measures of hospital quality, are omitted as they are generally time-invariant in the time period of the dataset used. The approach of this paper is to sacrifice the estimation of coefficients for hospital quality measures to improve the validity of the estimates of the travel time and waiting time coefficients with the use of hospital fixed effects (λ_j), which can control for all time invariant hospital characteristics.

We estimate LCMNL models with the $ln(traveltime)$ and $wait$ coefficients allowed to vary within classes. MNL models are estimated with practice characteristics as interactions with the $ln(traveltime)$ and $wait$ coefficients to capture observable heterogeneity in the coefficients.

We first estimate Model 1, a simple choice model with only waiting time and distance as attributes. This is the model in equation (7) but without any x_j^2 variables. Model 2 adds hospital fixed effects and allows us to see the effect of controlling for unobserved time-invariant hospital characteristics. This model is described by equation (8). Model 3 adds GP practice characteristics as interaction terms to allow for observable heterogeneity in coefficients at the practice level.

Models 4 and 5 are latent-class models, assuming a discrete distribution for practice unobserved heterogeneity. Model 5 includes a single practice characteristic - practice rurality - as a covariate determining the class probabilities (π's in equation (6)). We include this characteristic as it emerges from Model 3 as the most important practice characteristic.

The models are estimated in STATA 10 using the 'clogit' command for the MNL models and the 'gllamm' command (Rabe-Hesketh et al., 2002) for the latent-class models.

4 Results and discussion

Table 1 presents the descriptive statistics. The average hospital in the data performs 1037 cataract operations a year and the average practice admits 24 patients per year, showing that this is a common procedure. There are 30 hospitals and 1247 practices with each practice referring patients to 1.61 different hospitals on average

[INSERT TABLE 1 ABOUT HERE]

Two groups of statistics are presented for travel time: the travel time to the nearest hospital, and to the hospital actually visited. These are averaged over all admissions for the three years. The two means demonstrate that patients do not always visit the closest hospital to their practice. We can also see there is a much higher variation in the travel time to the hospital visited compared to the travel time to the closest hospital.

For waiting time we present an average of the median annual inpatient waiting time and mean annual outpatient waiting time at each hospital for each of the three years (NB: not all hospitals appear in all three years).

Two differences emerge between inpatient and outpatient waiting times. Firstly outpatient waiting time are much shorter, less than half the length of inpatient waits; secondly, outpatient waiting times have a smaller standard deviation, roughly a quarter the standard deviation of inpatient waiting time. These differences will influence the interpretation of coefficients and elasticities. For instance, we might interpret a large elasticity of waiting time to be less important if the standard deviation is small.

We also decompose the variation in waiting time into 'between' and 'within' variation. Crucially for the identification of models with hospital-fixed effects, there is substantial 'within' variation for both outpatient and inpatient waits (around 40-50% of the overall variation). This shows waiting times at each hospital are changing over the three years of data

Table 2 presents the descriptive statistics for the GP practice characteristics. There are some missing observations for some of the variables which affects the sample size in the model with interactions (Model 3). The units of the final three characteristics are transformed to give their coefficients a similar scale. Patient age is available for every observation (not just at practice level).

[INSERT TABLE 2 ABOUT HERE]

There is substantial variation in practice size with a standard deviation of nearly 2 FTE GPs around the mean of 2.88. Minorities of practice are training practices (22%) or are situated in rural areas (14%). The average patient is 75 years old.

The following tables of results present the Akaike and Bayesian Information criteria (AIC=$2k - 2lnL$, k =no. of parameters; BIC=$-2\ln L + k\ln(n)$, n =no. of observations) to allow some comparisons between models. These criteria reward models for having a high log-likelihood but penalise them for having many parameters. The lower the value of the AIC and BIC, the better.

Presenting the results of estimations, we indicate statistical significance in the conventional manner: *** indicates significance at the 1% level, ** indicates significance at the 5% level, and * indicates significance at the 10% level.

Table 3 presents results from the MNL models. Model 1 includes only the travel time and waiting time variables while Model 2 includes the 29 hospital fixed effects. Both models show a statistically significant negative effect of the travel time variable, indicating GPs prefer to refer patients to nearby hospitals. The size of the coefficient becomes slightly larger in Model 2, this may be due to correlations between unobserved time-invariant hospital specific attributes and travel time.

[INSERT TABLE 3 ABOUT HERE]

The waiting time coefficients are positive in Model 1 and negative in Model 2, suggesting the inclusion of hospital-specific fixed effects influences the estimated effect of waiting time. Waiting time may be positively correlated with time-invariant attributes of hospitals that attract patients (x_j^2 in (7)) biasing the waiting time coefficients in Model 1. Time-invariant attributes may include hospital quality. Here we do not model hospital quality explicitly, but simply control for all time invariant attributes using hospital fixed effects (in Models 2 to 5).

In Model 2, the waiting time coefficient is identified only by changes over time in hospital waiting times rather than cross-sectional variation between hospitals (see section 3.4).

Table 3 also shows results for Model 3, a simple MNL model where the travel time and waiting time variables are interacted with GP practice characteristics. The interactions are coded as differences from the sample means so the coefficients on travel time and waiting time can be directly compared to Models 1 and 2. The rural practice indicator has the most notable interaction effect, reducing the travel time coefficient and reversing the sign of both waiting time coefficients. Rural practices appear to value proximity less than urban practices. Patients in rural practices may value proximity less as they are used to travelling long distances to access facilities, and they their preferences for (lack) of proximity to amenities may have influenced their choice of location.

Among the other interaction terms, practices with more FTE GPs, higher list size per GP and more low-income patients (measured by the LISI), value proximity less. Practices with older patients value outpatient waiting time more.

Tables 4 and 5 present results for latent class models. We present results for models with only two classes. Attempts to estimate models with three or more classes resulted in one of the classes having a very small probability (less than 0.05) and large standard errors. Following Hole (2008) we take a pragmatic approach and select the more parsimonious and well-identified model with fewer classes. The estimates suggest one group of practices (Class 1) composing roughly 70% of practices in the sample with a high travel time coefficient and waiting time coefficients that are positive or close to zero. The second class of practices (Class 2) have a smaller travel time coefficient and larger (negative) waiting time coefficients.

[INSERT TABLES 4 AND 5 ABOUT HERE]

The results imply (1) substantial heterogeneity in coefficients between practices and (2) a negative correlation between the distribution of the travel time and waiting time coefficients across practices. We interpret the first class of practices to represent GPs who have strong preference for proximity but little or no preference over waiting time, and the second class of practices to represent GPs who have less preference for proximity and a stronger preference for lower waiting time.

Model 5 presents the same latent class model as Model 4 but where the class probabilities are a function of a practice characteristic: the rural practice indicator. We choose only one practice characteristic for estimation time to be manageable. We choose the rural practice indicator as this emerges from Model 3 as the most important interaction term, affecting all three attribute coefficients.

The results show a similar pattern to Model 4 with regard to the class probabilities and the negative correlation between the travel time and waiting time coefficients. The travel time and waiting time coefficients are also very similar in both models across the two classes.

The rural practice indicator increases the probability of a practice belonging to class 2, the class that has less preference for low travel time and more preference for low waiting time. The interaction effect of rurality in this model agrees with model 3 for travel time, but not for waiting time.

Models 4 and 5 suggest significant unobserved coefficient heterogeneity at the practice level, evidenced by significant improvements in the Akaike and Bayesian information criteria in these models compared to the simple MNLs. The rationale for this heterogeneity is twofold. Firstly, GPs act as agents for patients in choosing the hospital, and patient characteristics will vary between practices. Secondly, GPs are also influenced by their own preferences and characteristics which will influence the choice of hospital.

4.1 Waiting time elasticity estimates

Now we present estimates of the effect of a change in waiting time on hospital demand (the number of admissions) in the form of waiting time elasticities of demand. We calculate elasticities at the individual (admission) level and summarise the results with average elasticities and plots of the distribution of elasticities.

We define a probability weighted elasticity (similar to Howard, 2005) for each admission:

$$\widehat{E} = \sum_{j \in H} 100 * \frac{\widehat{P}_{j1} - \widehat{P}_{j0}}{\widehat{P}_{j0}} \widehat{P}_{j0} = \sum_{j \in H} 100 * \left[\widehat{P}_{j1} - \widehat{P}_{j0} \right] \tag{9}$$

where $\widehat{P}_{j0}$ is the predicted probability of choosing hospital j from the practice's choice set, and $\widehat{P}_{j1}$ is the same predicted probability when hospital j has a 1% higher waiting time. Table 6 presents the average of $\widehat{E}$ across all admissions for models 2, 3, 4 and 5.

The elasticities can be interpreted as the % change in demand (admissions) associated with a 1% change in waiting time. We can interpret the mean elasticities as the value of $\widehat{E}$ averaged across all hospitals where each hospital is weighted by the number of admissions it receives. The mean elasticities for all four models are close to -0.1. The mean and standard deviation of hospital waiting times are approximately 6 months and 3 months. Consider a hospital with a 6 month inpatient waiting time which increases by 3 months (50%), then an elasticity of -0.1 predicts admissions in that hospital will fall by 5%. If we think that outpatient and inpatient waiting time will both rise by 50 % (the former on average from two to three months) then hospital admissions would fall by 10%. In contrast, if as inpatient waiting time rises the outpatient waiting time falls by the same proportion then there will be no overall effect on admissions as the two waiting time effects will cancel each other out.

The inpatient and outpatient waiting time elasticities are similar despite the inpatient coefficient being substantially smaller than the outpatient coefficient in all of the models. The mean elasticities are much more consistent between outpatient and inpatient waiting time and across the different models in comparison with the coefficients, for which the outpatient and inpatient waiting times have more varied results. Comparing across different models, the elasticities are very similar despite the differences in the estimated coefficients.

The waiting time elasticity estimates are similar to those in the UK hospital demand literature, Martin et al. (2007) estimates elasticities in the range -0.07 to -0.24 for inpatient routine surgical procedures and ENT procedures. Our results are of a similar magnitude, however we must take care in comparing our results with the hospital demand literature due to the differences in the type of demand being estimated. The other papers mentioned estimate the aggregate demand for hospital admissions in general rather than modelling individual GP practice demand as a result of substitution between NHS hospitals.

Figures 1 to 4 plot the distribution of the waiting time elasticities of demand, $\widehat{E}$, across all admissions for Models 3 and 5, the MNL model including interactions, and the latent-class model with the rural indicator influencing class probabilities.

[INSERT FIGURES 1 TO 4 ABOUT HERE]

The elasticity distibutions share one feature: a peak of the distribution close to zero. This suggests there are a significant proportion of GP practices who are not responsive to hospital waiting time in choosing hospitals. All of the distributions also have a left tail indicating the practices who do take waiting time into account when choosing vary in the weight they put on waiting time. Figures 2 and 4 shows evidence of a bimodal distribution of waiting time elasticities influenced by the two latent classes of coefficients.

The average elasticities presented in Table 6, summarising the demand effect of changes in hospital waiting time, should be interpreted in light of these insights. An average elasticity of -0.1, for example is likely to be made up of a large number of GP practices with an elasticity of zero, with some GP practices with much higher elasticities of nearer -0.2 to -0.3.

Note that Figures 1 and 3 show a small minority of practices with positive waiting time elasticities. As Figures 1 and 3 relate to Model 3, these represent the minority of GPs whose characteristics give them a positive waiting time coefficient, including rural GPs (see Table 3). The model behind Figures 2 and 4 (Model 5) predicts no positive waiting time coefficients.

4.1.1 Comparison with Travel Time Elasticities

In this paper we concentrate on waiting time elasticities of demand as they are a policy focus in the NHS and act to clear the market for elective care services (Lindsay and Feigenbaum, 1984; Martin and Smith, 1999). Travel time elasticities are not as intuitive because the travel times between individual hospitals and populations of patients do not change. However as the results suggest travel time is much more important as a determinant of hospital choice than waiting time, it is interesting to compare the average waiting time elasticities with travel time elasticities.

We calculate travel time elasticities of demand in the same way as for waiting time, as summarised in equation (9), except for with travel time we use a 0.01 increase in $ln(traveltime)$ to simulate a 1% increase in travel time. Results are reported in the third column of Table 6. The travel time elasticities for this model have a mean of between -1.4 and -1.5 across all four models. Varkevisser et al (2010) calculate travel time elasticities for Dutch neurosurgical outpatient visits and estimate elasticities between -1.4 and -2.6. Our results are remarkably similar, given the difference in institutional arrangements and type of surgery

This result confirms that our models predict travel time has a much higher weight as a determinant of choice than waiting time. The travel time elasticity is approximately 10 to 15 times the size of the waiting time elasticity of demand. However we cannot readily interpret the travel time elasticity in the same way as we can the waiting time elasticity.

4.2 Robustness Checks

Table 7 presents Model R1 an estimate of Model 2, the MNL model with hospital fixed effects, where the choice set includes every hospital in the hospital set, not just the nearest 10 hospitals.

The three coefficients of interest are very similar to those in Model 2, when the choice set is restricted to the nearest 10 hospitals. The only notable difference is that the outpatient waiting time coefficient is larger in Model R1.

Whereas the models discussed in the main body of the paper include the nearest 10 hospitals and this robustness check included the 30 hospitals in the north-west region, theoretically we could include every hospital in England doing cataract operations in the analysis. Intuitively we might expect that including more far away hospitals in the choice set would alter the results, perhaps increasing the size of the travel time coefficient, especially if far-away hospitals are rarely chosen. However, we can see from the results of this robustness check that excluding far-away hospitals has little impact on the results as it excludes only largely irrelevant choices. As the results, especially the travel time coefficient, do not change much when increasing the choice set from 10 hospitals to 30, it seems unlikely that increasing the choice set further would have much impact.

We might suspect the strong effect of travel time is due to a default choice of the 'nearest hospital' for many GPs. Table 7 also presents model R2, an estimate of Model 2 including a dummy variable indicating the nearest hospital. The Closest Hospital variable is estimated as a positive and significant determinant of hospital choice. The Ln(travel time) coefficient falls, but only by 13%, from -4.91 to -4.23. Travel time is still a strong determinant of hospital choice after controlling for effect of the nearest hospital.

5 Conclusions

In this paper we apply latent-class multinomial logit models to the choice of hospital for cataract operation in the UK NHS. We concentrate on the trade-off between travel time and waiting time and calculate the waiting time elasticity of demand. The results show that travel time has a much stronger effect on the probability of hospital choice than waiting time.

Estimated waiting time elasticities are similar to those in the existing waiting time-demand literature, approximately -0.1. The results imply that the demand side of the hospital market reacts to differential waiting times, or changes in waiting times, between alternative hospitals. Fundamentally this implies waiting time acts (to some extent) as a 'price' of care for patients and rations the demand for care between alternative hospitals. From a policy perspective this result has implications for government regulation of waiting times: for example targets or maximum waiting times. Where waiting time clears the market for an elective procedure (eg cataract surgery) and different hospitals have different waiting times, forcing all hospitals to have the same waiting time (eg a maximum of 6 months), could results in an excess demand at some hospitals

We show how using latent-class logit models characterises the unobserved heterogeneity in GP practices choice behaviour. We find two distinct classes of GP practices, one of which is more reactive to waiting time, and values proximity less, than the other.

The estimated distributions of elasticities from the models show evidence for two groups of GPs: (1) GPs whose choice of hospital for their patients is not influenced at all by waiting time, (2) GPs whose choice of hospital for their patients is substantially influenced by waiting time, with elasticities ranging from approximately -0.1 to -0.5. Policies to encourage patient choice of provider based on waiting time may have most impact in the former group of GP practices which previously did not engage in choice.

The results also imply that rural GP practices are different, and are less likely to choose on waiting time. This result gives a different insight: in rural areas choice of hospital may be much less likely due to the lack of nearby hospitals. Any choice policies must recognise that patients in rural areas have more limited opportunities to choose between hospitals. For example, patients in rural areas could be offered subsidised transport to alternative hospitals.

Future research could develop a full model of waiting time and demand at individual hospitals allowing for substitution between different hospitals, as in this paper, and for the effect of waiting time on overall demand, as in the previous literature (Martin and Smith 1999, Martin et al 2007). Such a comprehensive model would allow accurate estimation of the overall effect of waiting time changes on demand, informing policies related to waiting time targets.

References

Bago d'Uva, T. (2006). Latent class models for utilisation of health care, *Health Economics* **15**(4): 329–343.

Bessho, S.-i. (2003). What are the determinants of hospital choice? Time, costs, option demands and communication, mimeo, University of Tokyo.

Borah, B. J. (2006). A mixed logit model of health care provider choice: Analysis of NSS data for rural India, *Health Economics* **15**: 915–932.

Brekke, Kurt, R., Siciliani, L. and Straume, O. R. (2008). Competition and waiting times in hospital markets, *Journal of Public Economics* **92**: 1607–1628.

Burge, P., Devlin, N., Appleby, J., Rohr, C. and Grant, J. (2004). Do patients always prefer quicker treatment? A discrete choice analysis of patients' stated preferences in the london patient choice project, *Applied Health Economics and Health Policy* **3**: 183–194.

Burns, L. R. and Wholey, D. R. (1992). The impact of physician characteristics in conditional choice models for hospital care, *Journal of Health Economics* **11**: 43–62.

Coombes, R. (2006). Patients get four choices for NHS treatments, *British Medical Journal* **332**: 8–8.

Damiani, M., Propper, C. and Dixon, J. (2005). Mapping choice in the NHS: Cross sectional study of routinely collected data, *British Medical Journal* **330**: 284–.

Dawson, D., Gravelle, H., Jacobs, R., Martin, S. and Smith, P. C. (2007). The effects on waiting times of expanding provider choice: Evidence from a natural experiment, *Health Economics* **16**(2): 113–128.

Deb, P. and Trivedi, P. (2002). The structure of demand for health care: latent class versus two-part models, *Journal of Health Economics* **21**(4): 601–625.

Department of Health (2008). Hospital waiting times/list statistics.
URL: *http://www.performance.doh.gov.uk/waitingtimes/*

Earwicker, S. C. and Whynes, D. K. (1998). General practitioners' referral thresholds and choices of referral destimation: An experimental study, *Health Economics* **7**: 711–722.

Goldman, D. and Romley, J. A. (2008). Hospitals as hotels: The role of patient amenities in hospital demand, *NBER Working Paper Series* **14619**.

Gravelle, H., Dusheiko, M. and Sutton, M. (2002). The demand for elective surgery in a public system: Time and money prices in the UK National Health Service, *Journal of Health Economics* **21 (3)**: 423–449.

Greene, W. and Hensher, D. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit, *Transportation Research Part B: Methodological* **37**(8): 681–698.

HES Online (2007). Differences between HES and NHS admitted patient waiting times data.

URL: *http:// www.hesonline.NHS.uk/Ease/servlet/ContentServer?siteID =1937categoryID=412*

Hodgkin, D. (1996). Specialized service offerings and patients' choice of hospital: The case of cardiac catheterization, *Journal of Health Economics* **15**(3): 305–332.

Hole, A. (2008). Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment, *Journal of Health Economics* **27**(4): 1078–1094.

Howard, D. H. (2005). Quality and consumer choice in healthcare: Evidence from kidney transplantation, Topics in Economic Analysis and Policy 5 (1).

Hynes, S., Hanley, N. and Scarpa, R. (2008). Effects on welfare measures of alternative means of accounting for preference heterogeneity in recreational demand models, *American Journal of Agricultural Economics* **90**(4): 1011.

Kjerstad, E. (2006). Patient mobility and the impact on waiting times, Nordic Health Economists Study Group.

Lindsay, C. M. and Feigenbaum, B. (1984). Rationing by waiting lists, *American Economic Review* **74 (3)**: 404–417.

Luft, H., Garnick, D., Mark, D., Peltzman, D., Phibbs, C., Lichtenberg, E. and McPhee, S. (1990). Does quality influence choice of hospital?, *JAMA* **263**(21): 2899–2906.

Mahon, A., Whitehouse, C., Wilkin, D. and Nocon, A. (1993). Factors that influence general practitioners' choice of hospital when referring patients for elective surgery, *British Journal of General Practice* **43**: 272–276.

Martin, S., Rice, N., Jacobs, R. and Smith, P. (2007). The market for elective surgery: Joint estimation of supply and demand, *Journal of Health Economics* **26**: 263–285.

Martin, S. and Smith, P. (1999). Rationing by waiting lists: an empirical investigation, *Journal of Public Economics* **71**: 141–164.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour, *in* P. Zarembka (ed.), *Frontiers in econometrics*, number 4, Academic Press, New York, pp. 105–142.

McGuire, T. g. (2000). Physician agency, *in* A. J. Culyer and J. P. Newhouse (eds), *Handbook of Health Economics*, chapter 9, pp. 461–536.

Monstad, K., Engesæter, L. and Espehaug, B. (2006). Patients preferences for choice of hospital, *HEB Working Paper 05/06* .

Ordnance Survey (2006).
URL: *http://www.ordnancesurvey.co.uk/gps/docs/ ProjectionandTransformationCalculations.xls*

Propper, C., Burgess, S. and Green, K. (2004). Does competition between hospitals improve the quality of care? - Hospital death rates and the NHS internal market, *Journal of Public Economics* **88**(7-8): 1247–1272.

Propper, C., Sutton, M., Whitnall, C. and F., W. (2008). Did ‘targets and terror’ reduce waiting times in england for hospital care?, *B.E. Journal of Economic Analysis and Policy (Contributions)* **8 (2)**.

Qian, D., Pong, R., Yin, A., Nagarajan, K. and Meng, Q. (2009). Determinants of health care demand in poor, rural China: the case of Gansu Province, *Health Policy and Planning* **24**: 324–334.

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature, *Stata Journal* **2**: 1–21.

Tay, A. (2003). Assessing competition in hospital care markets: The importance of accounting for quality differentiation, *RAND Journal of Economics* **34**(4): 786–814.

Varkevisser, M. and van der Geest, S. (2007). Why do patients bypass the nearest hospital? An empirical analysis for orthopaedic care and neurosurgery in the Netherlands, *The European Journal of Health Economics* **8**(3): 287–295.

Varkevisser, M., van der Geest, S. and Schut, F. (2010). Assessing hospital competition when prices donŠt matter to patients: the use of time-elasticities, *International Journal of Health Care Finance and Economics* **In Press**.

Whynes, D. K. and Reed, G. (1994). Fundholders’ referral patterns and perceptions of service quality in hospital provision of elective general surgery, *British Journal of General Practice* **44**: 557–560.

Windmeijer, F., Gravelle, H. and Hoonhout, P. (2005). Waiting lists, waiting times and admissions: an empirical analysis at hospital and general practice level, *Health Economics* **14**: 971–985.

Variable	Obs	Mean	S.D.	Min	Max
Number of hospitals	30				
Admissions per hospital-year	84	1037	768.95	82.7	327.2
Number of GP practices	1247				
Admissions per GP practice-year	3698	23.56	18.33	1	142
Hospitals referred to per GP practice-year	3698	1.61	0.63	1	5
Travel time to hospital visited (minutes)	87128	14.56	12.17	1.55	110.17
Travel time to closest hospital (minutes)	87128	12.23	7.63	1.55	69.87
Median annual inpatient wait across hospital-years (months)	84	6.51	3.17	0.375	16.96
Between-hospital variation			2.91	2.42	13.97
Within-hospital variation			1.50	3.26	10.31
Mean annual outpatient wait across hospital-years (months)	84	2.45	0.695	0.544	3.74
Between-hospital variation			0.648	0.601	3.29
Within-hospital variation			0.242	1.608	3.40

Table 1: Descriptive Statistics: Waiting time and travel time

Variable	Obs	Mean	S.D.	Min	Max
FTE GPs	1232	2.88	1.92	1	13
Training practice	1247	0.22	0.41	0	1
Rural practice	1247	0.14	0.35	0	1
List size/GP (000s)	1232	2.00	0.73	0	9.35
Patient age (10 years)	87128	7.51	1.04	0	10.3
Low Income Index (10%)	1225	1.52	0.91	0.2	58.7

Table 2: Descriptive Statistics: Practice characteristics

Variable	Model 1		Model 2		Model 3	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Ln(travel time)	-4.347***	0.182	-4.913***	0.217	-5.068***	0.164
x FTE GPs					0.308***	0.065
x Training practice					-0.469	0.361
x Rural practice					1.429**	0.562
x List size/GP (000s)					0.324**	0.157
x Patient age (10 years)					-0.233***	0.040
x Low Income Index (10%)					0.370**	0.158
Inpatient wait	0.056***	0.011	-0.060***	0.008	-0.069***	0.010
x FTE GPs					-0.002	0.009
x Training practice					-0.001	0.031
x Rural practice					0.081**	0.039
x List size/GP (000s)					-0.022	0.019
x Patient age (10 years)					-0.001	0.003
x Low Income Index (10%)					-0.004	0.017
Outpatient wait	0.068	0.050	-0.123***	0.046	-0.165***	0.051
x FTE GPs					0.022	0.034
x Training practice					-0.073	0.136
x Rural practice					0.352**	0.165
x List size/GP (000s)					0.152*	0.081
x Patient age (10 years)					-0.054***	0.014
x Low Income Index (10%)					-0.045	0.079
Hospital FEs	No		Yes		Yes	
No. of Parameters	3		32		50	
Observations	87128		87128		85425	
Log L	-74890		-57422		-54704	
AIC	149966		114908		109308	
BIC	149814		115208		108840	

Table 3: Results: Multinomial logit models

Variable	Model 4			
	Class 1		Class 2	
	Coeff.	S.E.	Coeff.	S.E.
Ln(travel time)	-8.674***	0.055	-3.075***	0.021
Inpatient wait	-0.016***	0.006	-0.109***	0.006
Outpatient wait	-0.124***	0.038	-0.270***	0.037
Class membership probability	0.671		0.329	
Constant	0.714***	0.064		
Hospital FEs	Yes			
No. of Parameters	36			
Observations	87128			
Log L	-47814			
AIC	95700			
BIC	96038			

Table 4: Results: Latent class multinomial logit model

Variable	Model 5			
	Class 1		Class 2	
	Coeff.	S.E.	Coeff.	S.E.
Ln(travel time)	-8.6731**	0.0549	-3.074***	0.021
Inpatient wait	-0.016***	0.0060	-0.109***	0.006
Outpatient wait	-0.125***	0.0375	-0.269***	0.037
Class membership probability	0.691		0.309	
Constant			-0.804***	0.070
Rural practice			0.590***	0.174
Hospital FEs	Yes			
No. of Parameters	37			
Observations	87128			
Log L	-47808			
AIC	95690			
BIC	96037			

Table 5: Results: Latent class multinomial logit model with practice rurality determining class probability

	Average elasticity of demand		
Model	Inpatient wait	Outpatient wait	Travel time
2 MNL	-0.121	-0.087	-1.481
3 MNL with interactions	-0.124	-0.095	-1.445
4 Latent-Class	-0.123	-0.129	-1.472
5 Latent-Class with rural	-0.124	-0.129	-1.461

Table 6: Average waiting time elasticities of demand

Variable	Model R1		Model R2	
	Coeff	S.E.	Coeff	S.E.
Closest Hospital			0.578***	0.129
Ln(travel time)	-5.026***	0.174	-4.23***	0.29
Inpatient wait	-0.057***	0.008	-0.051***	0.008
Outpatient wait	-0.163***	0.042	-0.120**	0.049
No. of Parameters	32		33	
Observations	87128		87128	
Log L	-65483		-56742	
AIC	131030		113550	
BIC	131330		113859	

Table 7: Multinomial logit model robustness checks

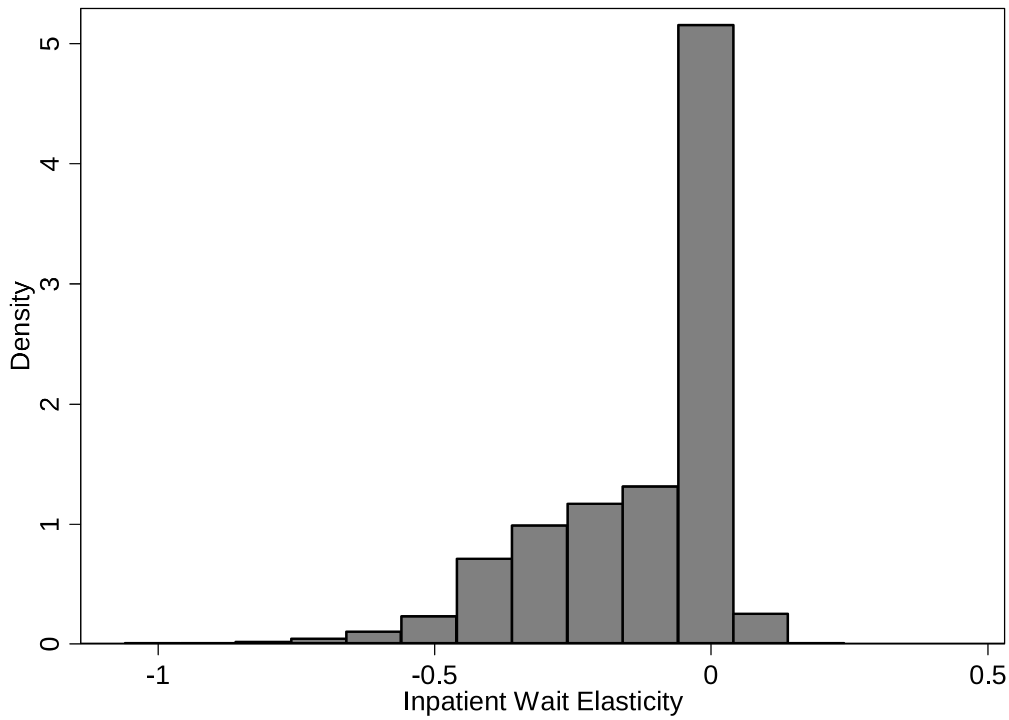

Figure 1: Model 3 (MNL) Inpatient Waiting Time Elasticity

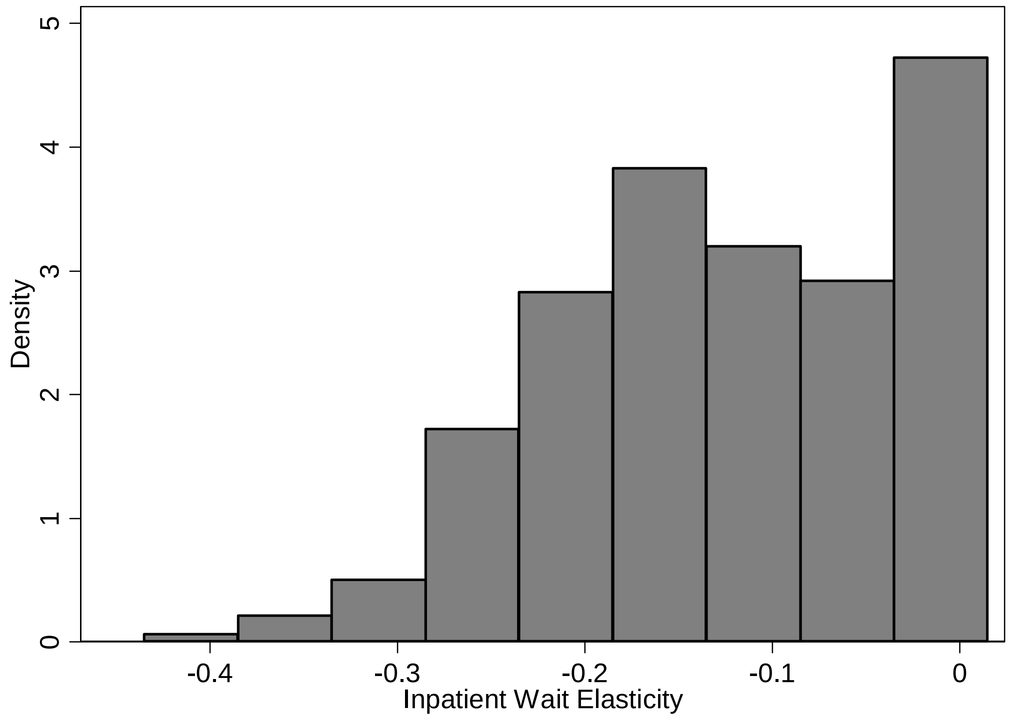

Figure 2: Model 5 (LCMNL) Inpatient Wait Elasticity

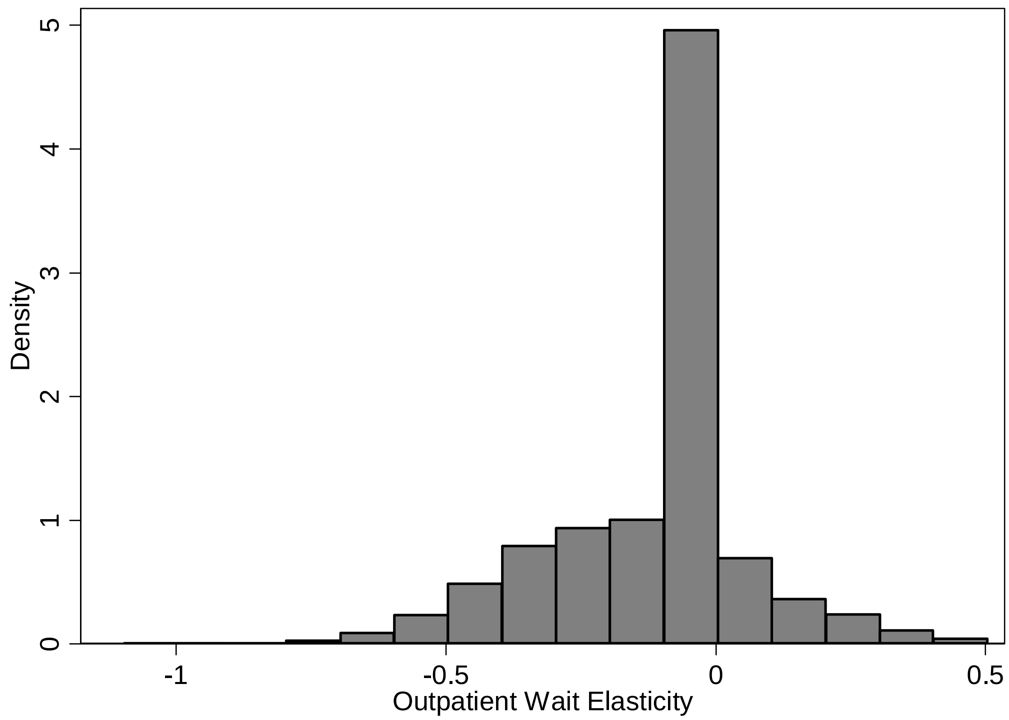

Figure 3: Model 3 (MNL) Outpatient Wait Elasticity

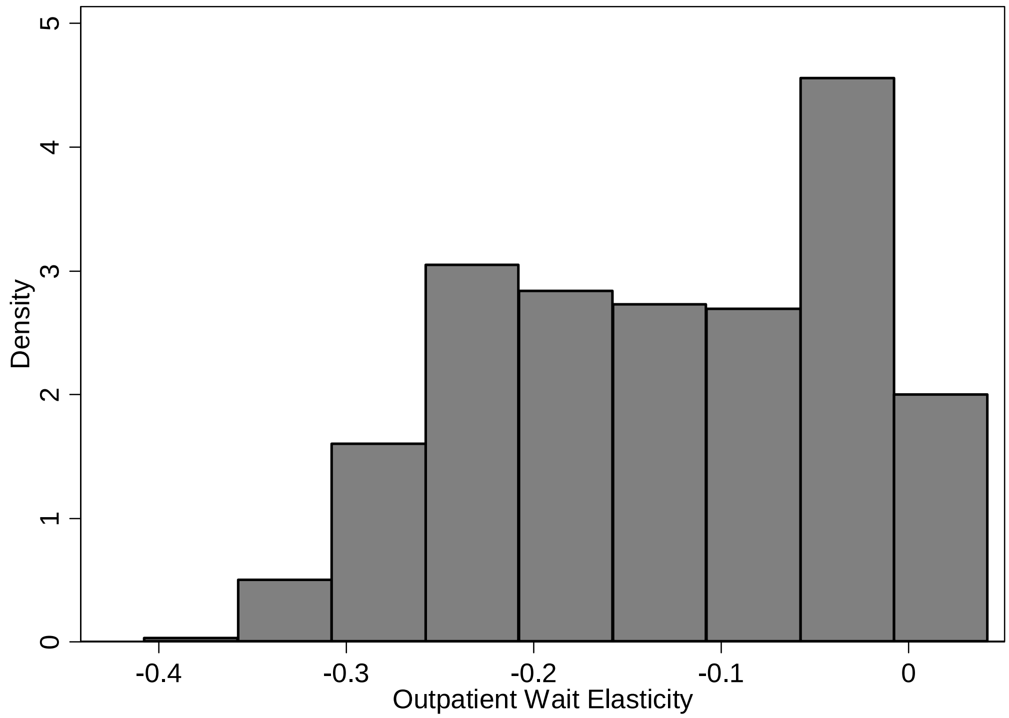

Figure 4: Model 5 (LCMNL) Outpatient Wait Elasticity