ECONOMIC DISCUSSION PAPERS

Efficiency Series Paper 06/2007

The effects of Stochastich Demand and Expense Preference Behaviour on Public Hospital Cost and Excess Capacity

Knox Lovell, Ana Rodríguez-Álvarez y Alan Wall

Departamento de Economía

Universidad de Oviedo

Available online at: www.uniovi.es/economia/edp.htm

UNIVERSIDAD DE OVIEDO

DEPARTAMENTO DE ECONOMÍA

PERMANENT SEMINAR ON EFFICIENCY AND PRODUCTIVITY

THE EFFECTS OF STOCHASTIC DEMAND AND EXPENSE PREFERENCE BEHAVIOUR ON PUBLIC HOSPITAL COST AND EXCESS CAPACITY

Knox Lovell♦, Ana Rodríguez-Álvarez♠ and Alan Wall♣

Efficiency Series Paper 06/2007

Abstract: The literature to date on the effect of demand uncertainty on public hospital costs and excess capacity has not taken into account the role of expense preference behaviour. Similarly, the research on expense preference behaviour has not taken demand uncertainty into account. In this paper we argue that both demand uncertainty and expense preference behaviour may affect public hospital costs and excess capacity and that ignoring either of these effects may lead to biased parameter estimates and misleading inference. To show this, we extend the analysis of Rodríguez-Álvarez and Lovell (2004) by incorporating demand uncertainty into the technology to account for the hospital activity of providing standby capacity or insurance against unexpected demand. We find that demand uncertainty in Spanish public hospitals affects hospital production decisions and increases costs. Our results also show that overcapitalization in these hospitals can be explained by hospitals providing insurance demand when faced with demand uncertainty. We also find evidence of expense preference behaviour. We conclude that both stochastic demand and expense preference behaviour should be taken into account when analysing hospital costs and production.

Key words: Hospital production, excess capacity, uncertain demand, bureaucracy, input distance function.

♣ Corresponding author. Departamento de Economía, Universidad de Oviedo, Spain.
e-mail: awall@uniovi.es
♠ Departamento de Economía, Universidad de Oviedo, Spain.
♦ University of Georgia, USA and University of Queensland, Australia.

1. Introduction

One of the characteristics of hospitals is that they face uncertain demand for the services that they offer. Faced with the possibility of upsurges in demand, a large part of which is unpredictable, hospital administrators will maintain a reservation capacity. This decision to reserve capacity may be motivated for a variety of reasons. Carey (1998), for example, points out that the public place a high value on health care availability and hospitals thus wish to prevent queuing or turning patients away. This reserve capacity will also ensure that there are additional staff available to serve patients and thus may be considered as a provision of quality. Whatever the precise reason for the decision to maintain reserve capacity, this decision will affect the structure of hospital production and costs. Gaynor and Anderson (1995) note that the standard theory of cost and production requires that production be technically efficient in the sense that the firm should be operating on the boundary of the production possibilities set or production "frontier". However, this property will not hold for hospitals who wish to provide something close to a universal provision of service.

The recognition that reservation capacity decisions affect the cost structure of hospitals has led a number of researchers to incorporate stochastic demand into their specifications of hospital costs. Friedman and Pauly (1981; 1983) were the first authors to incorporate stochastic demand into hospital cost functions, proposing a "latent penalty" model where hospitals minimize costs which include a deterioration of quality during periods of unexpectedly high demand. In particular, forecasted output is included in their specification of the cost function in addition to actual output using the argument that if a bed is forecast to be occupied then it will be staffed and thus add to costs. Gaynor and Anderson (1995), modifying a model by Duncan (1990), assume that hospitals maintain capacity in order not to deny care to patients and thus try to keep the "turn-away probability" below some target level. Carey (1998) combines considerations of costs and benefits of empty beds, deriving an equilibrium condition for the optimal number of "excess" beds and estimating a cost function which incorporates forecasted demand in accordance with the theoretical framework of Gaynor and Anderson (1995). Hughes and McGuire (2003) distinguish between elective admissions (predictable) and emergency admissions (unpredictable) to address the issue of production responses to demand uncertainty. Unpredictable demand is specified as the difference between actual and forecast emergency demand

and this variable is introduced into the specification of the cost function. Baker et al. (2004) estimate a cost function along the lines of Gaynor and Anderson (1995) using data on daily occupancy to investigate whether day-to-day variation in utilization has an impact on hospital costs. They find that increases in variance are associated with increases in hospital costs but that the effects are relatively modest. Smet (2004) estimates a generalized translog multi-product cost function with data on Belgian general care hospitals using a queuing theory indicator to capture standby capacity. His results show that providing standy capacity has a significant impact on total costs.

These studies share the common feature that they all specify empirical cost functions which differ from conventional cost functions in that they incorporate variables which capture demand uncertainty. The results show that demand uncertainty has an impact on hospital costs. Underlying these analyses is the fact that the hospitals are assumed to be operating off their production frontier, i.e. they are not producing in a technically efficient manner due to the fact that they maintain reserve capacity. However, the use of short-run cost functions implies the assumption of cost minimization with respect to variable inputs, and in public hospitals, which have a bureaucratic structure and a lack of incentives on the part of agents to introduce cost minimization criteria (see Rodríguez-Álvarez and Lovell, 2004), this assumption may be questionable.

An issue ignored in these papers, therefore, is the fact that excess capacity may also be due to other factors such as expense preference behaviour on the part of public hospital administrators. Expense preference behaviour occurs when managers use the budget to acquire inputs that provide both visible output and personal utility in the form of prestige, status, job security, etc (see Williamson, 1963; Migué and Bélanger ,1974; Lindsay and Buchanan, 1970; Lee, 1971; and Evans, 1971). Rodríguez-Alvarez and Lovell (2004) modelled public hospitals in a bureaucratic expense preference setting and found evidence of systematic allocative inefficiency in variable inputs and overcapitalisation. Given the absence of cost-minimisation behaviour, Rodríguez-Álvarez and Lovell (henceforth RL) used an input distance function to model hospital technology, exploiting the duality of this function with the shadow cost function in order to analyse the effect of expense preference on costs. RL ignored, however, the potential effect of stochastic hospital demand on excess capacity.

In this paper we pull together these two strands of literature by arguing that public hospital production and costs may be affected by *both* demand uncertainty and bureaucratic expense preference behaviour. To do so, we extend the RL model by introducing the effect of demand variability. We observe, as in RL, systematic allocative inefficiency. We also find that hospitals are overcapitalised but that this is explained by both expense preference behaviour and the desire of hospitals to maintain additional capacity to insure against demand uncertainty.

Two main implications arise from our research. On the one hand, we show that care must be taken to incorporate demand uncertainty and risk averse behaviour into studies dealing with systematic inefficiency in public hospitals, as there is a danger of unjustly labelling input usage which provides insurance against demand uncertainty as inefficiency. On the other hand, studies on the effects of demand uncertainty on hospital costs and production should also take account of the possibility of the existence of systematic inefficiency due to expense preference behaviour, as ignoring this may lead to an overestimation of the effect of demand uncertainty on input usage. In general, ignoring either of these two effects if both are present will lead to biased parameter estimates and misleading inference.

2. The effect of demand uncertainty on hospital production behaviour and cost

The theoretical model we use is a variation of that proposed by RL. In their model, hospital managers choose inputs in order to maximise their utility. Utility, in turn, is a function of hospital outputs, and given that inputs influence outputs, inputs will be choice variables in the utility function. Hospital demand is, however, stochastic, and the uncertainty associated with this demand will constrain behaviour. In particular, managers face social pressure to satisfy a large percentage of demand and must therefore ensure that they can provide health care to any given patient with the maximum feasible probability. Thus, we assume that managers are risk averse and produce in order to ensure a provision of service which minimises the probability of turning away patients (Duncan, 1990; Gaynor and Anderson, 1995). Given a manager's level of risk aversion, input choices will therefore depend on the level of demand uncertainty.

Introducing stochastic demand into the RL model, the utility maximization problem of the bureaucratic hospital manager can be expressed as follows:

$$\begin{aligned} &\max U = U(y, y_z, x, K) \\ &\text{s.t.} \quad \sum_{i=1}^{n} w_i x_i \leq P \\ &D_I (y, y_z, x, K) \geq 1 \\ &y^{MAX} - y \geq y_z (\sigma_y) \;\; , y_z'(\sigma_y) > 0 \end{aligned} \tag{1}$$

where the utility function depends on (exogenous) outputs (y and y_z), the choice of (endogenous) variable inputs, x, and quasi-fixed inputs K. *P* represents the exogenous budget, and w is an exogenous variable input price vector. D_I, the input distance function, captures the technology.

The difference with the Rodríguez-Álvarez and Lovell (2004) model is found in the definition of what the hospital produces, that is, in the output. As Smet (2004) neatly summarises, "hospitals should in addition to predicted patient care also provide sufficient standby capacity to service unexpected care. This option of insurance demand should be treated as a service provided by the hospital."

We assume that the hospital target includes, apart from the provision of the observed level of demand (y), an additional output, insurance demand (y_z) where we assume $\partial U(y,z,x,K)/\partial y_z > 0$ because greater insurance output increases the probability of satisfying the risk averse manager's minimum provision of service. Insurance demand also appears in the distance function representation of the technology. The larger the variability of the demand faced by the hospital, the more resources that are needed to maintain this insurance output. Thus, $y_z = y_z (\sigma_y)$ where σ_y is the standard deviation of demand (Gaynor and Anderson, 1995; Baker et al., 2004). The final constraint in (1) says that the sum of insurance output (y_z) and observed output (y) do not exceed the maximum possible output that could be produced if capacity were used to the full (y^{MAX}), with $y_z'(\sigma_y) > 0$ implying that the greater demand variability, the more insurance output the risk averse manager will produce.

Moreover, the hospital outputs (y and y_z) are different in that they consume different inputs and have different costs. Following Carey (1998) and Baker et al. (2004), in our short-run model we take fixed (or "quasi-fixed") inputs as given. The decision faced by hospital managers is: a) to contract in advance (i.e. before demand is realized) those inputs which are *not perfectably adjustable* taking into account the two objectives of the hospital, namely to face expected demand and provide "insurance" for unexpected demand; and b) contract *perfectly adjustable inputs* in response to actual demand. It is important to point out that while observed admissions (y) consume both types of variable inputs (imperfectly and perfectly adjustable), insurance demand (y_z) can be provided by contracting only imperfectly adjustable inputs. Thus, the two activities of the hospital consume different inputs and will therefore have different costs.

The technology is approximated by an input distance function which is illustrated in Figure 1 for the case where two variable inputs (x_1, x_2) are used to produce output level *y*. With the observed input combination given by the point *P* and the observed output level equal to y_0, the distance function is represented by the ratio OP/OA which is the reciprocal of the Farrell (1957) index of technical efficiency. It is the maximum radial contraction of the input vector that can be made while still permitting the firm to produce output level y_0. Note that the hospital is inefficient because it could reduce its input vector and still provide the two hospital outputs. At the existing input prices, the cost of producing y_0 is C_A.

Figure 1. The cost of demand uncertainty

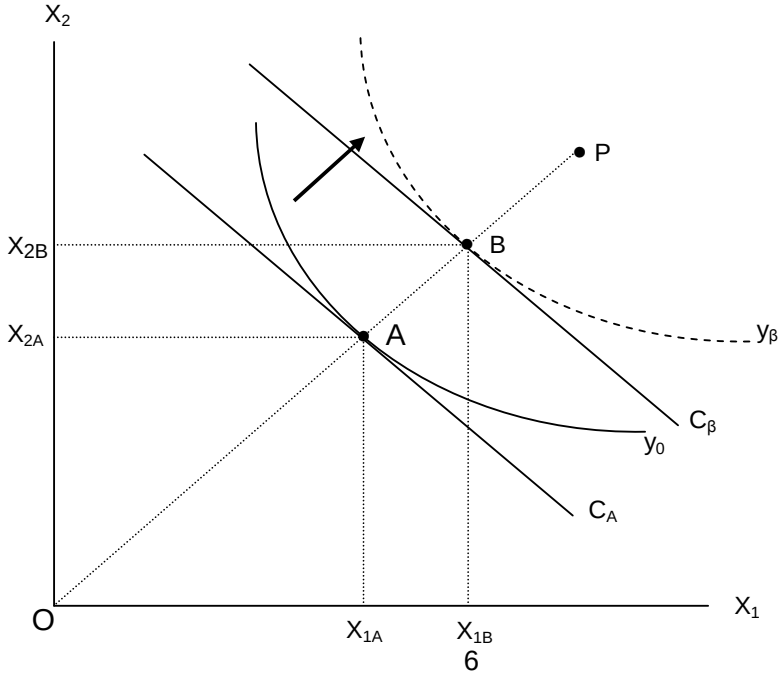

Figure 1 also illustrates the effect of demand uncertainty on hospital costs and efficiency measurement. Assume that the variable inputs in Figure 1 are not perfectly adjustable. Given the existence of demand uncertainty, the risk averse hospital manager will wish to keep the provision of service probability up to a level β and will thus contract in advance more of these inputs in order to cover this insurance demand. To guarantee this insurance demand output in a technically efficient way, the firm should choose the input combination represented by the point *B*. The new value of the distance function is OP/OB, and we see that the β-taking hospital is inefficient but less so than when we consider the hospital to be producing only the observed output.

The result of this is that the Farrell measure of technical efficiency increases from OA/OP to OB/OP. The lesson here is that if we consider only observed output we will overestimate the technical inefficiency of the hospital. To further illustrate this point, suppose the β-taking hospital contracts the input vector represented by point B and thus provides insurance output in an efficient way. If we do not take the insurance demand of the hospital into account and only consider actual output, as represented by actual discharges, we will label the hospital as inefficient - the distance function will take the value OB/OA > 1 instead of the correct value of OA/OA = 1. Finally, note that the minimum cost of producing when we take into account the insurance demand output is $C_\beta > C_A$. Moreover, it can be seen that $C_\beta - C_A$ represents the cost of producing the insurance output.

3. Estimating the cost of demand uncertainty

The question then arises as to how to calculate the effect of demand uncertainty on hospital costs. This can be done using the distance function by exploiting the duality between the input distance function and the (shadow) cost function defined by Shephard (1953). In the short-run case, this duality is represented as:

$$D(y, z, x,,K) = \min_{W^s} \{W^s x : C(y, y_z, W^s, K) = 1\} \quad (2)$$

$$C(y, y_z, w^s, K) = \min_x \{w^s x : D(y, y_z, x, K) = 1\} \quad (3)$$

where:

- w^s is the vector of input prices which minimises the cost of producing y and y_z, given x and K. This vector is known as the shadow price vector and coincides with the market price vector, w, when the choice of inputs, x, is optimum given y, y_z and K.
- $C(y, y_z, w^s, K)$ is the (shadow) short-run cost function and represents the minimum cost of producing y and y_z given x and K.
- $W^s = \frac{w^s}{C(y, y_z, w^s, K)}$ is the normalised shadow price vector.

It can be shown (see Appendix I) that

$$\frac{\frac{\partial C(y, y_z, w^s, K)}{\partial y_z}}{C(y, y_z, w^s, K)} = -\frac{\partial D(y, y_z, x, K)}{\partial y_z} \quad (4)$$

That is, in absolute terms the value of $\partial D(y, y_z, x, K)/\partial y_z$ is equal to the (normalized) marginal cost of uncertainty. Given that $C(y, y_z, w^s, K)$ is always positive, and given that a risk averse manager choses inputs to provide service to guarantee insurance demand ($\partial U(y, y_z, x, K)/\partial y_z > 0$) and therefore, $\partial C(y, y_z, w^s, K)/\partial y_z > 0$, the expected sign of the distance function with respect to z is negative, indicating that an increase in uncertainty increases costs. This is a testable hypothesis but there still remains the problem that the shadow short-run cost function $C(y, y_z, w^s, K)$ is unobservable so that (4) cannot be used to test the hypothesis. To resolve this problem, redefine the distance in logarithmic terms so that

$$\ln D(y, y_z, x, K) \geq \ln 1 \quad (5)$$

By definition,

$$\frac{\partial D(y, y_z, x, K)}{\partial y_z} = \frac{\partial \ln D(y, y_z, x, K)}{\partial \ln y_z} \frac{D}{y_z} \quad (6)$$

Also by definition, the normalized shadow cost function is that which makes the cost associated with a point on the isoquant optimum, given the input combination utilized. Thus:

$$C(y, y_z, w^s, K) = \frac{C}{D} \tag{7}$$

where $\frac{C}{D}$ is the observed cost evaluated at a point on the isoquant.

Using (4), (6) and (7), we obtain:

$$MC_z = \frac{\partial C(y, y_z, x, K)}{\partial y_z} = -\frac{\partial \ln D(y, y_z, x, K)}{\partial \ln y_z} \frac{C}{y_z} \tag{8}$$

Equation (8) indicates how the marginal cost of uncertainty, MC_z, can be estimated in monetary terms once a short-run input distance function has been estimated. The empirical specification of the distance function is the focus of the next section.

4. Empirical specification

In order to capture the effect of demand uncertainty we need an estimate of the variability of demand. Of course, this is unobservable but the mean and variance of demand conditional on past realizations can be used to estimate it. For a given level of risk aversion, desired insurance demand output is increasing in the mean or variance of the distribution of demand assuming a normal distribution (see Gaynor and Anderson, 1995). Therefore, we need to estimate a demand equation. We follow Freidman and Pauly (1983) and Hughes and McGuire (2003) by modelling a simple autoregressive (AR1) process where demand expectations are related to prior demand realizations. We also include bed size in the regression to control for hospital size. The demand equation to be estimated is thus specified as:

$$DEM_t = a D_h + b\, BED_t + \rho\left[DEM_{t-1} - \left(a D_h + b\, BED_{t-1}\right)\right] \tag{9}$$

where *DEM* represents demand, D_h are hospital dummy variables, and *BED* is the number of beds. The forecast demand from equation (9) is used to approximate the

expected demand and we use absolute value of the residual from (9) to approximate demand variability.

Once the expected demand and variability of demand are obtained, we then estimate the distance function in order to measure the effect of demand uncertainty on costs. To do so, we follow Rodríguez-Álvarez and Lovell (2004) by estimating a translog short-run input distance function jointly with the cost share equations, where the latter are obtained by differentiating the distance function (in logarithmic terms) with respect to each input. In a panel data setting the empirical model is specified as:

$$\ln 1 = B_0 + \sum_{r=1}^{m} \alpha_r \ln y_{rht} + \frac{1}{2}\sum_{r=1}^{m}\sum_{s=1}^{m} \alpha_{rs} \ln y_{rht} \ln y_{sht} + \sum_{i=1}^{n} \beta_i \ln x_{iht} +$$
$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \beta_{ij} \ln x_{iht} \ln x_{jht} + \frac{1}{2}\sum_{r=1}^{m}\sum_{i=1}^{n} \omega_{ri} \ln x_{iht} \ln y_{rht} + \xi_f \ln K_{ht} + \xi_{ff} \frac{1}{2}(\ln K_{ht})^2 +$$
$$\sum_{i=1}^{n} \xi_{fi} \ln K_{ht} \ln x_{iht} + \sum_{r=1}^{m} \xi_{fr} \ln K_{ht} \ln y_{rht} + \xi_{af} \ln K_{ht} \ln y_{z_{ht}} +$$
$$\pi_\alpha \ln y_{z_{ht}} + \pi_{\alpha\alpha} \frac{1}{2}(\ln y_{z_{ht}})^2 + \sum_{i=1}^{n} \pi_{\alpha i} \ln y_{z_{ht}} \ln x_{iht} + \sum_{r=1}^{m} \pi_{\alpha r} \ln y_{z_{ht}} \ln y_{rht} + \sum_{t=1}^{T-1} \gamma_t D_T + \eta_{ht} + \delta_h \qquad (10)$$

$$\frac{x_{iht} w_{iht}}{C_{ht}} = (\beta_i + a_i) + \sum_{j=1}^{n} \beta_{ij} \ln x_{jht} + \sum_{r=1}^{m} \omega_{ri} \ln y_{rht} + \pi_{\alpha i} \ln y_{z_{ht}} + \xi_{fi} \ln K_{ht} + (A_i - a_i) + \eta_{iht} \qquad (11)$$

where $y = (y_1,\ldots, y_m)$ is a vector of observed outputs, $x = (x_1,\ldots, x_n)$ is a vector of variable inputs, y_z is the variable capturing the degree of demand uncertainty represented by the absolute value of the residual from the demand equation (9), K is a quasi-fixed input, D_T are a set of time dummy variables, η_{iht} and η_{ht} are random disturbance terms which are distributed as i.i.d. $(0, \sigma^2)$. The δ_h represents individual hospital-specific fixed effects which permit technical inefficiency as well as unobserved, time-invariant differences between hospitals to be captured. Homogeneity of degree one in inputs is imposed by the following restrictions on the parameters: $\sum\beta_i = 1$; $\sum\beta_{ij} = 0$; $\sum\omega_{ri} = 0$; $\sum \pi_{\alpha i} = 0$; and $\sum\xi_{fi}=0$. The possibility of systematic allocative inefficiency is allowed for through the A_i terms, where $E(A_i) = a_i$. Finally, as there are endogenous inputs, estimation of the model is carried out using the same instrumental variable techniques as in RL.

5. Data and results

We use the same data as RL and we include a summary in Appendix I. We use three outputs: MED, which is the weighted sum (using weighted care units, or UPAs) of discharges in general medicine, surgery, obstetrics, paediatrics and intensive care; AM, which is the weighted sum outpatient visits and emergencies; and y_z, which is insurance demand and which is approximated by the residual of the demand equation (3) in the previous section). There are four variable inputs: care graduates, G; technical personnel, T; other personnel, RES; and supplies, S. The quasi-fixed input is the number of beds, BED. To capture the complexity of the hospitals we include the number of students, TEACH. We also include time dummies, D_T.

The results of the estimation of equation (9) are shown in Table 1.

Table 1. Demand estimation

Parameter	Estimate	t-statistic
RHO	0.6133	17.0268 **
BED	42.4744	16.4390 **
Adjusted R^2	0.9674	
Durbin-Watson	1.9021	

** statistically significant at 5%

The high R^2 indicates a good performance by the demand equation. Moreover, the autocorrelation between periods appears to be adequately captured, with the value of the parameter ρ being very similar to previous estimates in the literature.[1] From this estimation it is possible to obtain the difference between actual demand and predicted demand and we use this difference (in absolute values) as a measure of demand uncertainty.

The system of equations (10-11) was estimated, dropping one of the share equations, by iterative SURE and the results of the estimation are shown in Table 2.

[1] Our value of $\rho = 0.61$ is close to the value found by Hughes and McGuire of 0.66. Friedman and Pauly (1983) estimated average values of 0.62.

Table 2. Estimation of system of equations

Variable	Coefficients (a)	t-statistic	Variable	Coefficients (a)	t-statistic
L(MED)	-0.2643	-7.5120 **	L(G) L(G)	0.1139	6.1217 **
$L(y_z)$	-0.0104	-3.5057 **	L(MED) L(G)	0.0027	0.2120
L(AM)	-0.0531	-2.8940 **	$L(y_z)$ L(G)	-0.0002	-0.1957
L(G)	0.1631	4.1491 **	L(AM) L(RES)	0.0199	1.8281 *
L(T)	0.3422	6.8653 **	L(AM) L(T)	0.0170	2.3922 **
L(RES)	0.3739	6.6251 **	L(G) L(T)	-0.0644	-4.7876 **
L(S)	0.1206	3.9608 **	L(G) L(RES)	-0.0474	-2.4783 **
L(BED)	0.1310	2.5888 **	L(AM) L(S)	-0.0262	-2.9885 **
L(TEACH)	-0.0050	-1.0444	L(AM) L(BED)	0.2721	4.2430 **
L(MED).L(MED)	0.8858	8.6321 **	L(T) L(BED)	-0.0168	-1.5151
$L(y_z)\cdot L(y_z)$	-0.0050	-2.4425 **	L(RES) L(BED)	-0.0094	-0.5635
L(AM)·L(AM)	0.1232	2.1736 **	L(S) L(BED)	0.0295	2.1036 **
L(T).L(T)	0.0821	3.7136 **	L(G) L(S)	-0.0020	-0.1803
L(T) L(RES)	-0.0016	-0.0799	L(TEACH) L(TEACH)	0.0082	1.9971 **
L(T).L(S)	-0.0160	-1.5381	L(MED) L(TEACH)	-0.0491	-2.8650 **
L(RES).L(S)	-0.0696	-4.8560 **	$L(y_z)$ L(TEACH)	0.0037	1.6213
L(RES).L(RES)	0.1187	3.8249 **	L(AM) L(TEACH)	-0.0247	-2.1146 **
L(S)·L(S)	0.0876	6.2934 **	L(T) L(TEACH)	-0.0040	-2.1809 **
L(BED)·L(BED)	0.1266	0.8673	L(RES) L(TEACH)	-0.0009	-0.3362
$L(MED).L(y_z)$	-0.0034	-0.2359	L(S) L(TEACH)	0.0069	2.9385 **
L(MED) L(AM)	-0.3108	-4.6483 **	L(BED) L(TEACH)	0.0498	3.0168 **
L(MED) L(T)	-0.0002	-0.0181	$a_{graduates}$	0.0565	1.4360
L(MED) L(RES)	-0.0008	-0.0509	$a_{technicians}$	-0.1173	-2.3513 **
L(MED) L(S)	-0.0017	-0.1252	$a_{supplies}$	-0.1101	-1.9485 **
L(MED) L(BED)	-0.4408	-4.0232 **	$a_{other\ personnel}$	0.1709	5.4494 **
$L(y_z)$ L(AM)	0.0056	0.6775	γ_{89}	-0.0561	-4.7559 **
$L(y_z)$ L(T)	-0.0002	-0.2866	γ_{90}	-0.1276	-9.8942 **
$L(y_z)$ L(RES)	-0.0009	-0.7463	γ_{91}	-0.2078	-14.7001 **
$L(y_z)$ L(S)	0.0013	1.2719	γ_{92}	-0.2227	-13.9153 **
$L(y_z)$ L(BED)	-0.0032	-0.2144	γ_{93}	-0.2245	-12.8239 **
L(G) L(BED)	-0.0033	-0.2475	γ_{94}	-0.2162	-11.3647 **
L(AM) L(G)	-0.0108	-1.2649			

(a)Evaluated at the means of the data using parameter estimates of (4)-(5).

The variables are expressed in terms of deviations from their means. Thus, the function estimated is a Taylor series approximation to the true, but unobservable, distance function at the mean of the data and the first order coefficients can be interpreted as elasticities at the distance frontier for the mean hospital of the sample. At the sample mean, we have checked that the estimated distance function satisfies the properties required by theory: it is decreasing in outputs and non-increasing and concave in variable inputs.

The coefficient on the variable capturing demand uncertainty (y_z) is negative and significant. Thus, and in accordance with equation (2), our hypothesis that desired insurance demand increases hospital costs is confirmed as the sign of this coefficient

shows that more inputs are needed to insure against unexpected demand. More precisely, the marginal cost of uncertainty at the sample mean, interpreted as a one-unit increase in the standard deviation of hospital demand, is 412.51 euro -euro of the year 2006- (standard error=0.035**). The marginal cost of the other hospital output can be expressed as (see Section 2):

$$MC_y = \frac{\partial C(y, y_z, x, K)}{\partial y} = -\frac{\partial \ln D(y, y_z, x, K)}{\partial \ln y}\frac{C}{y} \qquad (12)$$

Applying equation (12) to the results obtained from our estimation, we find that the marginal cost of a medical stay is 768.89 euro (standard error=0.003**). In keeping with the discussion in Section 2, this cost is higher than the cost of insurance demand due to the fact that observed patients also consume perfectly adjustable inputs whereas demand uncertainty only affects imperfectly adjustable inputs. Finally, the cost of an external consultancy is 92.28 euro (standard error=0.018**).

Systematic allocative inefficiency can be calculated by analysing the parameters a_i. The results confirm, as in RL, the presence of systematic allocative inefficiency, thereby reaffirming the validity of the bureaucratic model. Moreover by comparing our results with those in RL we find that hospital demand uncertainty does not significantly influence the proportion in which inputs are used. It appears that hospitals scale up imperfectly variable inputs to service unexpected care.

6. Comparison of models

As our model takes into account the effect of both demand uncertainty and expense preference behaviour on hospital production and costs, it nests the RL model, which accounts for expense preference but not stochastic demand. It also nests a Gaynor and Anderson type model (in the sense of a model which accounts for stochastic demand, but not expense preference). That is, under the appropriate set of parameter restrictions, our model collapses to either the RL model (setting the coefficients on all the terms involving "y_z" equal to zero), the stochastic demand (Gaynor and Anderson type) model (setting the "a_i" parameters equal to zero), or to a base model which accounts for neither stochastic demand nor expense preference (setting both the "a_i" parameters and "y_z" coefficients equal to zero). This provides a convenient basis for

checking the validity of our model against the latter restricted models. On the basis of likelihood ratio tests, all three restricted models were rejected and our model was thus found to be a better representation of the technology for our sample of hospitals (see Table 3).

Table 3. Model selection tests

Hypothesis with regard to the general model	Value of test statistic	Number of restrictions
Demand uncertainty effects but no expense preference behaviour	30.62**	3
Expense preference behaviour but no demand uncertainty effects	19.58**	9
No demand uncertainty effects and no expense preference behaviour	49.06**	12

** Rejected at 5% level of level of significance. The critical value is a chi-square statistic with degrees of freedom equal to the number of restrictions.

The immediate implication is that a model which does not account for expense preference behaviour would, for our data set, be misspecified and therefore provide biased parameter estimates and misleading inference. Moreover, if demand uncertainty is not included, then excess capacity will be overestimated. To see this, following RL we can calculate the marginal cost or shadow price (r^S) of the capital input (BED), in the same way we calculate the marginal cost of uncertainty in (2), to obtain an index of overcapitalisation by comparing the shadow price and market price (r) of capital. A value less than 1 of the index $q = r^S/r$ shows overcapitalisation or excess capacity.

The values for the q index for our model and the restricted models are presented in Table 4. As the calculation of the q index involves parameters which have standard errors attached, we also check the significance of overcapitalisation by testing whether q is significantly lower than 1. Starting with a model in which both expense preference and demand uncertainty are ignored we find q = 0.57 and that this is less than 1 at the 5% level of significance, which would provide strong evidence that hospitals are overcapitalised. When expense preference and demand uncertainty are introduced, the degree of overcapitalisation falls: q rises to 0.77 and the hypothesis that this is lower than 1 would be soundly rejected even at the 10% level of significance, which shows that these expense preference behaviour and reactions to demand uncertainty explain overcapitalisation.

Table 4. Estimated overcapitalisation index

Model	Value of coefficient q	p-value for H_0: q < 1
Demand uncertainty effects and expense preference behaviour	0.77	0.19
Demand uncertainty effects but no expense preference behaviour	0.72	0.15
Expense preference behaviour but no demand uncertainty effects	0.64*	0.09
No demand uncertainty effects and no expense preference behaviour	0.57**	0.05

Note: We have tested the significance of the indices using the Wald test.

Thus, our calculations indicate that there is significant overcapitalisation in the short run in a model with neither expense preference behaviour nor stochastic demand effects. Our model shows that taking account of these effects provides an explanation for excess capacity.

7. Conclusions

Our research shows that the roles of both stochastic demand and expense preference behaviour should be considered when studying hospital production and costs. Ignoring either of these effects if they are present will lead to biased parameter estimates and misleading inference and if applied researchers have reason to suspect that these effects may be present they should test for them. We find evidence of both effects using data from Spanish general hospitals and it would interesting to see if the same applies for other datasets. Note, finally, that the issues involved here are not limited to hospitals and that our analysis could easily be applied to other sectors such as, for example, electricity distribution and transportation.

References

Baker, L.C., Phibbs, C.S., Guarino, C., Supina, D. and J.L. Reynolds (2004), "Within-year variation in hospital utilization and its implications for hospital costs", *Journal of Health Economics* 23, 191-211.

Carey, K. (1998), "Stochastic demand for hospitals and optimizing "excess" bed capacity", *Journal of Regulatory Economics* 14, 165-187

Duncan, G.D. (1990), "The effect of probabilistic demands on the structure of cost functions", *Journal of Risk and Uncertainty* 3, 211-220.

Evans R. G. (1971), "Behavioural Cost Functions for Hospitals", *Canadian Journal of Economics* 4, 198-215

Farrell, M. J. (1957): "The Measurement of Productive Efficiency", *Journal of the Royal Statistical Society, Series A, General*, 120, pp. 253-281.

Friedman, B. and M.V. Pauly (1981), "Cost functions for a service firm with variable quality and stochastic demand: the case of hospitals", *Review of Economics and Statistics* 63, 620-624.

Friedman, B. and M.V. Pauly (1983), "A new approach to hospital cost functions and some issues in revenue regulation", *Health Care Finance Review* 4, 105-114.

Gaynor, M. and G.F. Anderson (1995), "Uncertain demand, the structure of hospital costs, and the cost of empty hospital beds", *Journal of Health Economics* 14, 291-317.

Hughes, D. and A. McGuire (2003), "Stochastic demand, production responses and hospital costs", *Journal of Health Economics* 22, 999-1010.

Lee M. L. (1971), "A Conspicuous Production Theory of Hospital Behavior", *Southern Economic Journal* 38, 48-58.

Lindsay C. M. and J. M. Buchanan (1970), *The Organization and Financing of Medical Care in the United States (Part. I)*. In *British Medical Association* (ed.), Health Services Financing, London, 535-562.

Migué, J. L. and G. Bélanger (1974), "Toward a General Theory of Managerial Discretion". *Public Choice*, 17, 27-47.

Rodríguez-Álvarez, A. and C.A.K. Lovell (2004), "Excess capacity and expense preference behaviour in National Health Systems: an application to the Spanish public hospitals", *Health Economics* 13, 157-169.

Shephard, R.W. (1953), *Cost and production functions*, Princeton University Press, Princeton.

Smet, M. (2007), "Measuring performance in the presence of stochastic demand for hospital services: an analysis of Belgian general care hospitals" *Journal of Productivity Analysis* 27, 13-29.

Williamson O. E. (1963), "Managerial Discretion and Business Behavior" *American Economic Review* 53, 1032-1057.

APPENDIX I

The Lagrangian associated with equation (3) is:

$$L = w^s x + \lambda (1 - D(y, y_z, x, K)) \tag{A.1}$$

where λ is the Lagrange multiplier. From the envelope theorem:

$$\frac{\partial C(y, y_z, w^s, K)}{\partial y_z} = -\lambda \frac{\partial D(y, y_z, x, K)}{\partial y_z} \tag{A.2}$$

To gain further insight into the parameter λ, we begin with the first order conditions for a minimum associated with (A.1):

$$\frac{\partial L}{\partial x} = w^s - \lambda \frac{\partial D(y, y_z, x, K)}{\partial x} = 0 \tag{A.3}$$

$$\frac{\partial}{\partial \lambda} = 1 - D(y, y_z, x, K) = 0 \tag{A.4}$$

Multiplying equation (A.3) by x we obtain:

$$w^s x - \lambda \frac{\partial D(y, y_z, x, K)}{\partial x} x = 0 \tag{A.5}$$

Given that D is homogeneous of degree 1 in x, by Euler's theorem and (A.4) we have:

$$\frac{\partial D(y, y_z, x, K)}{\partial x} x = D(y, y_z, x, K) = 1 \tag{A.6}$$

Moreover, taking into account that w^s is the vector of shadow prices which minimizes costs given (y, y_z, x, K), solving for λ in (A.5) we have that:

$$\lambda = w^s x = C(y, y_z, w^s, K) \tag{A.8}$$

Finally, equation (A.2) can be expressed as follows:

$$\frac{\partial C(y, y_z, w^s, K)}{C(y, y_z, w^s, K)} = -\frac{\partial D(y, y_z, x, K)}{\partial y_z} \tag{A.9}$$

APPENDIX II

The data have been obtained from the statistics of the Spanish Ministry of Health and Consumption (*Instituto Nacional de Gestión Sanitaria*) and the *Instituto Nacional de Estadística* (*INE*). The sample is an unbalanced panel consisting of 385 observations on 67 general hospitals of the INSALUD g*estión-directa* over the period 1987-94 (with an AR1 model, we drop the first observation in each hospital). We have not been able to extend the sample as the EESRI has been modified starting form 1994, the change having involved the elimination of variables which form part of our study. The variables used are defined in Table A.1.[(1)] For more details see Rodríguez-Álvarez and Lovell, 2004.

Table A.1. Definition of variables

VARIABLE	TYPE	DESCRIPTION
MED	Output	Discharges (weighted sum by UPAS) in the following categories: general medical, psychiatric, tuberculosis, long stay, and others.
y_z	Output	Insurance demand, approximated by the residual of the demand equation (3)
AM	Output	Weighted sum by UPAS of outpatient visits (first and successive) and emergencies.
G	Input	Care graduate (doctors, pharmacists and others).
T	Input	Technicians (nurses, matrons, physiotherapists and other qualified personnel).
RES	Input	Other personnel.
S	Input	Expenditure on supplies (deflated).
BED	Quasi-fixed input	Number of beds.
TEACH	Variable capturing complexity	Number of medical students.
D_T	Time variable	Dummy variable for each year.

[(1)] As the input variables are endogenous, we estimated the system using an instrumental variable approach. As instruments, we employ the exogenous variables of the model and the following three variables that we also consider to be exogenous: childbirths, childbirths where the child weighs less than 2.5 kilograms, and the endowment of X-ray rooms of each hospital. To test the validity of using IV estimation, we have applied the Hausman (1978) test of exogeneity. The test statistic is asymptotically chi squared with r degrees of freedom (r = 24 being the number of parameters that are being tested). We find a value of 47.628, which is the highest that would correspond to the critical value corresponding to the conventional levels of confidence. The result of this test shows us, therefore, the aptness of using an instrumental variables approach.