Discussion Paper

Discussion Paper No 92-09

Contact Decision and Frequency Decision: An Econometric Model of the Demand for Ambulatory Services

Winfried Pohlmeier
Volker Ulrich

ZEW

Zentrum für Europäische Wirtschaftsforschung GmbH

Public Finance and Corporate Taxation Series

Contact Decision and Frequency Decision: An Econometric Model of the Demand for Ambulatory Services

by

Winfried Pohlmeier* and Volker Ulrich**

* *ZEW and Universität Mannheim*
** *Universität Mannheim*

June 1992

Abstract
The decision to contact a physician and the decision, how often to contact a physician, are based on different decision makers. We introduce a negative binominal distributed hurdle model that specifies the two stages of the decision process as different stochastic processes, and takes also care of the discrete nature of the data. Empirical results are based on a cross-section of the West-German Socioeconomic Panel. Specification tests reveal, that the two stages of the process have to be treated as distinct processes. Ignoring this distinction leads to serious misinterpretations.

Acknowledgements
The first author gratefully acknowledges financial support by the Deutsche Forschungsgemeinschaft. This study profited from helpful comments by M. Erbsland, D. Harhoff, F. Laisney, A. Mehnert, W. Ried, S. Siddiqui, V. Steiner and K. Winckler. For careful research assistance we like to thank C. Fernandes-Schlegel and D. DeVault. The usual disclaimer applies.

1. Introduction

Not only in Germany but almost world-wide, a relative increase in health care expenditures has taken place during the last two decades (see OECD 1990, p.10). Although health care is assigned a high rank in surveys, it is not so certain from an economic point of view whether the associated social benefit is as high as the opportunity cost. Most industrial nations have undertaken steps towards cost containment so that the rise in cost nearly equals the growth in national product. This orientation of annual increases of expenditures at the growth rate of the incomes of the insured population was implemented in order to stabilize health care premiums. Compared to expenditure analysis, little is known about the development of "prices" and "quantities (services)" which co-determine the corresponding expenditures. With catch phrases such as "cost explosion" being used, one could presume that price increases are responsible for the rise in expenditures. As far as Germany is concerned, it can be shown that the term "cost explosion" is misleading because not only prices but also quantities (benefits) have risen for all types of treatment services (see Ulrich, 1989, p. 131 and Ulrich and Wille 1989, p. 379). The number of an individual's visits to a physician during a certain period is one measurable quantity and plays a central role in the determination of health care expenditures.

This paper investigates the decision process underlying the demand for ambulatory medical services.[1] The decision to use a medical service can be characterized by two features which have been investigated separately in previous studies on health care demand. The first central feature is the two-part character of the decision process (see, e.g., Manning et al. 1981 or Leu and Doppmann 1986). At the first stage, the patient decides whether to visit the physician (contact analysis), whereas it is the physician who essentially determines the intensity of treatment (frequency analysis).[2] The consequence for the econometric specification is that the contact and the frequency analysis must be specified as two different stochastic processes.

1 For demand studies in health care see Grossman 1972 and 1982, Newhouse and Phelps 1974 and 1976, Newhouse 1981, Adam 1983, Breyer 1984, Leu and Doppmann 1986, Wagstaff 1986, Wedig 1988 also Manning et al. 1989.

2 This does not exclude the possibility that the patient herself may have an influence on the lenght of certain treatments. See Feinstein (1977) concerning the issue of compliance.

The second important feature is given by the discrete nature of this decision process. The variable to be explained, "number of physician visits," is, by definition, a discrete dependent variable that can take only non-negative integer values. There are some individuals that, during the survey, had no physician visits whereas others had single or multiple visits. The character of the dependent variable calls directly for the application of count data models. Up to now, applications of count data models for health economic issues have mainly emphasized estimation and/or test procedures and served for expository purposes (Cameron and Trivedi, 1986 and Larcher, 1991).

In this paper we develop a negative binomial distributed hurdle model which not only models the two part character of the decision process but also takes into account the discrete nature of the dependent variable "physician visits". Estimates are based on a subsample of 5,096 individuals from the second wave of the German Socioeconomic Panel (SOEP) for 1985. After presenting some theory which motivates the basic structure of the model in section 2, we derive the Negbin hurdle model in section 3. In the fourth section, we discuss the data and the construction of the variables. Major empirical results as well as the results for a variety of specification tests are reported in section 5, where we distinguish between visits to general practitioners (G.P.) and specialists. The paper concludes with recommendations for future research.

2. Decision phases of the use of medical services

In order to specify the demand for medical care as a two-part process, it seems reasonable to look for models that distinguish between the respective decision maker and the different sets of explanatory variables at each stage of the process. An approach to describe the demand for first visits would have to reflect the decision of the patient to visit the physician. Grossman's (1972) seminal model of the demand for health can be used to describe the contact decision (which is first stage of the process) since it solely results from the patient's utility maximization problem. Health is considered as a durable good that depreciates. By means of net investment, the stock of health capital can be accumulated by combining medical services and other inputs, e.g. time, to produce new health which counters the effects of ageing.

In this context, the demand for medical services is a derived demand, because services are not consumed per se but serve to maintain or improve a certain health status. Because the model reflects the individual's decision with respect to its own health, it is a suitable behavioural model to describe the demand for first contacts, which are almost always initiated by the patient (Wagstaff 1986). The typical form of the individual's demand function for medical services that results from the Grossman model (see Muurinen 1982, p. 10 and Wagstaff 1986, p. 201) is given by:

$$\ln M(t) = \beta_0 + \ln H(t) + \beta_1 \ln w(t) - \beta_2 \ln P_m(t)$$

$$+ \beta_3 t + \beta_4 X_1 + \beta_5 E + u(t).$$

The demand for medical services $(M(t))$ is determined by the latent variable "health status" $(H(t))$, the wage rate $(w(t))$, a price vector for medical services $(P_m(t))$, a time trend (t), a vector of environmental effects (X_1), and the level of education (E). The factor "health" enters the specification with a coefficient equal to unity, which can be explained by the character of health care demand as a derived demand: a higher demand contributes to the production of health. The coefficients on $\ln w(t), \ln P^m(t), t, X_1$ and E reveal the effects of changes in these variables on demand, which does not affect the optimum level of the stock of health capital directly. A higher wage rate leads to a substitution of time by medical services $(\beta_1 > 0)$, because time becomes relatively more expensive. For many western health care systems the price is close to zero so that the impact of $P_m(t)$ is neglected in the empirical analysis.[1] Using the terminology of the Grossman model, the rates of depreciation for the health capital stock increase with rising age, which implies that the β_3 – coefficient is positive. If environmental factors are damaging to health, their impact on $M(t)$ should be positive ($\beta_4 > 0$). Theory predicts a negative coefficient for β_5 if education contributes to a more efficient production of health. However, in structural empirical models that are closely linked to the underlying theory, the signs of many parameters contradict theoretical predictions very often, so that "the majority of the model's structural Parameters are in fact of the 'wrong sign'" (Wagstaff 1986, p. 216).

Models that aim at describing the behaviour at the second stage mainly concentrate on the physician who determines the treatment for each case and then fixes the frequency of repeated visits and/or referrals. The message of behavioural

---

1 In this case variables for time and purchase prices, which reflect the opportunity cost, should bear greater meaning.

models of the second stage is that the physician does not only determine treatment according to medical criteria alone, but also largely reacts to economic incentives. In Zweifel's (1982) model of physicians' behaviour, the physician does not only follow hippocratic goals (e.g. by maximizing the individual's health). The physician is also assumed to maximize utility with respect to his own leisure and consumption. No explicit demand function is derived in this model, but the physician is assumed to affect demand by setting strategic parameters like referrals, the wage rate, and the duration of treatment in order to achieve the best possible solution from his point of view. It can be shown that a decrease in the demand for first contacts reduces the tendency for referrals, i.e. the physician himself will treat more cases. Moreover, physicians with less healthy patients receive a higher wage rate on average. At the same time they spend more time on treatment, which provides weak evidence of the hippocratic oath. Furthermore, the introduction of co-payments in health insurance reduces the patient's demand for first visits. In this case, the model predicts a decrease in the physician's wage rate, and simultaneously, an increase in the time he/she spends per case. This result leads to a re-interpretation of the hypothesis of a "supplier induced demand". The fact that increasing co-insurance rates do not reduce the total cost of treatment seems to give support to the conclusion that physicians are able to create their own demand by producing additional and unnecessary services. By this demand inducement, the physician counters the increasing pressure on his wage rate. In Zweifel's model, though, we find that the decline in the wage rate is offset by an increase in the intensity with which the patient is treated. Therefore, the increase in total cost per case may be the result of ethical behavior and can't be solely explained by the traditional inducement hypothesis.

3. The Negbin Hurdle Model

Basic Models for Count Data

Count data models represent a natural starting point to estimate the demand for ambulatory services, measured as the number of physician visits during a given time interval. Unlike the more popular approaches for qualitative endogenous variables such as logit and probit, which are based on the idea of a continuous threshold-crossing latent dependent variable with an observable binary counterpart, count data approaches assume a dependent variable resulting from an underlying discrete probability function. This implies for our specific application, that we are mainly interested in the explanation of the number of visits to a physician per se and less in explaining the demand for ambulatory services as a latent variable that is proxied by the observable variable 'number of physician visits'.

Since the Poisson distribution has only one parameter implying equidispersion (equality of mean and variance) and hence is too restrictive for most empirical applications, we assume that our dependent variable results from a negative binomial data generating process. This has two main advantages: In the first place, the Negbin distribution is parametrically richer than the Poisson distribution (two parameters). Secondly, it nests the Poisson distribution parametrically. A Negbin process can be derived as a compound Poisson process where the parameter of the Poisson distribution is specified as a gamma distributed variable (gamma compounder). The interpretation of the Negbin distribution as a compound Poisson distribution allows the introduction of a stochastic error term capturing measurement errors and unobserved heterogeneity similar to the error term in the linear regression model.

Assuming a random variable Y, which can take only non-negative integer values $Y \in \{0, 1, 2, \ldots\}$, under the Poisson assumption the probability that exactly y counts are observed is given by:

$$(1) \qquad Pr\{Y = y \mid \lambda\} = \frac{e^{-\lambda}\lambda^{y}}{y!}, \qquad y = 0, 1, 2, \ldots$$

The negative binomial distribution for Y results from the assumption that λ is a gamma distributed random variable:

$$(2) \qquad \lambda \sim Gamma(\phi, v)$$

with:
$$E[\lambda] = \phi, \quad \phi > 0,$$
$$V[\lambda] = \frac{1}{\nu}\phi^2.$$

Integration over λ yields the negative binomial distribution for Y (see Cameron and Trivedi 1986):

$$(3) \qquad Pr\{Y = y\} = \int_0^\infty Pr\{Y = y \mid \lambda\} f(\lambda) d\lambda$$
$$= \frac{\Gamma(y+\nu)}{\Gamma(y+1)\Gamma(\nu)} \left(\frac{\nu}{\nu+\phi}\right)^{\nu} \left(\frac{\phi}{\nu+\phi}\right)^{y}$$

with:
$$E[Y] = \phi \quad , \quad \phi > 0,$$
$$V[Y] = \phi + \frac{1}{\nu}\phi^2 \quad .$$

Exogenous explanatory variables may be introduced by various means through the endogenization of ϕ and/or ν . A simple way to allow for constant over-dispersion has been proposed by Cameron and Trivedi with the Negbin I specification:

$$(4) \qquad E[Y] = \phi = \exp(x'\beta) \quad ,$$
$$(5) \qquad \nu = \frac{1}{\sigma^2}\exp(x'\beta) \quad ,$$

where x is a $k \times 1$ -vector of explanatory variables and β the corresponding parameter vector. Specifying the precision parameter ν as a linear function of the expected value nests the Poisson model and yields a variance function for Y of the following form:

$$(6) \qquad V[Y] = (1+\sigma^2)E[Y] = (1+\sigma^2)\exp(x'\beta).$$

It can easily be shown that under the null hypothesis $\sigma^2 = 0$, the Negbin distribution collapses to the Possion distribution and produces the well-known equidispersion property of the Poisson distribution. However, a test of the null $\sigma^2 = 0$ cannot be performed within the conventional trinity of the LM-, LR- and Wald-Test (or asymptotic t-test, respectively), since under the null the true parameter is on the boundary of the parameter space, which implies that the asymptotic normality property of the ML estimator does no longer hold. To our knowledge there are only a few studies that deal with the problem, for instance the papers by Chernoff (1954), Moran (1971) and Shapiro (1986), see also Lawless (1987). The restrictive assumption that overdispersion equals $(1 + \sigma^2)$ for every individual in the sample can be relaxed by endogenizing σ^2:

$$\sigma^2 = \exp(\bar{x}'\gamma) \quad , \tag{7}$$

This additional parameterization of equation (7) leads to a model of $2k$ parameters to be estimated, which nests the Negbin k - specification by Winkelmann and Zimmermann (1991) as well as the Negbin I- and II- specifications (Cameron and Trivedi 1986).

The Hurdle Specification

Health economic considerations point out that the decision to contact a physician and the decision about the length of the treatment by the physician are based on different decision processes, since the contact decision solely depends on the individual, while the frequency of visits is also based on the supply of therapeutical means by the physician.[1]

This implies that for an appropriate specification of a count data model, the contact decision and the frequency decision have to be treated as separate stochastic processes. The two processes can be driven by the same explanatory variables which, however, may have different interpretations depending on the stage of the decision process. For instance, in the first part of the decision process (contact decision), the explanatory variable 'physician density' represents an availability effect while in the second part of the process it is also likely to reflect competition among physicians and hence supplier induced demand in the narrow sense (i.e. additional demand for health caused by the physician).

Our econometric specification is based on the hurdle model for count data proposed by Mullahy (1986). Unlike Mullahy, we assume that the underlying distribution for both stages is Negbin I. For our specific empirical problem, this extension is important, since supply side effects are rarely well captured in household data at the micro level. Therefore, unobserved heterogeneity has to be accounted for. Our specification allows for explicit testing of distributional assumptions (e.g. against the Poisson or the Generalized Negbin version (7)) and for the equality of the two parts of the decision process.

Apart from its theoretical appeal, which seems relevant for numerous microeconomic applications, the hurdle specification is also interesting from a purely statistical point of view, since it allows over- and underdispersion at the individual level.

The hurdle specification rests on the basic assumption that the data generating process is driven by two sets of parameters. In the case of the Negbin hurdle specification, the binary outcome of the contact decision is governed by a binomial probability model. If the 'hurdle' is crossed and positive counts are observed, $Y \in N_+ = \{1,2,\ldots\}$, the data generating process is governed by a truncated-at-

1 For patients insured in the statutory health insurance system (i.e. more than 80 per cent of the German population) this statement mainly holds for the analysis of GP visits since in general the GP usually acts as a gatekeeper for visits to specialists. Privately insured patients do not face institutional restrictions to visit a GP first before consulting a specialist.

zero count model. For $\theta_1 = (\beta_1', \sigma_1^2)'$ and $\theta_2 = (\beta_2', \sigma_2^2)'$ as the parameter vectors to be estimated, the likelihood function for the hurdle specification is given by:

$$(8) \qquad L = \prod_{i \in \Omega_0} Pr\{y_i = 0 \mid x_i'\beta_1, \sigma_1^2\} \times \prod_{i \in \Omega_1} (1 - Pr\{y_i = 0 \mid x_i'\beta_1, \sigma_1^2\}) \frac{Pr\{y_i \mid x_i'\beta_2, \sigma_2^2\}}{Pr\{y_i \geq 1 \mid x_i'\beta_2, \sigma_2^2\}} \quad ,$$

where the first term in the second product indicates the probability of a contact, while the fraction is the probability of a positive count conditional on a contact. The sets Ω_0 and Ω_1 represent the subsamples of individuals without a visit to a physician and individuals with at least one visit to a physician. Let the binary variable d_i take on the value one if a contact has taken place and zero otherwise, then the likelihood function (8) can be expressed as the product of two parametrically independent likelihood functions:

$$(9) \qquad L = \prod_{i \in \Omega} Pr\{y_i = 0 \mid x_i'\beta_1, \sigma_1^2\}^{1-d_i} (1 - Pr\{y_i = 0 \mid x_i'\beta_1, \sigma_1^2\})^{d_i} \times \prod_{i \in \Omega_1} \frac{Pr\{y_i \mid x_i'\beta_2, \sigma_2^2\}}{Pr\{y_i \geq 1 \mid x_i'\beta_2, \sigma_2^2\}} \quad ,$$

where the first product is the likelihood for the binary process (contact vs. no contact) defined over the total sample Ω, and the second product is the likelihood of a truncated-at-zero Negbin model (defined for the sample of individuals with positive counts). Equation (9) reveals that estimates of the parameter vectors θ_1 and θ_2 can be obtained by separate maximization of the two log-likelihoods. If the two processes are the same ($\theta_1 = \theta_2$), the second term in the likelihood of the binary choice problem (probability of a visit to a positive count) and the denominator in the second likelihood are identical, which leads to the likelihood of a conventional negative binomial count data model. In this case the likelihood is similar to the likelihood of a tobit model. For $\sigma_1^2 = \sigma_2^2 = 0$, the specification reduces to the Poisson hurdle model.

Similar to the probit or the logit model, the binomial process in the first stage can also be interpreted in terms of a threshold-crossing binary choice model, in which the continuous latent variable is the individual's propensity to enter the second stage of the process, i.e., in this specific application, the patient's willingness to visit a physician. It can be shown that the binomial process in the first stage of a Poisson hurdle model is identical to the "Expit" binary choice model proposed by Pohlmeier (1989), which assumes that the latent variable is exponentially distributed. For the Negbin hurdle model, the first stage of the process can be interpreted as a binary choice model with a latent variable that follows a Burr VII distribution (for a proof see Pohlmeier 1992).

4. Data

Our data source is the second wave of the West German Socioeconomic Panel (SOEP) collected in 1985. The dependent variables are the number of visits to a general practitioner and the number of visits to a specialist in the last quarter before the interview. We define the number of visits to a specialist as the number of visits to any physician specialized in a certain field except gynaecology or pediatrics. Explanatory variables are conventional predisposing variables and variables capturing the access to medical services.

We restrict our attention to employed individuals only. Apart from standard questions appearing in every wave of the SOEP this allows us to use detailed information on the working conditions contained in the 1985 cross-section of the SOEP, which are important factors of the individual health status as shown by Pohlmeier and Ulrich (1992) on the basis of a latent structure model.

A distinct treatment of contact decision and demand for ambulatory services, potentially influenced by the physician, requires that variables reflecting the consumption and leisure preferences of the physician have to be included in the second stage of the estimation process. Unfortunately, household surveys, including the SOEP, do not contain information on supply side factors of the health care system. Since there is information on the state of residency, we are able to include variables like the physician density at the state level. For the German system of health care the opportunity cost of physician visits is of specific relevance since the actual pecuniary prices of ambulatory services are negligible. We use the size of the community of residence as a proxy for the opportunity costs of visiting a physician.

The total sample consists of 5,096 observations. For the second stage of estimation, 2,125 observations with a GP contact and 1,640 observations with at least one visit to a specialist are used. Unfortunately, our data do not contain information about the length of a specific treatment episode, which would allow us to distinguish between a true first contact in the time interval to be analyzed and a first contact in the quarter due to a therapy started in the preceding quarter. However, this problem seems to be of minor importance since more than 50 per cent of the individuals in the sample neither had a visit to a GP nor to a specialist. Less than 10 percent of the individuals in the sample had more than three visits to a physician, which indicates that the duration of a treatment usually does not exceed a quarter (see Table A2 in the appendix). For any conclusions concerning potential expenditure effects this interval of observation is particularly useful, since it corresponds to the accounting modalities of the statutory health insurance system. Persons insured by the statutory health insurance system receive on request a sickness voucher by the insurance company each quarter which is the basis for the renumeration of the physician chosen. The referral from a G.P. to a specialist is also based on a voucher, issued by the G.P. This procedure prevents individuals insured by the statutory system from "doctor shopping" at least within one quarter. For our analysis this means that for more than 90 per cent of the individuals in the sample multiple counts do not refer to multiple first contacts due to doctor shopping.

Table 1 informs about some descriptive statistics of variables used in our study, and Table A2 in the appendix contains the quantiles of selected variables. The distribution of the dependent variables, general practitioner visits and specialist visits, is skewed to the right, this means that 75% of the households had no or only one visit to a physician. Only 10% visited a physician three or more times. The median for the age variable is 40 years.

Table 1: Descriptive Statistics a)

Variable	Mean	Standard dev.	Min	Max
General practitioner visits	1.325	3.235	0	70
Specialist visits	1.238	3.368	0	49
Sex	0.368	0.482	0	1
Marital status	0.173	0.378	0	1
Age	39.89	11.13	18	65
Age^2	1715.1	897.7	324	4225
HH-Income	3363.6	2174	2500	60000
Chronic complaints	0.230	0.421	0	1
Private insurance	0.087	0.282	0	1
Education	1.985	3.128	0	40
Physically heavy labor	0.189	0.392	0	1
Stress	0.260	0.439	0	1
Variety on job	0.517	0.500	0	1
Self-determining	0.363	0.481	0	1
Control	0.170	0.370	0	1
Pop. < 5000	0.124	0.330	0	1
Pop. 5000-20000	0.247	0.431	0	1
Pop. 20000-100000	0.275	0.446	0	1
Physician density	0.260	0.049	0.212	0.466
Months of unemployment	0.240	1.300	0	12
Hospitalized >7 days	0.074	0.261	0	1
Sick leave >14 days	0.183	0.367	0	1
Degree of disabl. >20 %	0.058	0.234	0	1

a) Employees only; total number of observations: 5096; for the estimations in the second stage there remain 2125 observations with positive counts for the GP equation and 1640 observations for the specialist equation.

5. Results

Estimates and tests of the two stage model for the use of ambulatory services are based on two different explanatory variables: (i) the number of visits to a general practitioner and (ii) the number of visits to a specialist. All equations were estimated by maximum likelihood. Robust standard errors were computed for the estimated coefficients, so that estimates can be interpreted within the pseudo maximum likelihood methodology.[1] The nested models (Poisson hurdle model, Negbin model and Poisson model) were also estimated for both samples so that the nested versions could be used to test the Negbin hurdle specification using the Wald- and the Hausman principle. For tests against the Poisson distribution ($\sigma^2 = 0$) within the hurdle model, the Hausman test is of special attractiveness since it is based on the β-parameters and thus circumvents the boundary problem.[2] Appendices A3 and A4 survey the results of the specification tests. The hurdle specification can neither be rejected for the G.P. equation nor for the specialist equation (see the test vaules for the Negbin Hurdle model versus the Negbin model on the right branch of the diagram). The surprisingly high χ^2-values reveal that the contact and the frequency decision come from different models and have to be modelled separately. Mixing up both decision stages within one regression equation, which seems almost unavoidable with aggregate data due to the limited informational content of the dependent variable, leads to inconsistent estimates. The assumption of the Poisson distribution at the first stage cannot be rejected. This result does not primarily reflect the absence of unobserved components but rather the small informational content of the binary dependent variable in the first stage, which does not admit to discriminate between the Negbin and the Poisson distribution. Monte Carlo experiments by Blundell et al. (1989), substantiate that, with increasing information about the dependent variables, the diagnostic tests become more powerful. On the contrary, the Poisson distribution with t-values for σ_2^2 of 12.0 and 12.5 must be rejected at the second stage.

1 The variance-covariance matrix of the estimated coefficient vector is computed as $H^{-1} G H^{-1}$, with H as the Hessian of the Likelihood function (8) and G as the cross-product of the gradient evaluated at the maximum.

2 The Hausman test is not applicable to test the simple Negbin model against the Poisson since the Poisson model belongs to the linear exponential family, which implies consistent maximum likelihood estimates under the alternative.

Table 2 contains the results of the negative binomial distributed hurdle model for visits to a general practitioner and Table 3 the results for visits to a specialist. The first stage analyzes the contact decision. Estimates of the following additional use are contained in the second stage. Both tables also contain estimates of a more parsimonious specification for the second stage where some insignificant variables are neglected. Generally, the coefficients for the first stage are more precisely determined. Besides the substantial reduction in sample size by more than 58 and 67 per cent, one reason for this result is the type of explanatory variables in the SOEP which covers a broad spectrum of the socioeconomic background of the individual and allows a satisfactory description of first contacts. Major determinants of multiple visits like competition among physicians, can only be proxied by variables such as city population and physician density. In addition, various important variables, e.g. those which reflect the income and leisure goals of the physician, are neglected in the second stage. Moreover, there is no large variance of the dependent variable: During the survey quarter more than 90 percent of the individuals did not visit a general practitioner or a specialist more than three times (see Appendix A2).

In addition to the results of the specification tests, the estimation results also point to important differences of the two decision stages. Empirical approaches which ignore these differences do not only lead to inconsistent estimates, but are also bound to cause serious misinterpretations. To clarify this point, Appendix A5 contains estimates of the Negbin model where the distinction between the two stages of the decision process is neglected. For the simple Negbin estimates, the physician density has no significant impact on the number of G.P. visits but is significant in the specialist equation. Taking care of the decision process by the hurdle specification shows that the physician density has a significant impact on the number of G.P. visits (columns 4 and 6 in Table 2) while it has no significant impact on the contact decision itself. We interpret this interesting finding as an evidence for supply induced demand for general practitioners' ambulatory services, since the physician density only proxies the degree of competition of physicians on the second stage.

The Negbin estimates of the specialist equation yield a significantly positive coefficient on the physician density. However, since both stages of the decision process are entangled in one equation, the finding is hardly interpretable in terms of the true underlying causality. The hurdle estimates indicate that this finding is clearly dominated by the first stage process.

Table 2: Negative binomial distributed Hurdle model
Dependent variable: Number of visits to a G.P.

	1. Stage		2. Stage		2. Stage (reduced)	
Variable	coef.	(t-value)	coef.	(t-value)	coef.	(t-value)
Constant	0.63	(1.0)	-1.40	(-0.9)	-3.14	(-4.5)
Sex	0.18	(3.7)	0.22	(1.4)	0.19	(1.2)
Marital status	-0.14	(-1.8)	-1.24	(-1.5)	-0.83	(-1.4)
Age * 10^{-1}	-0.50	(-3.2)	-0.55	(-0.9)	0.27	(2.7)
Age^2 * 10^{-3}	0.73	(3.9)	0.88	(1.3)	-	-
HH-income * 10^{-4}	-0.47	(-2.2)	0.22	(1.3)	0.21	(1.1)
Chronic complaints	0.45	(8.4)	1.45	(3.9)	1.44	(3.7)
Private insurance	-0.20	(-2.1)	-0.56	(-1.3)	-0.61	(-1.5)
Education	-0.02	(-2.2)	-0.01	(-0.2)	-0.01	(-0.2)
Physically heavy labour	0.12	(2.2)	0.35	(2.0)	0.35	(2.1)
Stress	0.12	(2.3)	-0.02	(-0.1)	0.03	(0.2)
Variety on job	-0.11	(-2.2)	-0.19	(-1.3)	-0.21	(-1.4)
Self-determining	-0.08	(-1.6)	-0.50	(-2.4)	-0.51	(-2.4)
Control	0.08	(1.3)	0.16	(0.8)	-	-
Pop. <5000	0.46	(6.2)	0.14	(0.5)	0.14	(0.5)
Pop. 5000-20000	0.43	(7.0)	0.25	(1.1)	0.24	(1.1)
Pop. 20000-100000	0.20	(3.2)	0.22	(1.0)	0.21	(0.9)
Physician density	-0.22	(-0.4)	3.15	(2.4)	3.05	(2.4)
Months of unemployment	-0.05	(-2.5)	-0.08	(-0.5)	-0.05	(-0.4)
Hospitalized >7	0.23	(2.7)	0.03	(0.1)	0.06	(0.3)
Sick leave >14	0.35	(5.5)	0.97	(4.5)	0.97	(4.5)
Degree of disabil.>20	0.09	(0.9)	0.12	(0.7)	0.15	(0.9)
σ^2	2.49	(0.6)	4.08	(12.0)	4.09	(12.0)
log Lik partial	-3242		-3945		-3947	
log Lik total			-7187		-7189	
n	5096		2125		2125	

Like sex (female = 1), marital status (single = 1) is only significant at the first stage indicating that males and singles are more reluctant to contact the physician. But their behaviour is not different provided that a contact has taken place. The impact of age on demand is captured by the variables age and age^2. Age shows a convex relationship for both contact equations indicating that the probability of contacting a physician first decreases with age and then rises. However, for the specialist equation the relationship is only close to significance. Holding all other independent variables fixed, the minimum probability of a contact with a G.P. is reached at age 34. Both age variables are insignificant at the second stage. This unexpected result may be due to an overparameterization of the age effect. Hence, we estimated a second stage equation dropping all insignificant variables except the linear age term. In this more parsimonious specification of the second stage, the age effect turns out to be significant only for the G.P. equation. As was mentioned before, neglecting the two-part character of the decision process leads to a wrong conclusion: The length of treatment by a G.P. increases with age. However, the probability of contacting a G.P. reveals a convex relationship with respect to age. Somewhat surprisingly, there is no evidence that the duration of a treatment by a specialist increases with age.

The impact of household income (negative sign) is only significant at the first stage. High income earners visit a general practitioner less frequently. When the general practitioner treats only trifling health problems, opportunity cost might play a large part. High income earners show fewer visits to a general practitioner because they face higher opportunity costs. There are controversial opinions on the sign of the income effect. Van de Ven and van der Gaag (1982) find a negative indirect effect of income on demand. A high income results in a higher demand (direct effect), but on the other hand this leads to a higher level of health which reduces demand (negative indirect effect) so that the total effect is a priori undetermined.

The coefficient on the variable 'chronic complaints' reveals at all stages the expected positive and significant sign. The binary variable for the type of medical insurance is only significant at the first stage. Households that are privately insured have fewer first visits to a general practitioner. This fact provides at least a clue that deductibles and the cost refund principle have a measurable influence on demand and contradicts the thesis of an ineffective demand side incentive mechanism. Possibly, the deductibles for privately insured patients seem to make these patients more reluctant to contact a G.P., if minor health complaints occur. However, we do not find any evidence that the length of a treatment - which is approximately measured here as the number of visits to a physician - is different for privately and publicly insured patients. The variable education is only significant at the first stage. Education could eventually correlate with medical knowledge so that a higher educated person tends to favour specialists over general practitioners. It also seems possible that education increases the "non-market productivity" (Wagstaff 1986, p.216) , i.e. people with a higher education can improve their health more efficiently and therefore have lower marginal cost of health.

Except for the variable "self-determination", the working place variables are all significant for the G.P. equation at the first stage with theoretically expected signs. In comparison to the impact of other binary variables, the effect of working conditions on the contact probability is weak and does not play any part in the second stage of the G.P. equation. Satisfaction with job conditions, which are reflected in little control, a high degree of self-determination, and a lot of variety, also leads to fewer physician visits. The impact of other variables reflecting the working conditions such as shift work, night shift, and hazardous environmental conditions, were tested using the LM-Test. None of these variables turned out to have a significant impact, with the same holding true for work experience.

If job condition variables represent decisive determinants of individual health and for the demand for ambulatory services, a statistical confirmation at the second stage (and eventually for the demand for specialist services) would be expected. Nevertheless, we leave it to the reader to interpret our results with regard to the working place characteristics as determinants for one to go on the sick.[1] Again, the results of the simple Negbin estimation (appendix A5) lead to the misleading interpretation that working conditions are significant determinants of health status.

The estimates for the three dummy variables for community size are based on the reference category "community with at least 100,000 residents". In comparison, households in small and middle sized cities visit a general practitioner more often. It is not surprising that individuals who were seriously ill in the previous year (hospitalized > 7, sick leave > 14) do require more treatment by general practitioners and specialists. Hospitalization mainly affects the contact probability. Employees who faced unemployment in the previous year are more reluctant to contact the G.P. After a contact has taken place, we cannot find any significant behavioural differences between the previously unemployed and the reference group.

In discussing the estimation results for the specialist equation, we focus on the differences compared to the equation for a general practitioner (Table 3). Household income has a positive coefficient at the first stage, i.e., a higher income

1 In order to obtain sick leave payments, employees who are sick for more than 3 days are required to present the employer a notice of illness from a physician.

enables more first contacts with specialists. As was mentioned before, theory does not preclude a positive relation between income and demand. In contrast to patients with statutory medical insurance, private patients can choose their general practitioner and specialists freely, so that the "gatekeeper"-function of the general practitioner is irrelevant. The coefficient of education is, as could be expected, positive at the first stage.

Variables describing the patient's working conditions are of no importance to explain the demand for specialist at either stage. It is more likely that they represent the dissatisfaction with the working conditions and the resulting contacts to a general practitioner rather than pointing to a specific illness. At both stages there is a positive connection between large city and specialist visits. In comparison to large cities, the small and middle cities have fewer specialists. At the first stage physician density is significant. Both variables express the overproportionate supply of specialists in the large city as well as the decreasing opportunity cost of use.

Table 3: Negative binomial distributed Hurdle model
Dependent variable: Number of specialist visits

	1. Stage		2. Stage		2. Stage (reduced)	
Variable	coef.	(t-value)	coef.	(t-value)	coef.	(t-value)
Constant	-0.43	(-0.5)	-1.76	(-0.7)	-1.20	(-1.6)
Sex	0.74	(14.3)	0.35	(1.7)	0.35	(1.8)
Marital status	-0.24	(-2.8)	-0.64	(-1.0)	-0.64	(-1.1)
Age * 10^{-1}	-0.33	(-1.8)	0.32	(0.3)	0.11	(1.1)
Age^2 * 10^{-3}	0.33	(1.5)	-0.24	(-0.2)	-	-
HH-income * 10^{-4}	0.22	(2.3)	0.16	(1.1)	0.17	(1.3)
Chronic complaints	0.76	(13.0)	0.97	(2.2)	1.05	(2.2)
Private insurance	0.17	(1.8)	-0.37	(-0.7)	-0.25	(-0.6)
Education	0.02	(2.8)	-0.06	(-0.7)	-0.05	(-0.9)
Physically heavy labour	-0.03	(-0.4)	-0.21	(-0.8)	-	-
Stress	0.05	(0.8)	0.22	(1.0)	-	-
Variety on job	0.09	(1.7)	0.21	(0.9)	-	-
Self-determining	0.06	(1.1)	-0.13	(-0.5)	-	-
Control	0.02	(0.3)	0.13	(0.6)	-	-
Pop. <5000	-0.53	(-5.7)	-0.91	(-1.9)	-0.87	(-1.8)
Pop. 5000-20000	-0.37	(-5.3)	-0.30	(-1.0)	-0.34	(-1.3)
Pop. 20000-100000	-0.21	(-3.2)	-0.31	(-1.1)	-0.31	(-1.2)
Physician density	0.93	(1.9)	0.42	(0.3)	0.38	(0.3)
Months of unemployment	-0.01	(-0.4)	-0.00	(-0.0)	-	-
Hospitalized >7	0.18	(1.8)	0.25	(1.1)	0.24	(1.1)
Sick leave >14	0.37	(5.1)	1.13	(2.9)	1.09	(2.7)
Degree of disabil.>20	0.25	(2.5)	0.13	(0.5)	0.14	(0.6)
σ^2	2.15	(0.5)	5.69	(12.5)	5.71	(13.1)
log Lik partial	-2885		-3415		-3417	
log Lik total			-6299		-6302	
n	5096		1640		1640	

6. Conclusions

On the basis of a negative binomial hurdle model, the determinants of the demand for medical services as measured by the number of physician (general practitioner and specialist) visits in one quarter are estimated. Using these dependent variables within a count data framework permits the analysis of a real quantity variable and, in contrast to an econometric model of discrete choice, an exact identification of the model parameters. Moreover, the hurdle approach proposed allows us to separate and quantify the determinants of medical demand with regard to contact and frequency decisions.
Without repeating all empirical results, at this point, we would like to emphasize only the results that should be considered for future empirical approaches to analyse the demand for ambulatory medical services:

(i) Contact and frequency decisions for general practitioners as well as for specialists are different stochastic processes and are to be modelled separately. Ignoring these differences leads to inconsistent parameter estimates and to economic misinterpretations.

(ii) Household data are well suited to quantify the determinants of the contact decision. Since information on supply side aspects is limited, estimates for the second stage suffer from unobserved heterogeneity. Not surprisingly, specification tests call for stochastic specifications that are flexible with respect to unobserved heterogeneity and missing variables.

(iii) As we mentioned, the estimation results for the first stage are qualitatively similar to the results of previous studies where no hurdle specification is made and other estimation procedures are applied. Hence, we may conclude that the estimates of empirical approaches that do not distinguish between the two stages are dominated by the first stage process. Because of the opportunity to use real quantity variables as the dependent variables, count data models should be seen as an alternative to discrete choice models.

(iv) The findings reveal some empirical evidence in favour of supplier induced demand in the case of ambulatory services provided by general practitioners: The variable physician density is insignificant in the first stage where it solely reflects the patients opportunity costs to contact the physician. In the second stage, where

the variable also proxies competition among physicians, we find a positive impact. Failure to find empirical support for the hypothesis in similar studies might be due to an inappropriate modelling of the microeconomic decision process.
Our results indicate two important aspects for future health economic research. From a theoretical point of view, an attempt should be made to combine models of the demand for first contacts with models that describe the behaviour of physicians (e.g. Zweifel 1982), in order to find an adequate unified theoretical representation of the demand for health care. Regarding the empirical modelling, it appears to us that the development and application of appropriate hurdle models for panel data are of special interest. This would offer a more appropriate treatment of the unobserved components. Panel estimators for count data like those proposed by Hausman, Hall and Griliches (1984) should enable future research to separate individual differences in health endowments from physician induced differences in the frequency of visits.

References

Adam, H. (1983): *Ambulante ärztliche Leistungen and Ärztedichte. Zur These der anbieterinduzierten Nachfrage im Bereich der ambulanten ärztlichen Versorgung*, Berlin.

Blundell, R., S. Peters and R.Smith (1989): "The Impact of Grouping on Non-Normality Diagnostics", University College London.

Breyer, F. (1984): Die Nachfrage nach medizinischen Leistungen. Eine empirische Analyse von Daten aus der Gesetzlichen Krankenversicherung, Berlin et al.

Cameron, A. C. and P.K. Trivedi (1986): "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests", *Journal of Applied Econometrics*, 1, 29-53.

Chernoff, H. (1954): "On the Distribution of the Likelihood Ratio", *Annals of Mathematical Statistics*, 25, 573-578.

Feinstein, A. R. (1977): "Consequences of 'Compliance Bias'", in Clinical Biostatistics, A. R. Feinstein (Ed.), St. Louis, 122-133.

Grossman, M. (1972): The Demand for Health - A Theoretical and Empirical Investigation, New York.

Grossman, M. (1982): "The Demand for Health After a Decade", *Journal of Health Economics*, 1, 1-3.

Hausman, J., B. H. Hall and Z. Griliches (1984): "Econometric Models for Count Data with an Application to the Patents - R & D Relationship", *Econometrica*, 2, 909-938.

Larcher, B. (1991): "A Poisson Process of Which the Parameter Contains a Non-Stationary Error. Application to the Analysis of a Series of Deaths in a Large Hospital", in *Econometrics of Health Care*, G. Duru and J.H.P. Paelinck (Eds.), Dodrecht et al.

Lawless, J.F. (1987): "Regression Methods for Poisson Data", *Journal of the American Statistical Association*, 82, 808-815.

Leu, R. E. and R. J. Doppmann (1986): "Gesundheitszustandsmessung und Nachfrage nach Gesundheitsleistungen in der Schweiz", in *Informations- und Planungsprobleme in öffentlichen Aufgabenbereichen*, E. Wille (Hrsg.), Frankfurt/M. et al., 1-90.

Manning, W. G. et al. (1981): "A Two-Part Model of the Demand for Medical Care: Preliminary Results from the Health Insurance Study", in *Health, Economics, and Health Economics,* J. van der Gaag and M. Perlman (Eds.), Amsterdam et al., 103-123.

Manning, W. G. et al. (1989): "The Taxes of Sin: Do Smokers and Drinkers Pay their Way?", *Journal of the American Medical Association, 261,* 1604-1609.

Moran, P.A.P. (1971): "Maximum-Likelihood Estimation in Non-Standard Situations", *Proceedings of the Cambridge Philosophical Society,* 70, 441-450.

Mullahy, J. (1986): "Specification and Testing of some Modified Count Data Models", *Journal of Econometrics*, 33, 341-365.

Muurinen, J.-M. (1982): "Demand for Health. A Generalised Grossman Model", *Journal of Health Economics*, 1, 3-28.

Newhouse, J. P. (1981): "The Demand for Medical Care Services: A Retrospect and Prospect", in *Health, Economics, and Health Economics,* J. van der Gaag and M. Perlman (Eds.), Amsterdam et al., 85-102.

Newhouse, J. P. and C. E. Phelps (1974): "Price and Income Elasticities of the Demand for Medical Care Services", in *The Economics of Health and Medical Care,* M. Perlman (Ed.), London, 139-161.

Newhouse, J. P. and C. E. Phelps (1976): "New Estimates of Price and Income Elasticities of Medical Services", in *The Role of Health Insurance in the Health Care Sector,* R. N. Rossett (Ed.), New York, 261-309.

OECD (1990): *Health Care Systems in Transition, The Search for Efficiency*, Paris.

Pohlmeier, W. and V. Ulrich (1991): "Determinanten des Gesundheitszustands: Ein empirischer Ansatz zur Outputmessung im Gesundheitswesen bei partieller Information", *Zeitschrift für Wirtschafts- und Sozialwissenschaften*, forthcoming.

Pohlmeier, W. (1989): Simultane Probit- und Tobitmodelle: Theorie und Anwendungen auf Fragen der Innovationsökonomik, Heidelberg et. al.

Pohlmeier, W. (1992): Notes on the Interpretation and Identification of Count Data Hurdle Models, Deparment of Economics, University of Mannheim, mimeo.

Shapiro, A. (1986): "Asymptotic Theory of Overparameterized Structural Models", *Journal of the American Statistical Association*, 81, 141-149.

Statistisches Bundesamt (1986): Fachserie 12, Gesundheitswesen, Reihe 1, Stuttgart und Mainz.

Ulrich, V. (1989): Preis- und Mengeneffekte im Gesundheitswesen, Frankfurt/M. et al.

Ulrich, V. and E. Wille (1989): "Zur Ausgabenentwicklung der gesetzlichen Krankenversicherung: Preis- und Mengeneffekte von Behandlungsarten", *Finanzarchiv*, NF, 47 (3), 361-403.

van de Ven, W. P. M. M. and J. van der Gaag (1982): " Health as an Unobservable: A MIMIC Model of the Demand for Health Care", *Journal of Health Economics,* 1, 157-183.

Wagstaff, A. (1986): "The Demand for Health. Some New Empirical Evidence", *Journal of Health Economics,* 5, 195-233.

Wedig, G. J. (1988): "Health Status and the Demand for Health", *Journal of Health Economics,* 7, 151-163.

Winkelmann, R. and K.F. Zimmermann (1991): "A New Approach for Modelling Economic Count Data", *Economics Letters*, 37, 139-143.

Zweifel (1982): "Ein ökonomisches Modell des Arztverhaltens", Berlin et al.

A 1: Description of variables

General practitioner visits:
	Number of visits to a general practitioner in the last quarter (before time of survey).
Specialist visits:
	Number of visits to a specialist in last quarter (before time of survey) with exception of visits to gynaecologist, pediatrician or dentist medicine.
Sex:
	0 = Male; 1 = Female.
Marital status:
	0 = other; 1 = single.
Household income:
	net monthly income, after income and social taxes; including regular monthly subsidies such as housing allowances and other transfers. -
Chronic illness:
	Chronic illness or complaints for at least one year, yes = 1, no = 0.
Age:
	Age in years, survey year minus year of birth.
Private medical insurance:
	Private medical insurance in the last year; yes = 1, other = 0.
Education:
	Number of years of education beyond age sixteen.
physically heavy labour job:
	Person has a position where heavy physical work is required; agree completely = 1, other = 0.
Stress:
	Position involves a high level of stress; agree completely = 1, other = 0.
Variety on the job:
	Job has a lot of variety: agree completely = 1, other = 0.
Self-determining :
	Individual can plan and carry out his job tasks; agree completely = 1, other = 0.
Control:
	Work performance is strictly controlled; agree completely = 1, other = 0.
Population < 5000
	Place of residence has less than 5000 residents = 1, other = 0.
Population 5000 - 20000
	Place of residence has between 5000 and 20000 residents = 1, other = 0.
Population 20000 - 100000
	Place of residence has between 5000 and 20000 residents = 1, other = 0.
Physician density:
	Number of licensed physicians per 100000 residents in the place of residence, Source: Statistisches Bundesamt (1986).
Months of unemployment:
	Duration of unemployment in months in the previous year.

Hospitalized:

More than 7 days hospitalized in previous year, yes = 1, other = 0.

Sick leave:

Missed more than 14 work days because of sickness in previous year, yes = 1, other = 0.

Degree of disability:

Degree of disability is greater than 20% = 1, other = 0.

A 2: Quantiles of some selected variables

Variable	10%	25%	50%	75%	90%
G.P. visits	0	0	0	1	3
Specialist visits	0	0	0	1	3
Age	24	31	40	48	55
Household income	1800	2300	3000	4000	5000

A 3: Specification Tests within the Negbin Hurdle Model: GP equation [1]

	null hypothesis			
alternativ hypothesis	P_1-Nb_2	Poisson Hurdle	Negbin	Poisson
Negbin Hurdle	$W_{(1)}=0.3$ $H_{(22)}=0.8$		$W_{(23)}=86.9$ $H_{(23)}=136.0$	
P_1-Nb_2		$W_{(1)}=144.4$ $H_{(1)}=125.8$		
Poisson Hurdle				$W_{(22)}=1487$ $H_{(22)}=4756$
Negbin				$W_{(1)}=334.6$

[1] $W_{(k)}$-Wald test with k degrees of freedom, $H_{(k)}$-Hausman test with k degrees of freedom. P_1-Nb_2 is a hurdle model with Poisson assumption in the first stage and Negbin assumption in the second stage. The H-test for Nb-hurdle vs P_1-Nb_2 is based on β_1. H-tests for Poisson hurdle vs Poisson are based on β_2 and Negbin hurdle vs Negbin on θ_2 respectively.

A 4: Specification Tests within the Negbin Hurdle Model: Specialists equation [1]

	null hypothesis			
alternativ hypothesis	P_1-Nb_2	Poisson Hurdle	Negbin	Poisson
Negbin Hurdle	$W_{(1)}=0.3$ $H_{(22)}=s$		$W_{(23)}=56.5$ $H_{(23)}=64.6$	
P_1-Nb_2		$W_{(1)}=156.4$ $H_{(1)}=92.4$		
Poisson Hurdle				$W_{(22)}=2251$ $H_{(22)}=s$
Negbin				$W_{(1)}=792.1$

[1] $W_{(k)}$-Wald test with k degrees of freedom, $H_{(k)}$-Hausman test with k degrees of freedom. P_1-Nb_2 is a hurdle model with Poisson assumption in the first stage and Negbin assumption in the second stage. The H-test for NB-hurdle vs P_1-Nb_2 is based on β_1 . H-tests for Poisson hurdle vs Poisson are based on β_2 and Negbin hurdle vs Negbin on θ_2 respectively. S-Difference of the covariance matrices for the Hausman test is singular.

A 5: Estimate of the Negbin I-Model

	general practitioners		specialists	
Variable	coef.	(t-value)	coef.	(t-value)
Constant	0.57	(1.7)	-0.04	(-0.1)
Sex	0.18	(4.1)	0.70	(14.2)
Marital status	-0.17	(-2.4)	-0.25	(-3.4)
Age * 10^{-1}	-0.50	(-3.5)	-0.34	(-2.4)
Age^2 * 10^{-3}	0.74	(4.3)	0.37	(2.1)
HH-income * 10^{-4}	-0.40	(-2.0)	0.14	(2.2)
Chronic complaints	0.54	(11.0)	0.78	(14.0)
Private insurance	-0.22	(-2.5)	0.15	(1.7)
Education	-0.02	(-2.5)	0.02	(2.4)
Physically heavy labour	0.14	(2.7)	-0.03	(-0.4)
Stress	0.10	(2.1)	0.06	(1.0)
Variety on job	-0.11	(-2.4)	0.10	(1.9)
Self-determining	-0.13	(-2.8)	0.05	(0.9)
Control	0.08	(1.4)	0.04	(0.6)
Pop. <5000	0.44	(6.4)	-0.56	(-6.3)
Pop. 5000-20000	0.41	(7.3)	-0.37	(-5.6)
Pop. 20000-100000	0.20	(3.5)	-0.22	(-3.6)
Physician density	0.32	(0.6)	0.91	(2.1)
Months of unemployment	-0.05	(-2.6)	-0.01	(-0.7)
Hospitalized >7	0.19	(2.4)	0.20	(2.1)
Sick leave >14	0.43	(7.4)	0.44	(6.3)
Degree of disabil.>20	0.11	(1.4)	0.24	(2.6)
σ^2	3.42	(14.9)	5.36	(16.8)
log Lik	-7242		-6329	
n	5096		5096	