MANNHEIM RESEARCH INSTITUTE FOR THE ECONOMICS OF AGING

MODELLING SOCIOECONOMIC AND HEALTH DETERMINANTS OF HEALTH CASE USER: A SEMIPARAMATRIC APPROACH

Jürgen Maurer

145-2007

L13, 17_D-68131 Mannheim_Phone +49 621 181-2773/1862_Fax +49 621 181-1863_www.mea.uni-mannheim.de

Modelling socioeconomic and health determinants of health care use: A semiparametric approach

Jürgen Maurer[*]

March 2007

Abstract

This paper suggests bivariate semiparametric index models as a tool for modelling the interplay of socioeconomic and health characteristics in determining health care utilisation. These models allow for a fully nonparametric relationship between socioeconomic status, health care need and care utilisation. The only parametric restriction imposed is that multiple socioeconomic and health indicators can be aggregated into two distinct indices that measure the broader concepts of socioeconomic status and health care need, respectively. We demonstrate the usefulness of this class of models based on an illustrative empirical example. The estimations highlight complex interactions of socioeconomic status and health care need in determining care use, which may be difficult to grasp via standard parametric modelling approaches.

Keywords: Semiparametric Methods, Health Care Utilisation

[*] Mannheim Research Institute for the Economics of Aging (MEA), University of Mannheim, L13, 17, 68131 Mannheim, Germany. Email: juergen.maurer@iue.it. I would like to thank Roger Klein and Frank Vella for sharing some of their GAUSS routines with me. I would also like to thank Martin Salm, Emily Sinnott, seminar participants at the University of Mannheim, participants of the 15th European Workshop on Econometrics and Health Economics in Thessaloniki (especially my discussant, Eddy van Doorslaer), the editor, Andrew Jones, as well as an anonymous referee for very helpful comments and suggestions. Of course, all remaining errors are my own responsibility.

1 Introduction

Empirical assessments of the effects of socioeconomic status (SES) on health care use are of great interest in the health policy sphere, as ensuring socioeconomic equity of the health care system is often considered a high priority. Since health care is instrumental for improving and maintaining individual health and functioning, a lack of adequate medical attention among the socially disadvantaged may cause or at least exacerbate socioeconomic gradients in health.[1] However, not all individuals require the same intensity of health care. The amount of care delivered to a specific individual should rather depend on her own health characteristics and risk factors as well as on the availability of (cost-effective) treatment technology for improving or managing these. The latter consideration highlights the important distinction between health care need on the one hand and common notions of ill-health on the other. Particularly, it permits "the non-ill to be said to be in need of medical care, in the sense that their health in the future could be better than it would otherwise be if they received (preventive) care now" (Wagstaff and van Doorslaer (2000), p.1813). The allocation of health care resources should therefore appropriately reflect individual differences in health care need, but not be dependent on the patients' socioeconomic status.

The distinction between health care need and socioeconomic status as legitimate and illegitimate sources of differences in medical care utilisation is central to almost all empirical studies of equity in health care utilisation.[2] Although this conceptual dichotomy is omnipresent in equity assessments of health care delivery, modelling it in a flexible, yet parsimonious way poses several challenges for the applied researcher. Parametric models may often appear too restrictive to incorporate fully flexible SES-need-interactions, as they usually impose strong a priori assumptions on permissible functional forms. Without extensive prior knowledge, nonparametric approaches seem better suited for modelling the interplay of SES and health care need with respect to care use. Yet, actual applications of these methods are frequently impractical. Modelling SES and health care need typically requires the use of multiple indicators to capture the multi-facetted nature of either concepts, and nonparametric methods have well-known difficulties in handling high-dimensional problems.

This paper suggests the use of bivariate semiparametric index models as a potentially powerful tool for modelling the dichotomy of SES and health-care need in health care

[1] A survey of the extensive literature on the health-SES nexus is beyond the scope of this paper. Adler and Newman (2002), Adler et al. (2000), Deaton (2003), Marmot (2005), Marmot and Wilkinson (1999) and Smith (1999, 2003, 2004) provide excellent introductions to key aspects of the subject.

[2] See e.g. Wagstaff and van Doorslaer (2000) for an overview.

delivery. A blend of parametric and nonparametric approaches, semiparametric index models combine the two in order to mitigate the dimensionality issues of nonparametric modelling, while maintain its flexibility. Specifically, while this class of models places some parametric structure on the constituents of SES and care need respectively, it retains a fully nonparametric approach with respect to how these concepts may interact to determine care use. The model is therefore especially well-suited for assessing observable heterogeneity in the effects of SES on health care use across different levels of care need.[3]

The reminder of the paper is organized as follows: Section 2 reviews parametric and nonparametric strategies for modelling the SES and care need interactions with respect to health care use, and highlights their respective advantages and disadvantages.[4] Section 3 then tries to strike a balance between these considerations by introducing the class of bivariate semiparametric index models. Section 4 brings the proposed approach to life by means of a simple empirical illustration. Specifically, we use a bivariate semiparametric model to analyse the effects of education and health care need on the average number of yearly doctor visits of older Italian men. The section describes how such a model can be estimated and highlights some of the most important findings revealed by this modelling strategy. Section 5 concludes with a short summary of the paper and potential directions for future research.

2 Modelling SES and health care need interactions

This section provides a brief conceptual review of alternative approaches for modelling the SES-need dichotomy in care use. We first consider parametric approaches for modelling the conditional expectations of care use m_i given multiple measures of SES S_i and health care need H_i.[5] We then move to nonparametric generalizations of these approaches, as these do not rely on any a prior restrictions regarding the way S_i and

[3] Manski (2005) notes, for example, the importance of assessing the effects of *observable* heterogeneity for policy targeting. This is due to the fact that policy makers can usually only discriminate between individuals based on *observable* characteristics in assigning different treatments.

[4] This section gives a review of various well-known modelling strategies and considers some peculiarities of applications in health economics. A more extensive overview of different assumptions implied by parametric, semiparametric and nonparametric modelling approaches can be found in Powell (1994) among others.

[5] Throughout the paper, our discussion will focus on modelling the conditional expectations $E\left[m_i|S_i, H_i\right]$. We note, however, that this is only done for expositional ease and because most models concentrate on the conditional mean as their main parameter of interest. In fact, the approaches suggested here can also be generalized to other features of care use - such as access to care - that may be of independent interest (see e.g. Maurer (2007) for an example).

H_i are allowed to affect m_i. However, fully nonparametric models are often difficult to employ in practice. Their application to multi-dimensional data requires very large samples, and these may not available in many applications.

2.1 Parametric models

In parametric approaches, the researcher assumes a functional form for the relationship between health care use m_i, SES S_i, and health care need H_i, which is determined a priori up to some finite-dimensional parameter vector β. This single parameter β fully characterizes a specific data generating process within the supposed parametric family. A parametric specification for the conditional expectation of m_i, given S_i, and H_i will therefore have the form

$$E\left[m_i|S_i, H_i\right] = m\left(S_i, H_i, \widetilde{\beta}\right) \tag{1}$$

where $m(\cdot,\cdot,\cdot)$ denotes a fixed deterministic family of regression functions assumed by the researcher and $\widetilde{\beta}$ a free parameter to be estimated from the data.[6] Of course, the form of $m(\cdot,\cdot,\cdot)$ may be chosen after prior inspection of the data, but the functional form itself is not allowed to be determined by the data directly. Finding an appropriate parametric specification seems especially difficult for typical health care data, which feature numerous peculiarities such as "excess zeros" or "overdispersion". As a response, a myriad of different functional forms for $m(\cdot,\cdot,\cdot)$ has been used in practice. These range from simple linear regression models over more sophisticated nonlinear specifications such as the negative binomial model up to even more complex approaches like two-part models, which combine different parametric families such as logit or probit models with say truncated negative binomial models to obtain an expression for $m(\cdot,\cdot,\cdot)$ that depends on a finite-dimensional parameter vector β only.[7] Such parametric approaches have some well-known merits and shortcomings. Particularly, if the functional form $m(\cdot,\cdot,\cdot)$ is correctly specified, parametric maximum likelihood estimators for β will be root-N-consistent, asymptotically normal and efficient. If the specification of $m(\cdot,\cdot,\cdot)$ is incorrect, however, the resulting estimate $m\left(S_i, H_i, \widehat{\beta}\right)$ will generally be inconsistent for $E\left[m_i|S_i, H_i\right]$, and may therefore be seriously misleading about how average levels of care use vary by SES and care need. It is easy to see that this issue is indeed a cause for concern for the applied researcher. The multitude of different models used in the past already highlights that an a priori picking of the correct functional form

[6] Throughout the paper, we always use tildes to indicate flexible parameters, which are determined in the estimation rather than a priori.

[7] See e.g. Jones (2000) for an overview.

for $m(\cdot,\cdot,\cdot)$ is far from trivial. In this context, it is also worth noting that choosing $m(\cdot,\cdot,\cdot)$ does not only require deciding on the right distributions for the stochastic error terms, but also taking a stance on how potentially important interactions between S_i and H_i can be accurately reflected in the model. The various different treatments adopted in the previous literature again illustrate the difficulties in resolving these issues using prior reasoning alone. For example, some studies simply impose additivity in S_i and H_i, and thereby rule out any SES-need interactions in the use of health care. Yet, other approaches include linear interactions in the model to allow for some linear interdependence in the way S_i and H_i affect m_i, whereas a third strand of the literature employs arbitrarily discretised values of S_i and then runs separate regressions for each SES group, corresponding to a model with full SES-need interactions at the SES-group level however these are defined.[8]

The strong requirement of extensive prior information as well as the arbitrariness of parametric assumptions have long been identified as key shortcoming of parametric approaches. As McFadden notes in his famous 1985 presidential address to the Econometric Society, parametric modelling "interposes an untidy veil between econometric analysis and the propositions of economic theory, which are most abstract without specific dimensional or functional restrictions". This argument seems to apply with even stronger force to the empirical analysis of equity in health care utilisation, since there is rarely any formal theory that could be used to justify specific parametric restrictions ex ante.

2.2 Nonparametric models

Fully nonparametric models resolve the arbitrariness of parametric approaches, which stems from the need to choose a functional form for $m(\cdot,\cdot,\cdot)$ ex ante. Formally, a model for the conditional expectation of care use m_i, given SES S_i, and care need H_i can be written as

$$E[m_i|S_i, H_i] = \widetilde{m}(S_i, H_i) \tag{2}$$

Rather than restricting $m(\cdot,\cdot,\cdot)$ to lie in some a priori specified parametric family, nonparametric methods treat $\widetilde{m}(\cdot,\cdot)$ as an infinite-dimensional unknown parameter to be estimated from the data. Apart from mild regularity conditions, these methods do not impose any restrictions on the relationship between S_i, H_i and $E[m_i|S_i, H_i]$. Thus, "the main strength of nonparametric over parametric regression is the fact that it

[8] See e.g. Jones (2000) and Wagstaff and van Doorslaer (2000) for further references.

assumes no functional form for the relationship, allowing the data to choose, not only the parameter estimates, but the shape of the curve itself" (Deaton (1997), p.193). A variety of approaches, such as kernel or series methods can be used to estimate $\widetilde{m}(\cdot,\cdot)$.[9] While such a general estimation approach provides a desirable safeguard against the potential adverse effects of parametric misspecification, the price of this added flexibility and robustness are much greater data requirements for actual implementation. Particularly, the precision of fully nonparametric estimators is often poor and their rate of convergence is usually slower than in parametric models. This is especially true in higher-dimensional problems like the one considered here, where both SES and health care need may only be measurable via multiple indicators. Hence, there is a practical trade-off between the use of parametric or nonparametric methods for estimation based on finite samples. Given this trade-off, it would seem desirable to combine the two approaches to retain some flexibility and robustness in modelling the function $E[m_i|S_i, H_i]$, but mitigate the "curse of dimensionality" associated with nonparametric regression. Semiparametric models do this by introducing parametric components to the model to attain some dimensionality reduction. Of course, the plausibility of such an approach depends on whether such parametric elements seem justifiable by theory, as indicated by the McFadden quote. We use the nature of the policy discourse to motivate a bivariate semiparametric framework for modelling $E[m_i|S_i, H_i]$, which ought to strike a reasonable balance between precision, flexibility and coherency with theory.

3 Bivariate semiparametric index models

Moving from fully nonparametric to semiparametric regression requires the introduction of some theoretically justifiable parametric elements in (2). At first, this appears challenging, as we have not provided a formal model on how SES and health care need bring about specific levels of care use. Moreover, looking at the previous literature as well as the policy debate, the only apparent theoretical differentiation is their categorization into legitimate and illegitimate sources of differences in medical care utilisation, what we have labelled SES S_i on the one hand, and health care need H_i on the other. At the conceptual level, the discourse therefore appears merely dichotomous, with additional complexities solely arising from the use of multiple measures to capture "need" and "non-need" determinants of health care use. In other words, while SES and health

[9] See for example Härdle (1990), Härdle and Linton (1994), Deaton (1997) or Yatchew (2003) for accessible introductions to the topic.

care need appear sufficiently well-defined for theory, they seem inherently difficult to actualize when bringing theory to the data. The class of bivariate semiparametric index models suggested here reflects this dichotomous structure of the debate - aggregating illegitimate and legitimate sources of differences in care use into one-dimensional concepts of SES and care need, respectively.[10] The analysis then proceeds fully nonparametrically when estimating how expected care use varies with SES and health care need. Specifically, it does not impose any additional functional form assumption on $m(\cdot,\cdot,\cdot)$ beyond the two index restrictions that allow aggregation of the multiple SES and care need indicators.

To understand how this procedure works, assume that the concept of SES can be measured as a linear combination of say education, income or wealth, while health care need may be measured as a linear combination of multiple need indicators such as age, diseases, functioning measures, respectively. We can then represent SES and health care need as two indices I_i^{SES} and I_i^{HCN} constructed as a linear combination of the multiple SES and need indicators S_i and H_i that are measured in the data. We thus obtain

$$I_i^{SES} = S_i\delta \tag{3}$$

$$I_i^{HCN} = H_i\beta \tag{4}$$

where δ and β denote specific aggregation parameters pertaining to the broader concepts of SES and health care need, which are now measured via I_i^{SES} and I_i^{HCN}, respectively. Assuming that expected health care utilisation $E[m_i|S_i,H_i]$ depends on S_i and H_i through the two one-dimensional indices I_i^{SES} and I_i^{HCN} only, we obtain a semiparametric model of the form

$$E[m_i|S_i,H_i] = E\left[m_i|\widetilde{I_i^{SES}},\widetilde{I_i^{HCN}}\right] = \widetilde{m}\left(\widetilde{I_i^{SES}},\widetilde{I_i^{HCN}}\right) = \widetilde{m}\left(S_i\widetilde{\delta},H_i\widetilde{\beta}\right) \tag{5}$$

where $\widetilde{\delta}$, $\widetilde{\beta}$ and $\widetilde{m}(\cdot,\cdot)$ denote unknown (finite- and infinite-dimensional) parameters that need to be estimated from the data. The only parametric restrictions involved in this model are the index assumptions (3) and (4). Particularly, the approach remains fully nonparametric with respect to $\widetilde{m}(\cdot,\cdot)$, and does therefore not incorporate any a priori constraint on how $\widetilde{I_i^{SES}}$ and $\widetilde{I_i^{HCN}}$ affect $E[m_i|S_i,H_i]$.

Being a hybrid of parametric and nonparametric regression, the semiparametric dou-

[10] Ichimura and Lee (1991) have introduced multiple index model into the theoretical econometrics literature, which have been subsequently applied to diverse research questions with bipartite structures such as supply and demand (Stern (1996), Maurer and Pohl 2007)) or interactions of macro and micro determinants of the income distribution (Farré-Olalla and Vella (2006)) to name just a few.

ble index model in (5) features obvious similarities with both approaches. Like in parametric models, (5) includes some finite-dimensional unknown parameter vectors $\widetilde{\delta}$ and $\widetilde{\beta}$. Beyond $\widetilde{\delta}$ and $\widetilde{\beta}$, however, the model also includes the infinite-dimensional parameter $\widetilde{m}(\cdot,\cdot)$ which represents a fully flexible link function for mapping SES $\widetilde{I_i^{SES}}$ and care need $\widetilde{I_i^{HCN}}$ into expected care use $E[m_i|S_i,H_i]$. Like in the nonparametric approach, the functional form of $\widetilde{m}(\cdot,\cdot)$ is not at all constrained a priori, but in fact is estimated from the data. On important advantage of this modelling strategy is that it flexibly incorporates observable heterogeneity in the effects of SES $\widetilde{I_i^{SES}}$ for different levels of health care need $\widetilde{I_i^{HCN}}$. Hence, SES-gradients are allowed to vary freely across the need distribution, which may reveal some useful information for policy design.

The class of semiparametric index models does, however, not only inherit the advantages of both approaches, but also their disadvantages. On the one hand, semiparametric estimators for $E[m_i|S_i,H_i]$ are generally consistent under a wider range of circumstances than their parametric counterparts and therefore more robust to potential misspecification. At the same time, they are usually more precise than their nonparametric counterparts, due to the built-in dimensionality reduction implied by the index restrictions. Specifically, the parametric components of the model, $\widetilde{\delta}$ and $\widetilde{\beta}$, can be estimated with the usual parametric rate of convergence, whereas the estimate for the conditional expectation $\widetilde{m}\left(S_i\widetilde{\delta},H_i\widetilde{\beta}\right)$ converges at the (slower) rate of a nonparametric estimate of a conditional mean function with two arguments, and thus much faster than if S_i and H_i would be treated fully nonparametrically. On the other hand, semiparametric estimators may be considerably less efficient than their fully parametric counterpart, at least if the latter can be correctly specified based on prior information. Also, unlike the fully nonparametric approach, semiparametric estimation may still lead to inconsistent estimates, if the parametric part of the model - here the two index restrictions - are inaccurate. We therefore consider semiparametric models as complementary to the other two approaches.

4 Empirical illustration

This section presents an empirical example to bring the above concepts to life, and show the applicability of such semiparametric index models to typical survey data on health care use.

4.1 Data

To illustrate the kinds of insights that a semiparametric double index approach may deliver, we estimate a model for health care utilisation of older Italian men using data from the first wave of the Survey of Health, Ageing and Retirement in Europe (SHARE) collected in 2004.[11] SHARE is a multidisciplinary, cross-national micro data base containing information on health and socioeconomic status of some 22,000 Continental Europeans aged 50+ from ten European countries. Yet, for the sake of this simple illustration, we restrict our sample to the 1017 male respondents from Italy.[12]

4.2 Model specification

We measure health care utilisation as the total number of doctor visits during the last twelve months. This measure includes both visits to GPs as well as specialists, but does not account for inpatient care as a potential substitute. For SES, we use years of education as our proxy variable. As we are considering an elderly population, alternative SES measures such as income or wealth appear heavily confounded by the respondent's labour market status as well as typical life-cycle trajectories of asset holdings, respectively.[13] It is, however, important to note, that the suggested model could easily incorporate multiple SES indicators if desired. Specifically, these indicators would then form an actual SES-index, replacing our one-dimensional SES-measure based on education. The corresponding index coefficients would then be estimated in the same way as the index coefficients of the care need index, to which we turn now. Our model uses a large number of health indicators to comprehensively capture individual differences in health care need. In fact, the abundance of health measures available in SHARE makes it an ideal data source for the kind of exercise considered here.

Apart from age and dynamometer-measured maximum grip strength, we include

[11] This paper uses data from the early release 1 of SHARE 2004. This release is preliminary and may contain errors that will be corrected in later releases. The SHARE data collection has been primarily funded by the European Commission through the 5th framework programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life). Additional funding came from the US National Institute on Ageing (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, Y1-AG-4553-01 and OGHA 04-064). Data collection in Austria (through the Austrian Science Foundation, FWF), Belgium (through the Belgian Science Policy Office) and Switzerland (through BBW/OFES/UFES) was nationally funded. The SHARE data set is introduced in Börsch-Supan et al. (2005); methodological details are contained in Börsch-Supan and Jürges (2005).

[12] Results from a more comprehensive semiparametric cross-country comparison of care utilization using data for both sexes and all ten initial SHARE countries can be found in Maurer (2007).

[13] See Maurer (2007) for a more detailed justification of this choice. Banks et al. (2002) provide for a more comprehensive theoretical discussion of the issues involved. An empirical assessment can, for example, be found in Vos (2004).

a set of 15 binary indicators for different doctor-diagnosed health conditions. These ought to capture various aspects of individual health care need, including acute conditions such as heart attacks or stroke as well as chronic diseases like diabetes or mere risk factor such as hypertension, all of which require a different form and intensity of disease management. Table 1 presents basic descriptive statistics for all variables used in the analysis. The unconditional mean of the number of doctor visits in the last twelve months is 7.23. The respondents have an average education level of 7.6 years of schooling. An additional noteworthy feature of the data is our set health controls features considerable heterogeneity in terms of prevalence rates. While some conditions such as hypertension or arthritis are quite prevalent among older Italian men, others such as hip fractures or Parkinson are fairly rare.

4.3 Estimation

As our empirical illustration employs many health care need indicators but only one SES measure[14], the general double index model for estimating $E\left[m_i|S_i,H_i\right]$ based on (5) simplifies to

$$E\left[m_i|S_i,H_i\right]=\widetilde{m}\left(S_i,H_i\widetilde{\beta}\right) \tag{6}$$

and thus requires estimation of only one index parameter $\widetilde{\beta}$ in addition to nonparametric link function $\widetilde{m}\left(\cdot,\cdot\right)$.[15]

We use Ichimura and Lee's (1991) multiple index extension of Ichimura's (1993) semiparametric least squares estimator for single index models to estimate the index coefficient $\widetilde{\beta}$ in (6) along with the nonparametric link function $\widetilde{m}\left(\cdot,\cdot\right)$. Specifically, Ichimura and Lee's estimator for $\widetilde{\beta}$ is based on minimizing a semiparametric least squares criterion function of the form

$$SSR\left(\widetilde{\beta}\right)=\sum_{i=1}^{N}\left(m_i-\widetilde{m}\left(S_i,H_i\widetilde{\beta}\right)\right)^2 \tag{7}$$

[14] From the derivations that follow, it is straightforward to see how the estimation can also handle richer index specifications for S_i. Yet, for reasons outlined in the specification section, as well as in Maurer (2007), we prefer the specification based on education alone, which seems still sufficient to illustrate the main features of bivariate semiparametric models.

[15] Note that, as always in semiparametric index models, separate identification of the components of $\widetilde{m}\left(S_i,H_i\widetilde{\beta}\right)$ can only be attained up to location and scale (see e.g. Horowitz (1998)). This is why the first argument of $\widetilde{m}\left(\cdot,\cdot\right)$ does not contain any parameter and there is also no intercept in $\widetilde{\beta}$ (to normalize location) while its first element is set equal to one (to normalize scale).

where an estimate of $\widetilde{m}\left(S_i, H_i\widetilde{\beta}\right)$ for any given candidate parameter vector $\widetilde{\beta}$ is constructed via bivariate nonparametric kernel regression of m_i on S_i and $H_i\widetilde{\beta}$, respectively.[16] The estimate of $\widetilde{m}\left(S_i, H_i\widetilde{\beta}\right)$ is thus computed via

$$\widetilde{m}\left(S_i, H_i\widetilde{\beta}\right) = \frac{\sum_{j=1}^{n} K_{h_i}\left(X_i - X_j\right) m_j}{\sum_{j=1}^{n} K_{h_i}\left(X_i - X_j\right)} \tag{8}$$

with $X_i = \left(S_i, H_i\widetilde{\beta}\right)$, i.e. a two-dimensional vector consisting of SES S_i and health care need $H_i\widehat{\beta}$, respectively, and $K_{h_i}\left(X_i - X_j\right) = \det(h_i)^{-1} \cdot K\left(h_i^{-1}\left(X_i - X_j\right)\right)$ for some bivariate kernel function $K(\cdot)$ and matrices of local bandwidths h_i regulating the degree of smoothing in the two directions of the $\left(S_i, H_i\widetilde{\beta}\right)$-space.

4.4 Selected results

As highlighted in the above discussion, the semiparametric modelling approach features two main estimation parameters - the finite-dimensional parameter $\widetilde{\beta}$ as well as the infinite-dimensional parameter $\widetilde{m}(\cdot,\cdot)$. Our presentation of the results reflects this bipartite structure, starting with a brief discussion of what we estimate as health care need $H_i\widehat{\beta}$ before turning to the semiparametric estimate for the conditional expectations $E\left[m_i|S_i, H_i\right]$ that also incorporates an estimate $\widehat{m}(\cdot,\cdot)$ of the nonparametric link function $\widetilde{m}(\cdot,\cdot)$.

Table 2 presents the estimates for the parameter vector $\widehat{\beta}$, which aggregates all health indicators into the one-dimensional health care need index $H_i\widehat{\beta}$. Of course, we cannot investigate the effects of these health controls on actual care utilisation without knowledge of the unspecified mapping $\widehat{m}(\cdot,\cdot)$. However, we can check whether their aggregation into a single care need index is consistent with our prior expectations regarding their relative signs. For identification, the index does not include an intercept and we have normalized the coefficient of age to 0.01 to fix its location and scale. Given this normalization, we would expect that all of our health controls, with the exception of maximum grip strength, enter the model with a positive sign.[17] In fact, a remarkably consistent

[16] The actual estimator requires some additional adjustments such as trimming of the criterion for observations were the data is sparse as well as the use of so-called higher order kernels or local bandwidth selection (local smoothing). While the actual estimations incorporate trimming and local smoothing, these technicalities are omitted from the discussion here for the sake of brevity and expositional ease.

[17] This presumes of course that both higher age and the prevalence of a health condition indicate more care need, whereas higher grip strength indicates less need for medical attention. So far, we do not know how the care need index maps into actual utilization, but we can assess this relationship once we turn to our estimate of $\widetilde{m}(\cdot,\cdot)$.

pattern emerges. Almost all of our health conditions enter the index positively, whereas maximum grip strength has indeed the expected negative coefficient. Specifically, the only health conditions that enter the model with a very small and statistically insignificant negative coefficient are "having been diagnosed with high cholesterol" and "having ever been diagnosed with cataracts", two conditions for which we would also not have expected a large effect on care need.[18] Also, the relative sizes of the estimated coefficients seem largely in line with prior expectations. Given this encouraging first glimpse on the constituents of care need, we can now turn to a more comprehensive assessment of the effects that our health controls on actual care use. Also, and arguably more interestingly, we can assess the effect of our SES measure - education - on medical care utilisation, and how its effects vary across the distribution of health care need. To do so, we present our estimate of the nonparametric function $\widehat{m}\left(\cdot,\cdot\right)$ which links years of education S_i and the aggregated care need index $H_i\widehat{\beta}$ with the expected number of doctor visits.

Since semi- and nonparametric methods leave it to the data to choose the shape of the regression function $\widetilde{m}\left(\cdot,\cdot\right)$, we can of course not come up with a corresponding estimate $\widehat{m}\left(\cdot,\cdot\right)$ out-of sample. It is therefore important to first clarify the relevant support of S_i and $H_i\widehat{\beta}$ over which we can estimate the conditional expectation function $\widetilde{m}\left(\cdot,\cdot\right)$ nonparametrically. Figure 1 presents bivariate density estimates for the joint distribution of S_i and $H_i\widehat{\beta}$ to highlight the relevant support of the data. The joint distribution of S_i and $H_i\widehat{\beta}$ in the sample is concentrated at five to six years of education and a care need index value of around 0.4, but covers a fairly wide range of educational attainment and health care need. In the discussion of our estimates, we will make sure to just consider points that lie well inside the support of the data to avoid invalid out-of-sample prediction as well as spurious results due to a lack of sufficient data in the tails.

Figure 2 presents our semiparametric estimate of the conditional expectations function $E\left[m_i|S_i,H_i\right]$, which is the main object of interest in this study. Exploiting the semiparametric structure of (6), the two plots display $\widehat{m}\left(\cdot,\cdot\right)$ as a function of its two arguments, education S_i and health care need $H_i\widehat{\beta}$. The figure reveals several interesting patterns regarding the interplay of S_i and $H_i\widehat{\beta}$ in the determination of care use. Particularly intriguing is the vast observable heterogeneity in the effects of educational attainment across different points in the care need distribution. For low levels of care need, say $H_i\widehat{\beta}=0.1$, the average number of doctor visits is only slightly increasing with

[18] This is especially true for the latter, which can basically be completely cured by one-time surgery.

increasing levels of educational attainment. Among these relatively healthy respondents, we estimate only a small positive education gradient, with conditional means of $\widehat{m}\,(2, 0.1) = 3.12$ for two years of education and $\widehat{m}\,(14, 0.1) = 3.88$ for fourteen years of education, respectively. Yet, the effects of education change dramatically as we move along the distribution of care need. Specifically, the slightly positive education gradient gradually reverses, being almost zero at $H_i\widehat{\beta} = 0.44$ before reemerging as a pronounced negative gradient for those with higher levels of care need. For example, for respondents with a care need level of $H_i\widehat{\beta} = 1$, the average number of doctor visits in the past year varies from 13.26 to 9.46 for two and fourteen years of education, respectively. We thus find a fairly strong dependence of the effects of SES on care use for different levels of care need, indicating important SES-need interactions that seem to call for explicit consideration when modelling SES-gradients in care use.

5 Conclusion

The present paper suggests semiparametric double index models as a potentially valuable tool for applied researchers analysing the interplay of SES and health care need in determining medical care utilisation. The previous literature has mostly employed parametric methods, in which the functional form of the regression function is a priori restricted by the researcher. While these approaches surely have certain advantages in terms of estimation efficiency and inference if correctly specified, they are also known to perform poorly if the assumed parametric structure is inaccurate. Semiparametric approaches on the other hand, combine parametric modelling with nonparametric estimation. This approach seems especially advantageous if there is little a priori knowledge about potentially complex features in the data, but some parametric structure can nonetheless be justified by theory. The semiparametric double index approach suggested here is motivated by these observations. With regard to the micro-determinants of health care use, almost all studies feature a bipartite conceptual distinction between SES and care need as illegitimate and legitimate sources of differences in medical care utilisation, even if either concept can only be actualized via multiple proxies. The parametric component of our suggested modelling approach therefore assumes that multiple SES and care need indicators can be aggregated into a single SES and health care need index respectively. Given these one-dimensional SES and care need measures, the analysis proceeds fully nonparametrically. Particularly, we do not assume any specific functional form with respect to how SES and care need interact in bringing about certain intensities of care use.

Our approach thus allows for observable heterogeneity in SES-gradients across the care need distribution, which may deliver important insights for targeting health policies.

We demonstrate how the suggested method works based on actual data on health care use of older Italian men taken from the first wave of SHARE. Particularly, our example considers a typical research question in equity analysis, namely the conditional effects of education on health care use, given multiple controls for care need. Our illustration turns out to be an interesting case of partially offsetting education gradients which vary considerably across the distribution of health care need. Particularly, our estimates indicate that the highly educated healthy respondents consume more care than their less educated counterparts. The education gradient then gradually reverses with the less educated using considerably more care when sick than the well educated respondents. Similar to Abasolo et al. (2001), we can interpret our regression results in terms of inequity in health care delivery. Yet, our approach has the additional advantage that we can explicitly consider how SES-gradients may vary across the need distribution. Our example has highlighted that such observable heterogeneity may indeed be important and by itself informative for targeting policies. In addition, it may also pose some challenges for conventional parametric approaches as to how to account for such heterogeneity.

Admittedly, the paper leaves some open issues for future research. Firstly, it would be most useful to also attempt an empirical comparison of the various parametric models with the semiparametric approach suggested here, which may deliver further insights on the practical relevance of some of the theoretical concerns that have served as its motivation. Secondly, investigating the usefulness of the proposed estimator for formal approaches of measuring and testing for inequity in health care delivery seems another promising route for further work. As it stands, we nonetheless deem semiparametric estimation as a promising tool for flexibly modelling SES-need interactions in the delivery of care, even if additional evidence on its relative performance is surely desirable.

References

Abasolo, I, Manning, R, Jones, A 2001, Equity in utilization of and access to public-sector GPs in Spain, *Applied Economics*, **33**: 349-364.

Adler, N, McEwen, B, Stewart, J, Marmot, M (eds) 2000, *Socioeconomic status and health in industrial nations: social, psychological, and biological pathways,* Annals of the New York Academy of Sciences: New York.

Adler, N, Newman, K 2002, Socioeconomic disparities in health: pathways and policies, *Health Affairs*, **21**: 60-76.

Banks, J, Blundell, R, Marmot, M, Nazroo, J 2002, Economic measures in health surveys, IFS London.

Börsch-Supan, A, Brugiavini, A, Jürges, H, Mackenbach, J, Siegrist, J, Weber, G (eds) 2005, *Health, ageing and retirement in Europe – first results from the survey of health, ageing and retirement in Europe*; MEA Mannheim.

Börsch-Supan, A, Jürges, H (eds) 2005, *Health, ageing and retirement in Europe – methodology*; MEA Mannheim.

Deaton, A 1997, *The analysis of household surveys: a microeconometric approach to development policy.* Johns Hopkins University Press: Baltimore.

Deaton, A 2003, Health, inequality, and economic development. *Journal of Economic Literature*, **41**: 113-158.

Ferré-Olalla, L, Vella, F 2006, Macroeconomic conditions and the distribution of income in Spain, *IZA Discussion Paper* No. 2512.

Härdle, W 1990, *Applied nonparametric regression.* Cambridge University Press: Cambridge.

Härdle, W, Linton, O 1994, Applied nonparametric methods. In *Handbook of Econometrics*, Volume 4: 2295-2339, Engle,R, McFadden D (eds) Elsevier Science: Amsterdam. Science: Amsterdam.

Horowitz, J 1998, *Semiparametric methods in econometrics.* Springer: New York.

Ichimura, H 1993, Semiparametric least squares (SLS) and weighted SLS estimation of single index models. *Journal of Econometrics*, **58**: 71-120.

Ichimura, H, Lee, L 1991, Semiparametric least squares estimation of multiple index models: single equation estimation", In *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, 3-49, Barnett, W, Powell, J Tauchen, G (eds) Cambridge: Cambridge University Press.

Jones, A 2000, Health econometrics. In *Handbook of Health Economics*, Volume 1: 265-344, Culyer A, Newhouse J. (eds) Elsevier Science: Amsterdam.

Manski, C 2005: *Social choice with partial knowledge of treatment response.* Princeton University Press: Princeton.

Marmot, M 2005, *Status syndrome.* Bloomsbury: London.

Marmot, M, Wilkinson, R 1999, *Social determinants of health.* Oxford University Press: Oxford.

Maurer, J 2007, Socioeconomic and health determinants of health care utilization among older Europeans: a semiparametric assessment of equity, intensity and responsiveness for ten European countries , MEA Mannheim.

Maurer, J, Pohl, V 2007, Who has a clue to preventing the flu? MEA Mannheim.

Powell, J 1994, Estimation of semiparametric models. In *Handbook of Econometrics*, Volume 4: 2443-2521, Engle,R, McFadden D (eds) Elsevier Science: Amsterdam.

Smith, J 1999, Healthy bodies and think wallets: the dual relationship between health and socioeconomic status. *Journal of Economic Perspectives*, **13**: 145-166.

Smith, J 2003, Consequences and predictors of new health events. *IFS Working Paper* WP03/22.

Smith, J 2004, Unraveling the ses-health connection. *Population and Development Review*, **30**: 108-132.

Stern, S 1996 Semiparametric estimates of the supply and demand effects of disability on labor force participation. *Journal of Econometrics*, **71**: 49-70.

Vos, S 2005, Indicating socioeconomic status among elderly people in developing societies: an example from Brazil, *Social Indicators Research*, **73**: 87-108.

Wagstaff, A, van Doorslaer, E 2000, Equity in health care finance and delivery. In *Handbook of Health Economics*, Volume 1: 1803-1862, Culyer A, Newhouse J. (eds) Elsevier Science: Amsterdam.

Yatchew, A 2003, *Semiparametric regression for the applied econometrician*, Cambridge University Press: Cambridge.

Table 1: Summary statistics

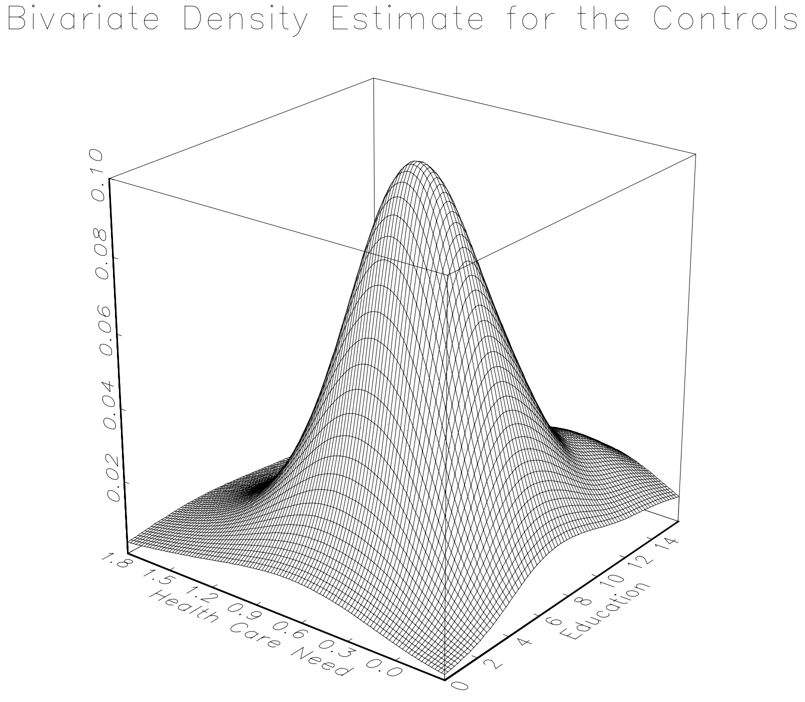
Variable	Mean	Standard Error	Minimum	Maximum
Number of doctor visits	7.230	11.280	0	98
Years of education	7.600	4.340	0	22
Age	64.752	8.580	50	94
Maximum grip strength	39.942	10.680	7	70
Asthma	0.044	0.206	0	1
Cancer	0.025	0.155	0	1
Cataracts	0.048	0.214	0	1
Cholesterol	0.176	0.381	0	1
Diabetes	0.122	0.327	0	1
Heart attack	0.112	0.316	0	1
Hip fracture	0.012	0.108	0	1
Hypertension	0.354	0.478	0	1
Lung disease	0.080	0.271	0	1
Osteoporosis	0.015	0.121	0	1
Parkinson	0.005	0.070	0	1
Stroke	0.030	0.172	0	1
Ulcer	0.078	0.268	0	1
Other condition	0.127	0.333	0	1
Number of observations	1017			

Table 2: Parameter estimates (health care need index)

Variable	Coefficient	Standard Error
Age	0.0100	--------
Maximum grip strength	-0.0120	0.0064
Asthma	0.1431	0.1244
Cancer	0.2997	0.2301
Cataracts	-0.0023	0.0994
Cholesterol	-0.0044	0.0730
Diabetes	0.0834	0.0547
Heart attack	0.7112	0.2781
Hip fracture	0.3410	0.3959
Hypertension	0.4009	0.1762
Lung disease	0.3128	0.1413
Osteoporosis	0.1521	0.2000
Parkinson	0.2280	0.7824
Stroke	0.1387	0.1210
Ulcer	0.0966	0.1065
Other condition	0.1361	0.0953

Figure 1: Bivariate Density Estimates for the Controls

A. Surface Plot



B. Contour Plot

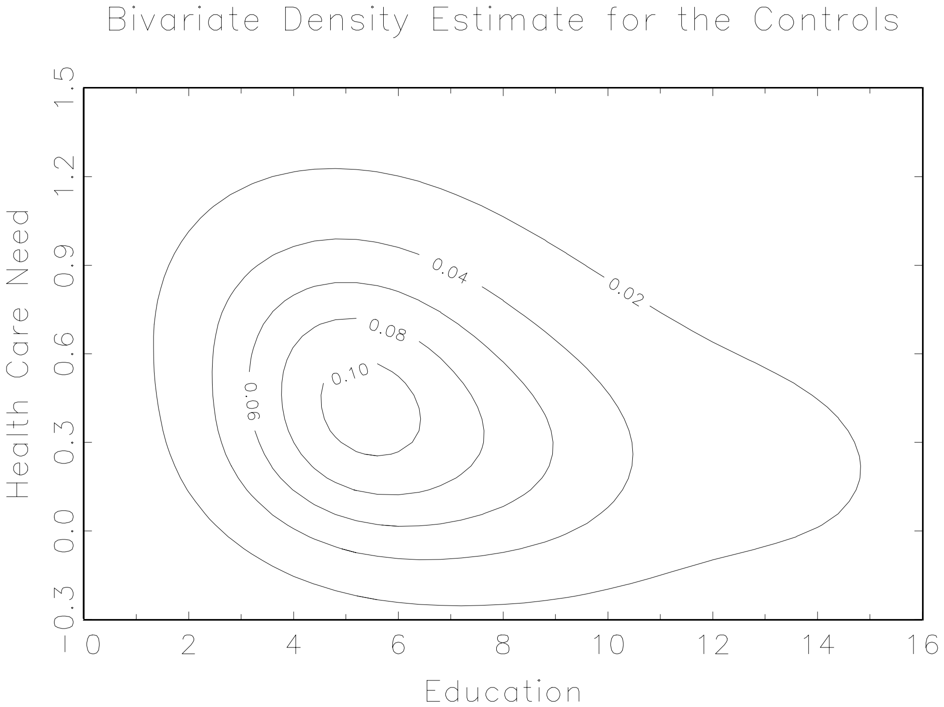
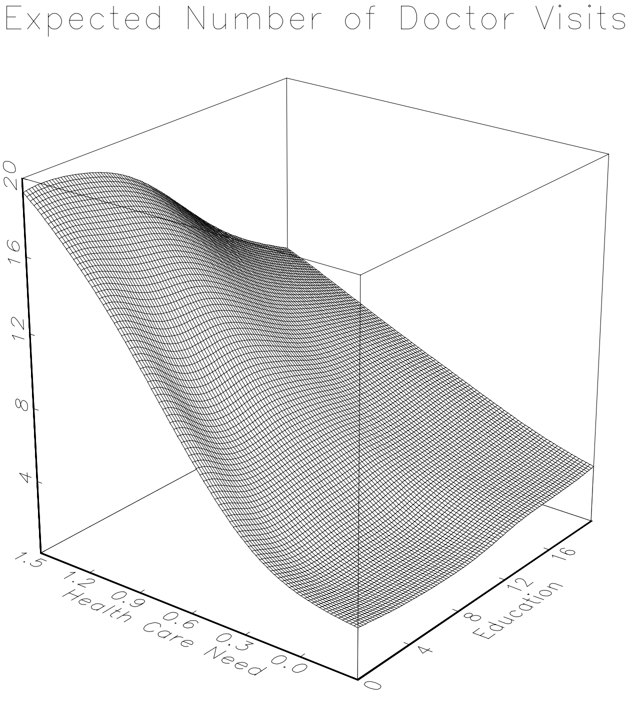
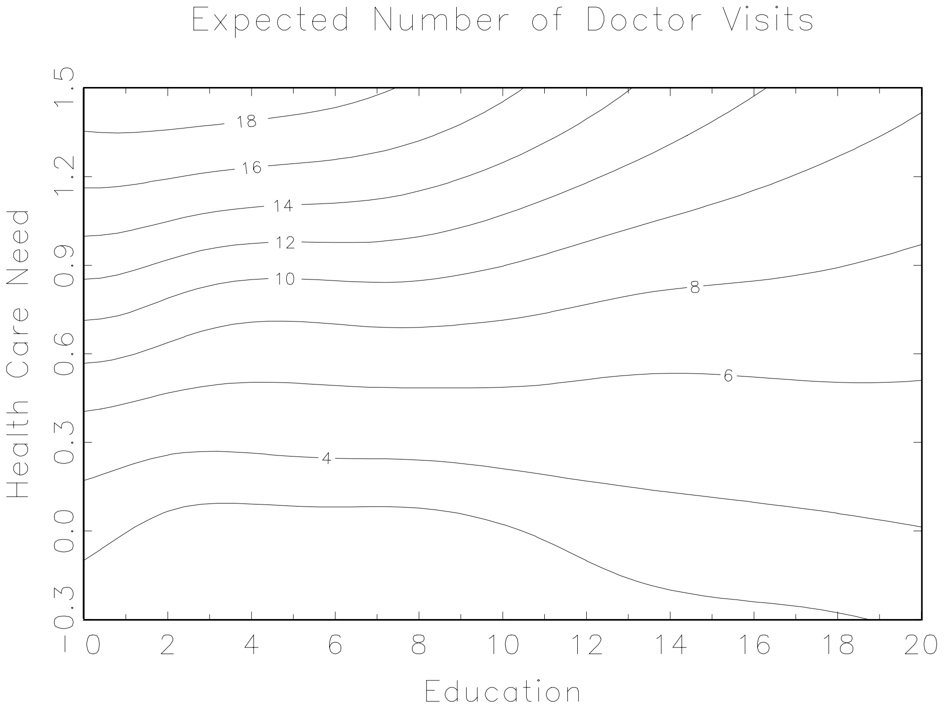


Figure 2: Estimated Conditional Expectations Function

A. Surface Plot



B. Contour Plot



Discussion Paper Series

Mannheim Research Institute for the Economics of Aging Universität Mannheim

To order copies, please direct your request to the author of the title in question.

Nr.	Autoren	Titel	Jahr
133-07	Axel Börsch-Supan	Über selbststabilisierende Rentensysteme	07
134-07	Axel Börsch-Supan, Hendrik Jürges	Early Retirement, Social Security and Well-Being in Germany	07
135-07	Axel Börsch-Supan	Work Disability, Health, and Incentive Effects	07
136-07	Axel Börsch-Supan, Anette Reil-Held, Daniel Schunk	The savings behaviour of German households: First Experiences with state promoted private pensions	07
137-07	Hendrik Jürges, Mauricio Avendano, Johan Mackenbach	How comparable are different measures of self-rated health? Evidence from five European countries	07
138-07	Hendrik Jürges, Kerstion Schneider	What can go wrong will go wrong: Birthday effects and early tracking in the German school system	07
139-07	Hendrik Jürges	Does ill health affect savings intentions? Evidence from SHARE	07
140-07	Hendrik Jürges	Health inequalities by education, income, and wealth: a comparison of 11 European countries and the US	07
141-07	Hendrik Jürges	Healthy minds in healthy bodies. An international comparison of education-related inequality in physical health among older adults	07
142-07	Karsten Hank, Stephanie Stuck	Volunteer Work, Informal Help, and Care among the 50+ in Europe: Further Evidence for ‘Linked’ Productive Activities at Older Ages	07
143-07	Jürgen Maurer	Assessing Horizontal Equity in Medication Treatment Among Elderly Mexicans: Which Socioeconomic Determinants Matter Most?	07
144-07	Jürgen Maurer	Socioeconomic and Health Determinants of Health Care Utilization Among Elderly Europeans: A Semiparametric Assessment of Equity, Intensity and Responsiveness for Ten European Countries	07
145-07	Jürgen Maurer	Modelling socioeconomic and health determinants of health care use: A semiparametric approach	07