Semiparametric Estimation of Social Tariffs: An application using the SF-6D value sets.

Ildefonso Mendez[1]
University of Murcia

Jose M. Abellan
University of Murcia

Jorge E. Martinez
University of Murcia

Fernando I. Sanchez
University of Murcia

[1]Address for correspondence: Departamento de Economía Aplicada, Facultad de Economía y Empresa, 30100-Espinardo (Murcia), Spain. Phone: +34 968363732. Fax: +34 968363745. Email: ildemm@um.es.

Abstract

This paper presents a semiparametric approach to model health state values with important advantadges over the traditional parametric one. Our method makes no assumption on the distribution of health state values, accomodates covariates in a flexible way, eschews parametric assumptions on the relationship between the outcome and the regressors and it allows for an undetermined amount of heterogeneity in the estimates. Additionally, it produces estimates for the population of interest even if the sample is not representative for that population with regard to many discrete and continuous individual characteristics that affect health state values. The estimates obtained using the semiparametric method are higher in absolute value than the regression estimates, particularly so when adjusting for the distribution of individual characteristics in the Spanish population and when analyzing the valuation effect of small departures from full health. These results suggest that the standard method underestimates the value that the Spanish population assigns to a given departure from full health and, in particular, to small departures.

JEL Classification: I1.

Keywords: Preference-based Health Measure; SF-6D; Semiparametric Methods.

1 Introduction

Preference-based measures of health status are increasingly being used to evaluate the outcomes of health care interventions and to inform resources allocation decisions. A number of health state descriptive systems have been designed for calculating a single index value for every state defined within the system. Since most descriptive systems define more health states than it is feasible to elicit direct valuations for in an empirical study, choices have to be made about how best to estimate values for all states from direct observations on a subset of those states.

The standard estimation model uses a set of dummy indicator variables describing health states in terms of their level of severity in different dimensions of health to explain the individual valuations obtained. Under the assumption of normally distributed errors, a regression of health state values on the set of dummy variables identifies the valuation effect of departures from full health. The estimates are then used to predict the value associated to the health states not directly valued.

However, the standard model presents some limitations that are likely to result in biased estimates at the sample and population levels. First, health state valuation data are skewed, truncated, non-continuous and hierachical (Brazier et al., 2002) and, thus, the normality assumption is not likely to hold. Second, the standard model does not provide any guidance on how to control for individual characteristics that affect health state values. The traditional way of accomodating covariates to the standard model is to introduce them linearly and estimating an expanded regression equation. This approach, which implicitely assumes that covariates only affect the estimated intercept, is rejected by the evidence in Dolan and Roberts (2002) and Kharroubi et al. (2007). They find that respondent's characteristics impact on the value they give to health states and that this effect varies with the severity of the health state at examination. Furthermore, the alternative of introducing covariates additively contravenes both the goal of estimating one preference-based tariff for the whole community (Dolan, 1997) and the theoretical requirement of the intercept being equal to unity.

The debate on the valuation effect of the covariates is related to that on whose values should count when evaluating health state intervention outcomes and informing resources allocations decisions. The common recommendation of using the preferences of the whole population (NICE, 2003; Gold et al., 1996) calls for obtaining population valid estimates of the coefficients of interest, that is, to adjust for the distribution of

covariates in the population.

The standard practice of defining samples that are representative for the population with regard to the sex and age interval distribution of individuals does not guarantee that there are no sample selection biases in the estimates of interest. Non-response issues and the elimination of respondents providing inconsistent answers results in sometimes relevant differences between the initial sample design and the size of the estimation sample. This in turn might result in systematic differences between the theoretical and empirical distribution of sex and age intervals in the sample. Moreover, there might be other individual characteristics that affect health state values whose sample distribution is not neccesarily that in the population even if age intervals and sex are equally distributed in both instances. For example, Dolan and Roberts (2002) find that marital status and the usual activities dimension of own health affect health state values. Additionally, Kharroubi et al. (2007) find that the individual's employment status, educational level and own physical and social functioning condition health state valutions.

Additionally, the standard model does not provide the user with the appropiate statistical tools for testing and correcting for the population validity of the estimates. The traditional way of analyzing the representativeness of the estimates is to look for differences in the univariate descriptive statistics of the covariates between the estimation sample and the population. This approach, which implicitely recognizes the relevance of individual characteristics aside the respondent's sex and age interval, is not a formal test and, thus, it raises doubts as to in how many covariates do we have to find a significant difference of a given magnitude between the sample and population means for the estimates not to be valid at the population level. Moreover, finding no significant difference in the sample and population mean of a continuous variable is not necessarily very informative about the presence of differences in other moments of the distribution. Furthermore, multivariate distributions can differ and condition the population validity of the estimates even if univariate statistics are identical in the estimation sample and the population of interest.

The traditional way of recovering the population validity of the estimates once relevant differences are found between the sample and population distribution of the covariates is to include corrective weights. The weights are intended to adjust for the population distribution of individual characteristics where relevant discrepancies are detected. However, this procedure suffers from the course of dimensionality problem, that is, it becomes less feasible as the number of variables used to construct the

weights increases.

This paper presents a new approach for estimating preference-based measures of health that makes no assumption on the distribution of health state values, allows for an undetermined amount of heterogeneity in the estimates of interest, accomodates covariates in a flexible way, satisfies the theoretical requirement of the intercept being equal to unity, formally tests for differences in the distribution of individual characteristics between the estimation sample and the population of interest, estimates tariffs using the distribution of individual characteristics in the population of interest and, as opposed to the nonparametric Bayesian model in Kharroubi et al. (2005), it provides the user with a simple table of estimated coefficients that defines the estimated preference function, resulting in efficiency and transparency gains. Despite all these theoretical advantadges, the technical complexity of the semiparametric method that we present is only slightly higher than that of the standard regression estimator.

The paper has four more sections. Section 2 presents the semiparametric model and compares its properties to those of the standard parametric model. Section 3 describes the data used in the estimation. Section 4 presents and discuss the estimation results and Section 5 concludes.

2 Modelling

The standard model of health state valuations is

$$Y_{ij} = \alpha + \beta' Z_j + \varepsilon_{ij} \tag{1}$$

where Y_{ij} is the utility that individual i assigns to health state j, Z_j is a vector of dummy indicator variables Z_{kw} that equal one if health state j reaches level of severity w in dimension k and zero otherwise, for $w = 2, 3, .., W_k$ and $k = 1, 2, ., K$, α is the intercept and ε_{ij} is a zero mean error term.[1] Model (1) is the 'main effects' model, as opposed to other specifications that also control for level effects and/or interactions between the elements of Z_j. The model is estimated using the Ordinary Least Squares (OLS) or the Random Effects (RE) estimators, that is, assuming that ε is normally distributed.[2]

[1] For the SF-6D descriptive system $K = 6$ and W_k ranges from 4 to 6. Equivalently, $K = 5$ and $W_k = W = 3$ for the EQ-5D.

[2] Most papers use the RE estimator since it takes account that the same individual values several health states, increasing the efficiency of the estimates relative to the OLS estimator.

The main advantadge of the standard model is that it is easy to implement. However, it presents some limitations that are likely to overcome its benefits. First, the normality assumption is not likely to hold given the skewed, truncated, non-continuous and hierachical nature of the data (Brazier et al., 2002), thus, resulting in biased estimates.[3] Second, parametric models like (1) are severely limited in the way they control for individual characteristics. The standard approach of introducing individual characteristics additively and estimating an expanded version of equation (1) contravenes both the goal of estimating one preference-based tariff for the whole community (Dolan, 1997) and the requirement of the intercept being equal to unity.[4] The result of the estimated intercept deviating from unity follows from the unrealistic assumption that the valuation impact of a deviation from full health in a given dimension is the same for respondents with different background characteristics and is also independent of the level of severity in other dimensions.[5]

Additionally, the standard model estimates are also likely to be sensitive to differences in the covariate distributions for respondents valuing different levels of severity in different dimensions. That is the case because the standard model is a regression model and regression estimators rely heavily on extrapolation when differences in the covariate distributions for compared individuals are large.

The estimators that we propose make no assumption on the distribution of health state values, accomodate covariates in a flexible way, eschew parametric assumptions on the relationship between the outcome and the regressors, restrict the intercept to unity and allow for the effect of Z_{kw} to be heterogeneous in individual characteristics and in the severity of the departure from full health in the other dimensions. Additionally, and as opposed to the standard model, the proposed estimators produce estimates for the population of interest even if the estimation sample is not

[3] Dolan et al. (1996) find evidence that the distribution of health state valuations obtained using the time trade-off method was non-normal for each health state. Johnson et al. (1998) find departures from normality when estimating US-based population weights using the EQ-5D questionnaire. Diagnostic tests in Brazier et al. (2002) reveal non-normal residuals in the estimation of a preference-based measure of health for the UK general population using the SF-36. Many other studies, like Tsuchiya et al. (2002) and Lamers et al. (2006), simply provide no formal test of the underlying distributional assumption.

[4] There are strong theoretical reasons for restricting the intercept to unity since it captures the utility associated to full health, which equals one on the conventional full health-death scale used to estimate QALYs. See Brazier et al. (2002) for a discussion on this topic.

[5] The parametric way of relaxing this assumption is to further expand equation (1) by introducing interactions between the elements of Z_j and the variables that measure individual characteristics. However, this approach further contravenes the estimation of an unique population tariff and complicates both the interpretation of the estimates and the correct specification of the regression model.

representative for that population with regard to many discrete and continuous individual characteristics that affect health state values. The traditional approach of using corrective weights for the standard model estimates to be representative for the population of interest suffers from the so-called course of dimensionality problem. That is, corrective weights can only be defined over a reduced set of discrete covariates.[6] Moreover, the corrective weights approach does not provide a formal test for whether differences in the distributions of covariates between the sample and the population of interest are significant or not. The estimators that we propose overcome these limitations.

The identification strategy is presented focusing on β_{kw}, the coefficient associated to Z_{kw} that measures the average health state valuation impact of moving from level of severity 1 to level of severity w in dimension k. Let X_i be a vector of characteristics of individual i that potentially affect his valuations. The estimation sample is restricted to respondents valuing levels of severity 1 and w in dimension k and the individual and health state subscripts i and j are dropped out to simplify the notation. The effect of interest for the sample with individual characteristics x that value health states with level of severity w' in dimension k' for $k' \neq k$ and $w' = 2, 3, .., W_{k'}$ is[7]

$$\beta_{kw}(x, z_{k'w'}) = E[Y/Z_{kw} = 1, X = x, Z_{k'w'} = z_{k'w'}] - E[Y/Z_{kw} = 0, X = x, Z_{k'w'} = z_{k'w'}]$$

where $z_{k'w'} = \{0, 1\}$. Equivalently, the effect for an individual randomly drawn from the estimation sample is

$$\beta_{kw} = E[\beta_{kw}(h)] = E[Y/Z_{kw} = 1] - E[Y/Z_{kw} = 0] \tag{2}$$

where H comprises X and the set of dummy variables $Z_{k'w'}$ for $k' \neq k$ and $w' = 2, 3, .., W_{k'}$ and expectations are defined over the distribution of H in the estimation sample.

The estimators that we propose for β_{kw} share features with both treatment effects and missing data estimators. We present the estimators in the context of the treatment effects literature and highlight the connections with the missing data estimators when appropiate. The treatment effects literature provides answers to questions concerning the efficacy of a particular programme or policy initiative. The programme

[6] In most cases, the weights are constructed using dummy variables for the sex of the respondent and the age interval he/she belongs to. That is the case, among others, in Tsuchiya et al. (2002) and Kharroubi et al. (2007).

[7] Existence of expectations is assumed throuhout.

at investigation is the treatment, the collective that receives the treatment is the treatment group and those not receiving the treatment are the control or comparison group. In this setting β_{kw} is the average valuation effect of a binary treatment that consists in valuing health states where dimension k reaches level of severity w instead of level 1. The causal interpretation of β_{kw} follows from the assumption that unobserved individual characteristics do not affect health state valuations or their overall average impact is zero.[8] This assumption is also required for the OLS estimate of β_{kw} to have a causal meaning.

Among the available treatment effects estimators we choose the so-called Inverse Probability Weighting (IPW) estimators for three reasons.[9] First, they are easy to implement and provide consistent and in some cases asymptotically efficient estimates of the parameter of interest under standard fairly regularity conditions. Second, they exhibit the best overall finite sample perfomance among the broad class of treatment effect estimators analyzed in Busso et al. (2008). This is particularly relevant for the problem at hand given that the sample used to estimate β_{kw} is modest in size in some empirical applications. Finally, weighting estimators can be used to assess the effect of changes in the distribution of X on the outcome of interest (DiNardo et al., 1996) and, thus, they allow to estimate preference functions for the population of interest from non-representative samples.

Some additional notation is needed at this point. Let $p_{kw}(h) = P(Z_{kw} = 1/H = h)$ be the conditional probability of receiving treatment given H. This variable is named the *propensity score* in the treatment effects literature. The research value of the propensity score rests on its power to solve the dimensionality problem. Adjusting for between-groups differences on a high dimensional vector of covariates can be either difficult or impossible. Rosenbaum and Rubin (1983) show that the propensity score captures all of the variance on the covariates relevant for adjusting between-group comparisons, that is, treated and control units with the same value of the propensity score have the same distribution of the elements in H.[10]

Additionally, the following overlap assumption on the joint distribution of treatments and covariates is necessary for the estimation problem to be well defined:

[8] The assumption is known as *selection on observables* (Barnow, Cain, and Goldberger, 1981), *strong ignorable treatment assignement* (Rosenbaum and Rubin, 1983) or *conditional independence assumption* (Lechner, 1999) and is also implicit in the standard model.

[9] Imbens (2004) provides an overview of the estimators used in the treatment effects literature under the selection on observables assumption.

[10] Properly speaking, Rosenbaum and Rubin (1983) show that if treatment is randomized conditionally on the observed covariates then it is randomized conditional on the propensity score.

$0 < P(Z_{kw} = 1/H) < 1$. Formally, it requires that for a given value of H there is some fraction of the estimation sample in the treatment and control groups to be compared. That is, a necessary condition for the effect of Z_{kw} to be identified is that no other regressor predicts treatment status perfectly. This assumption has implications on the number and selection of health states valued in the sample in order to identify β_{kw}. In particular, it states that there is no level of severity w' in dimension k', for $w' = 1, 2, 3, .., W_{k'}$ and $k' \neq k$, valued only by respondents of a given treatment status. Otherwise, the effect of interest cannot be separatelly identified from that for $Z_{k'w'}$ unless we rely on extrapolation. This is precisely what the standard regression model does. As previously discussed, extrapolation results in biased estimates if differences between the covariate distributions of respondents of different treatment status are relevant and the parametric relationship between the outcome and the regressors is not properly specified.[11]

As an illustrative example, we analyze whether the common support condition holds in the estimation of the EQ-5D tariffs for Holland, Japan, the United Kingdom and Spain in Lamers et al. (2006), Tsuchiya et al. (2002) , Dolan (1997) and Badia et al. (2001), respectively. In particular, we check whether there is overlap in the distribution of $Z_{k'w'}$ for $k' \neq k$ and $w' = 1, 2, 3, .., W_{k'}$ for respondents valuing levels of severity w and 1 in dimension k, for $w = 2, 3, .., W_k$ and $k = 1, 2, ., K$. We find that the validity of the Dutch and Japanese estimates rests on whether the corresponding regression models were correctly specified or not since the common support condition is satisfied in just one out of the ten estimated coefficients.[12] Conversely, the overlap requirement is met in six of the ten estimates in Dolan (1997) and Badia et al. (2001). The difference with the former two studies is in the number of health states for which direct valuations are elicited. While Lamers et al. (2006) and Tsuchiya et al. (2002) use the values obtained for 17 health states to estimate 10 coefficients, Dolan (1997) and Badia et al. (2001) uses 42 health states for the same number of coefficients. Thus, the estimates in the latter two studies are more likely to be robust to the misspecification of the regression model than those in Lamers et al. (2006) and Tsuchiya et al. (2002).

[11] Some studies have reported evidence of misspecification when using the standard model. A non-exhaustive list includes Dolan (1997) and Johnson et al. (1998). Brazier et al. (2002) express their surprise with the result of no specification problems according to the Ramsey RESET test given the skewness of their *RE* estimation residuals. Many other studies simply provide no formal test of misspecification of the standard model.

[12] No mispecification test is reported in these studies.

Moving back to the estimation problem, the expectations in (2) are written as

$$E\left[Y/Z_{kw}=t\right]=\int Y f\left(Y/H\right) g\left(H/Z_{kw}=t\right) dh,\ \text{for } t=\{0,1\} \tag{3}$$

This expression makes it clear that each expectation is calculated using the distribution of H in the collective of respondents of a given treatment status. However, for β_{kw} to be identified we need the same distribution of H in the two expectations. In particular, we use the distribution of H in the estimation sample (respondents valuing levels of severity 1 or w in dimension k). Formally, let $g\left(H\right)$ and $g\left(H/Z_{kw}=t\right)$ be the joint density of H in the estimation sample and in the collective of respondents with treatment status t, respectively, and observe that by definition

$$g\left(H\right)=\frac{g\left(H/Z_{kw}=t\right)P\left(Z_{kw}=t\right)}{P\left(Z_{kw}=t/H\right)},\ \text{for } t=\{0,1\}$$

That is, the distribution of H in the collective of respondents with treatment status t can be changed for the distribution in the estimation sample $g\left(H\right)$ by simply introducing the appropiate weighting function λ_t in (3)

$$\begin{aligned} E\left[Y/Z_{kw}=t\right] &= \int \underbrace{\frac{P\left(Z_{kw}=t\right)}{P\left(Z_{kw}=t/H\right)}}_{\lambda_t} Y f\left(Y/H\right) g\left(H/Z_{kw}=t\right) dh \\ &= \int Y f\left(Y/H\right) g\left(H\right) dh,\ \text{for } t=\{0,1\} \end{aligned}$$

The effect of interest can now be written as

$$\beta_{kw}=E\left[\frac{Z_{kw}Y}{p_{kw}}\right]-E\left[\frac{\left(1-Z_{kw}\right)Y}{1-p_{kw}}\right] \tag{4}$$

This equation suggests immediately the following estimator of β_{kw} which we name as the $IPW1$ estimator

$$\widehat{\beta}_{kw,\,IPW1}=n^{-1}\sum_{i=1}^{n}\frac{Z_{kwi}Y_{ij}}{\widehat{p}_{kwi}}-n^{-1}\sum_{i=1}^{n}\frac{\left(1-Z_{kwi}\right)Y_{ij}}{1-\widehat{p}_{kwi}} \tag{5}$$

This estimator identifies the effect of interest if the estimation sample is representative for the population, that is, if there are no selection biases in the definition of the estimation sample. In that case, the estimator can be implemented by simply obtaining an estimate of the propensity score and pluggin the fitted values for the estimation sample into (5). However, since we cannot be sure *a priori* than the

sample distribution of the elements of X is that in the population, we improve on the latter estimator by accounting for the probability that an individual randomly drawn from the population is in the estimation sample.[13] We do so by rewritting expression (3) so that the two expectations are averaged over the distribution of X in the population. Obviously, the feasibility of this approach rests on whether we have an external representative sample that contains information on X.[14] Conditioned on the availability of such a sample, the effect of interest is written as

$$\beta_{kw} = E\left[\frac{\overline{P}_s D_{kw} Y}{p_{kw} p_s}\right] - E\left[\frac{\overline{P}_s \left(1 - D_{kw}\right) Y}{\left(1 - p_{kw}\right) p_s}\right] \tag{6}$$

where D_s is a binary indicator variable that equals one if the individual is in the estimation sample and zero if he/she is in the external representative sample, $p_s(x) = P(D_s = 1/X = x)$ is the conditional (on X) probability of being in the estimation sample and $\overline{P}_s = P(D_s = 1)$ is the proportion of respondents in the estimation sample. The set of indicator variables $Z_{k'w'}$ for $k' \neq k$ and $w' = 1, 2, 3, .., W_{k'}$ is not included in $p_s(x)$ since representativeness is analyzed with regard to the distribution of individual characteristics.

The sample analog of expression (6) is the $IPW2$ estimator of β_{kw} that is calculated as

$$\widehat{\beta}_{kw,\, IPW2} = n^{-1} \sum_{i=1}^{n} \frac{Z_{kwi} Y_{ij}}{\widehat{p}_{kwi} \widehat{p}_{si}} - n^{-1} \sum_{i=1}^{n} \frac{\left(1 - Z_{kwi}\right) Y_{ij}}{\left(1 - \widehat{p}_{kwi}\right) \widehat{p}_{si}} \tag{7}$$

This equation suggests a simple two-step method to estimate β_{kw}. First, estimate discrete choice models for the two propensity scores and compute the fitted values for the estimation sample.[15] Second, plug the fitted values into the sample analog of (7). Under this scheme, a simple weighted average of the outcome variable recovers the effect of interest.[16] The $IPW1$ estimator weights-down (up) the distribution of

[13] On the one hand, deviations from the original sample design due to, for example, nonresponse issues or to the exclusion of inconsistent respondents might result in non-representative samples. On the other hand, it is difficult or even impossible to define representative samples for the population of interest with regard to the whole set of covariates that have been found to be correlated with health state valuations (Dolan and Roberts, 2002; Kharroubi et al., 2007).

[14] The availability of such a sample is not likely to be a problem in most developed countries. For example, the Census and the European Community Households Panel provide us with the distribution of many sociodemographic, employment and health related individual and household characteristics in the Spanish population. In the empirical section of the paper we use data from the European Community Household Panel for Spain.

[15] The common support condition is tested and imposed in this first step.

[16] The weighting scheme for

health state values for respondents of a given treatment status for those values of the covariates that are (under) over-represented among respondents with that treatment status. Additionaly, the $IPW2$ estimator weights-down (up) the distribution of health state values for sample respondents for those values of the covariates that are (over) under-represented among individuals in the external representative sample.

These two estimators can also be interpreted in the related framework of imputation for missing data.[17] To appreciate this, we follow Rubin (1974) and define β_{kw} in terms of potential outcomes. Let Y_t be the valuation that individual i would have given had he received treatment status t. We only observe the realized outcome $Y = D_s D_{kw} Y_1 + D_s (1 - D_{kw}) Y_0$ but want to know about the effect of the treatment for an individual randomly drawn from the population of interest (β_{kw}). In this setting, β_{kw} is the difference between the population averages of Y_1 and Y_0, which we label μ_1 and μ_0. We only observe Y_1 for treated individuals in the estimation sample and the probability of a 'complete case' i is $p = p_s(x) \times p_{kw}(h)$. As Lunceford and Davidian (2004) point out, weighting by the inverse of the product of propensity scores allows observation i to count for him/herself and $(p^{-1} - 1)$ other 'missing' subjects with like covariates h in estimating μ_1.

The propensity score $p_s(x)$ adjusts for the distribution of individual characteristics in the population. It is a generalization of the traditional corrective sample weights used to recover the representativeness of the standard model estimates once relevant differences are observed between the sample and population distributions of the elements in X.[18] Both methods are equivalent if X includes only dummy indicator variables. In that case, the corrective weights used in the standard framework provide a nonparametric estimate of the propensity score $p_s(x)$. The propensity score allows us to overcome the dimensionality problem in the construction of sample weights and, thus, to control for potential sample selection biases in as many discrete and continuously measured individual characteristics as necessary.

As discussed in Imbens (2004), the estimator in (7) is not an attractive estimator for β_{kw} since the weights for observations of a given treatment status t do not necessarily add up to unity. Indeed, these weights add up to 1 conditioned on treatment

[17] Each of the terms in (7) approximates the average outcome for units of a given treatment status using a weighted sample mean estimator of Horvitz-Thompson type. Horvitz and Thompson (1952) introduced this type of estimator to analyze samples drawn without replacement with unequal selection probabilities from finite universes.

[18] Tsuchiya *et al.* (2002) and Brazier (2006) introduce corrective weights to reflect the non-representative age and sex distribution of their respondents in the standard model and in a non-parametric Bayesian method, respectively.

status t in expectation terms, but because the variance of the sum is positive, the corresponding sample analog is likely to deviate from one. Thus, we normalize the weights to unity and obtain the following $IPW2$ estimator:

$$\widehat{\beta}_{kw,\,IPW2} = \left(\sum_{i=1}^{n} \frac{Z_{kwi}}{\widehat{p}_{kwi}\widehat{p}_{si}}\right)^{-1} \sum_{i=1}^{n} \frac{Z_{kwi}Y_{ij}}{\widehat{p}_{kwi}\widehat{p}_{si}} - \left(\sum_{i=1}^{n} \frac{(1-Z_{kwi})}{(1-\widehat{p}_{kwi})\,\widehat{p}_{si}}\right)^{-1} \sum_{i=1}^{n} \frac{(1-Z_{kwi})\,Y_{ij}}{(1-\widehat{p}_{kwi})\,\widehat{p}_{si}} \tag{8}$$

The $IPW1$ and $IPW2$ estimators can be implemented by producing non-parametric estimates of the propensity scores and plugging the fitted values into (8). However, the number of observations required to attain an acceptable precision for this type of non-parametric estimator increases rapidly with the dimension of X. Moreover, a non-parametric estimate of a conditioned on particular values of X version of these estimators may be difficult to interpret if the dimension of X is larger than two. Furthermore, the net gains of moving from the standard model to an alternative estimator decrease as the implementation of the proposed estimator becomes more challenging. Thus, we focus on semiparametric approximations to the $IPW1$ and $IPW2$ estimators where the propensity scores are parametrically estimated using standard discrete choice models like the logit and probit models.

The $IPW1$ and $IPW2$ estimators are members of a class of semiparametric consistent estimators developed in Robins et al. (1994) for general missing data problems. Robins et al. (1994) show that the estimator within the class having the smallest large-sample variance combines regression on the covariates and propensity score weighting. Contrary to the parametric standard model, the regression model in the semiparametric efficient estimator is incorporated only as a way of gaining efficiency over the $IPW1$ and $IPW2$ estimators, that will still be consistent. The asymptotically efficient estimator is doubly robust in the sense that it provides consistent estimates of β_{kw} if either the propensity score or the regression model are correctly specified. Anyway, the double robust estimator cannot be implemented here because we cannot regress health state values on H in each subsample with treatment status t given that respondents do not value health states defined by $Z_{k'w'}$ for any $k' \neq k$ and any $w' = 2, 3, .., W_{k'}$. Indeed, we can just regress health state values on X and some elements of Z for respondents with treatment status t, where that subset of elements of Z is likely to vary with treatment status t and also with the estimation subsamples used to identify each element of β. Anyway, as shown in Busso et al.

(2008), the small sample properties of the double robust estimator are close to those for propensity score weighting estimators like $IPW1$ and $IPW2$, with the former estimator being slightly more variable and more biased than those implemented in this study.

The consistency and large sample properties of the $IPW1$ and $IPW2$ estimators are derived in the Appendix using the theory of M-estimation and the requirement of the intercept being equal to unity is satisfied using the transformed outcome variable $Y^* = Y - 1$ instead of the original one.

3 Data

A representative sample by gender and age groups of the general population has been used in this study. The sample was divided into 17 subsamples (n = 60) each of them representative of the general population in terms of gender and age groups. In our study we have applied an inter-sample design, that is, the set of 78 health states has been split up in subsets of lower size that have been distributed among subsamples for their valuation. Each of the 17 groups of respondents valued a different subset of five health states so that representativeness by age and gender is held for all the states evaluated. Moreover, all the subsets of five states include a range of severity from mild to more severe health conditions.

The questionnaire.

Each interview began with an introduction in which the SF-6D classification system was explained to the individuals through a 'tutorial'. Once the respondent had been understood the meaning of the dimensions and levels of the instrument, he/her was presented the following tasks:

a) Visual analogue scale (VAS) evaluations of 5 hypothetical health states defined by the SF-6D. The states were labelled as V, W, X, Y, Z and presented in the form of cards. In these cards, besides the statement describing the level on each of the dimensions, the number of the level (from 1 to 4, 5 or 6, depending on the dimension) appeared on a coloured gradient that went from green (no problem) to red (a serious problem). The extremes of the thermometer were 'the worst possible health state' and 'the best conceivable health state'. Respondest had to valued in the scale the five states and the state 'dead'.

b) Probability lottery equivalence (PLE) evaluation of the five hypothetical health states. Respondents were asked for the value of p that made them indifferent

between the following prospects:

$$(FH, p, Death)\tilde{}(FH, 0.5, h)$$

Values of p > 0.5 are expected in most cases, which implies that the respondent assumes a certain risk of death (1-p) < 0.5 in exchange for avoiding a 50% risk of living in the intermediate health state for the rest of his/her life. Nevertheless, some health sates may be considered worse than death, so p would be less than 0.5. Initially, p was fixed in 0.5 to know if the respondent considers that the health state is better or worse than death because. In the fist case, the individual should prefer the lottery on the right side, whereas the lottery on the left should be chosen if the state is considered worse than death. Although the framing of the method is the same in both cases, it is not the same the way in which the value of p is changing in order to reach the indifference point.

In the final part of the interview information about the individual's health state and his/her socioeconomic characteristics (sex, age, studies, income level, etc.) was collected. Three instruments were used to ask the respondents how healthy they felt: the EQ-5D self-report questionnaire, the SF-36 (v.2) questionnaire and a visual scale similar to that presented previously for the valuation of the hypothetical states.

Within each of the 17 subsets of five health states, some of them can be logically ranked, so that a health state can be viewed as 'logically' worse (or better) than another health state. That is the case when a health state has equal or higher levels than another state in each of its six dimensions. Should a respondent assign higher VAS evaluations or PLE evaluations for states that were logically worse than other which received lower valuations, he/she is considered inconsistent and therefore excluded from the analysis. Additionally, individuals who refused to assume any amount of death risk in PLE questions, at least in three out of five elicitations, were also excluded.

4 Estimation results

We first analyze the effect of the respondents' characteristics on health state values and then we compare the estimates of β obtained using the parametric and semiparametric estimators in the preceeding section.

4.1 The valuation effect of individual characteristics

In Table 2 we present OLS and RE estimates of the effect of individual characteristics on health state values. The estimates of β are discussed in the next subsection. The reported estimates indicate that there is a relevant correlation between health state values and the respondents' background characteristics even after adjusting for differences in the severity of the health states being valued. In particular, we find that health state values are significantly affected by the age and marital status of the respondent and by other household level characteristics like household income and the number of children at home.[19] According to RE estimates, health state valuations are primarily affected by the age and marital status of the respondent and by the level of household income.

The estimated non-linear effect of age implies that valuations increase slowly from the age of 18 to about the age of 47, fall slowly up to about 70 and then fall sharply in later years. This means that a 20 years old individual gives about the same value than an otherwise equivalent 70 years old individual. This non-linear association between the age of the respondent and health state values was also found in Dolan and Roberts (2002) and Kharroubi et al. (2007) for the United Kingdom Time Trade-Off and Standard Gamble valuations of the EQ-5D and SF-6D, respectively.[20]

The estimates in Table 2 indicate that there is a positive, monotone and quantitatively relevant correlation between household income and health state values. The valuations of respondents whose household income ranges between 2.000 and 3.000 euros per month are, on average, 0,040 higher than those of respondents whose household income is below 1.500 euros per month. That difference amounts to 0,053 if the latter group of respondents is compared to those whose household income is above 3.000 euros per month.

The positive association between household income and health state values can be interpreted in the light of the results in Lubetkin et al. (2005). They find a positive, significant and relevant association between personal income and health-related quality of life in a large sample of the United States general population using the EQ-5D. That is, *ceteris paribus* and on average terms, high-income people enjoy better health than low-income people and, thus, they are more likely to assign a

[19] The significant OLS estimates obtained for the sex and educational level of the respondent are not confirmed in the RE estimation.

[20] The age that maximizes health state valuations is about 45 years in Dolan and Roberts (2002) and between 60 and 65 years in Kharroubi et al. (2007).

low chance to the event of a bad health outcome when it is presented to them. Moreover, even if respondents judge the plausability of health states independently of their disposable income, the negative consecuences of the realization of a bad health outcome are likely to be quite different for low- and high-income individuals. The positive found for household income in Table 2 is compatible with the hypothesis that respondents value health states according to the expected utility losses in case of realization of that health state.

The estimates in Dolan and Roberts (2002) confirm the relevance of the respondent's marital status for health state valuations. However, the estimated coefficient in that study is of lower magnitude and opposite direction than that in Table 2. While we find that the valuations of married or cohabiting people are, on average, 0,067 higher than the valuations of single people, Dolan and Roberts (2002) find that the average valuation of the latter collective is 0,006 higher than that of the former one.

Regarding the effect of children, Kharroubi et al. (2007) find no significant association between the presence of children under 16 years in the household and health state values. We reach the same conclusion when controlling for whether there is a child aged under 12 years in the household or not.[21] However, when we allow for the effect of children to vary with the number of children in the household we find that having three or more children exerts a negative, quantitatively relevant and significant effect on health state values.[22]

Although existing studies disagree on the magnitude and sign of the effect of some covariates, they provide robust evidence on the importance of accounting for individual and household characteristics when estimating preference-based value functions.[23]

4.2 Semiparametric estimates

The semiparametric estimates of β are presented in Table 3. For comparability purposes, the OLS and RE estimates in Table 2 are reproduced in columns 1 and 2, respectively. Columns 3 and 4 present $IPW1$ and $IPW2$ estimates obtained using the set of covariates in Table 2. Additionally, in columns 5 and 6 we present $IPW1$ and $IPW2$ estimates calculated using a restricted version of matrix X that only

[21] These estimates are available upon request to the authors.

[22] The proportion of respondents with three or more children aged under 12 years is 5,87 percent.

[23] It is beyond the scope of this paper to explain the discrepancies in the results of existing studies. They might totally or partially reflect cross-country differences in the effect of interest or in the distribution of covariates, differences in the estimated specifications or in the estimation methods used.

includes two discrete variables: the respondent's sex and age interval.[24]

Let us first comment on the parametric estimates. There are no inconsistencies in the estimated coefficients and both the OLS and RE estimates indicate that being limited in the kind of work or other activities as a result of physical health ($RL2$) has no significant effect on health state valuations.[25] Moreover, there is no clear direction of change in the estimated β when taking into account that the same individual values several health states, that is, when moving from the OLS to the RE estimates. The average difference between the OLS and RE estimates of β_{kw}, for $w = 2, 3, .., W_k$ and $k = 1, 2, ., K$, is almost zero.

On the contrary, we find substantial differences between the parametric and the semiparametric estimates of β. First, the semiparametric estimates suggest that a departure from full health translates into a significantly lower valuation independently of the severity of the departure and the dimension of health considered. In other words, the semiparametric estimate of β_{kw} is significantly different from zero for any k and any w. Second, the semiparametric model produces estimates of β_{kw} of higher magnitude than those obtained using the standard model. That is the case in 21out of the 25 estimated coefficients which are significantly different from zero in both models. Moreover, the magnitude of the discrepancy between the standard and the semiparametric estimates is negatively related to the severity of the departure from full health. For example, the $IPW2$ estimates for the coefficients associated to levels 2, 3, 4 and 5 of the Vitality dimension are 188, 120, 84 and 56 percent higher, respectively, than the corresponding RE estimates.

As discussed in the preceeding section, there are important differences between the standard and the semiparametric models that can explain the discrepancies in the estimates of β. First, the samples used to estimate β_{kw} are different in both models. While any individual in the sample contributes to the estimation of β_{kw} in the standard model, the estimation sample in the semiparametric model is restricted to respondents valuing levels of severity 1 or w in dimension k. However, the practical relevance of this difference depends on whether the standard model estimates rests on extrapolation and on the differences in the covariate distributions between respondents. Since the overlap assumption holds in this application, we expect this

[24] The distribution of individuals characteristics in the Spanish population is approximated using those in the Spanish sample of the European Community Household Panel for the year 2001, the last year for which we have data.

[25] An inconsistency occurs if the coefficient estimated for Z_{kw} is not higher than that estimated for $Z_{kw'}$, for $w' > w$. Había excepciones.

argument to have little to do with the discrepancies between the parametric and semiparametric estimates in Table 3.

Another reason why the semiparametric estimates might differ from the standard regression estimates is that the two estimation strategies use different weighting schemes. As discussed in Angrist and Krueger (1998), while the semiparametric estimator combines covariate-value-specific estimates of the effect of interest, regression estimators produce a variance-weighted average of these effects.[26] That is, while regression estimators like the OLS and the RE estimators weight each covariate-specific estimate by $P(H = h/Z_{kw} = 1)(1 - P(H = h/Z_{kw} = 1))$, the weights underlying the semiparametric estimator are proportional to the propensity score. So respondents with a higher probability of being "treated" get the most weight in the semiparametric estimates. In contrast, regression estimators weight each covariate-specific estimate by the conditional variance of treatment assignment, which in this case is maximized when $P(Z_{kw} = 1/H) = 0.5$. Obviously, the difference in weighting schemes is of no importance if the estimated coefficient does not vary with the elements of H. Thus, the observed discrepancy between the standard and semiparametric estimates can be thought as evidence that the health state valuation impact of a given departure from full health is heterogeneous in the respondent's background characteristics and in the particular level of severity in other dimensions of health.

Next, the differences between the $IPW1$ and $IPW2$ estimates in columns 3 and 4 suggest that the distribution of the elements in X is imbalanced between the population and the estimation sample.[27,28] These differences are much lower in magnitude than those found when comparing from the standard to the semiparametric estimates. On average, the estimate of β_{kw} increases by 3 percent in absolute value when using the distribution of covariates in the Spanish population instead of that in the corresponding estimation sample.

The results of estimating the propensity score $p_s(x)$ can be used to identify the covariates whose sample distribution differs from that in the population.[29] A significant

[26] Angrsit and Krueger (1994) provide a general discussion on the weighting schemes underlying matching and regression estimators.

[27] This finding is common to the multiple specifications of the propensity score used to improve its balancing power.

[28] Finding no relevant discrepancies between $IPW1$ and $IPW2$ estimates indicates that either the effect of interest does not vary with individual characteristics or that is heterogeneity in the effect of interest but the distribution of the elements in X is the same in the sample and population of interest. However, the former scenario is rejected in the preceeding paragraph.

[29] Crump et al. (2008) show that limited overlap in the covariate distributions between treatment groups can led to imprecise estimates and can make estimates sensitive to the choice of specifications

coefficient in the logit estimates denotes that the distribution of the corresponding covariate is not balanced between the population and the estimation sample. As an illustrative example, Table 4 resumes the logit estimation of $p_s(x)$ for the estimation of the coefficient associated to variable $PF2$. We find that the sample distribution of any covariate but the respondents' sex is significantly different from the corresponding distribution in the Spanish sample of the European Community Household Panel for the year 2001. This is a common finding to the estimation of $p_s(x)$ for almost any of the $IPW2$ estimates in Table 3.[30]

The result that adjusting for the population distribution of the covariates results in relevant variations in the magnitude of the estimates contrasts with findings in Kharroubi et al. (2007) and Dolan and Roberts (2002). Both studies find that the standard model estimates are almost invariant to the inclusion of corrective weights that adjust to the age interval and sex distribution in the population. Interestingly, we reach the same conclusion when we restrict the covariates in X to a set of dummy indicator variables for the sex of the respondent and the age interval he/she belongs to.[31] As shown in columns 5 and 6 of Table 3, the resulting $IPW1$ and $IPW2$ estimates of β_{kw} are almost identical for any k and any w.

Interestingly, these estimates are close in magnitude to the $IPW1$ estimates in column 3 obtained adjusting for the distribution of the expanded list of covariates in the collective of "treated" respondents. On the contrary, we find relevant differences between the estimates in columns 5 and 6 and the $IPW2$ estimates in column 4 where we adjust for the population distribution of the expanded list of covariates. These results suggest that conditioning on a reduced set of covariates suffices to remove much of the bias due to compositional differences between the collectives of different treatment status compared in the estimation of β_{kw}.[32] However, they also suggest that the traditional approach of using corrective weights constructed using a reduced set of discrete variables does not exploit much of the variance in the distribution of

for the first step propensity score model. We address this issue by checking the sensitivity of the estimates to discarding all units with estimated propensity score outside the range $[0+k, 1-k]$, for $k = \{0.05, 0.1\}$. The estimates of β remain almost unchanged. These results are available upon request to the authors.

[30] These estimates and also those for the first step estimate of the propensity score $p_{kw}(x)$ are available upon request to the authors.

[31] We also obtain standard model estimates using corrective weights constructed using dummy indicator variables of the age and sex of the respondent. As in Kharroubi *et al.* (2007) and Dolan and Roberts (2002), these estimates are almost identical to the unweighted estimates in Table 2. These estimates are available upon request to the authors.

[32] Cochran

individual characteristics and, thus, it does not guarantees the population validity of the estimates. The continuous generalization of that procedure that we propose seems far more effective in removing sample selection biases and adjusting for the distribution of X in the population.[33]

Following Brazier and Roberts (2004), in Table 5 we present parsimonious consistent models estimated by aggregating levels of a given dimension if inconsistencies ocurred in the $IPW2$ estimates in column 4 of Table 3, that is, if the semiparametrically estimated coefficient for Z_{kw} is not higher than that estimated for $Z_{kw'}$, for $w' > w$. In particular, we obtained three inconsistencies. The estimate for $PAIN2$ is larger in absolute value than that for $PAIN3$, the coefficient for $MH4$ is smaller in absolute value than that associated to $MH3$ and, finally, the estimate for $VIT4$ is slightly larger in absolute value than that for $VIT5$. As opposed to the standard models, the semiparametric model does not require the full vector β to be reestimated when an inconsistency is detected. Previously discussed features of the semiparametric model and its differences with the standard model remain true when analyzing the consistent models.

The parsimonious models are used to estimate values for each of the 18.000 health states that can be defined using the SF-6D classification system. The resulting OLS, RE, $IPW1$ and $IPW2$ estimated tariffs are summarized in Table 6. First, we find that the OLS and RE estimates produce almost identical values. Conversely, and as expected given the preceeding discussion, there are substantial differences between the standard and the semiparametric predicted values. These discrepancies are observed in any of the distributional moments reported in Table 6 and they tend to be higher the lower is the value assigned to a particular health state, that is, the higher is its severity according to the respondents' preferences. For example, the difference between the zth percentiles of the RE and $IPW1$ distributions of values lowers from -0,160 for $z = 10$ to -0,135 for $z = 90$. In relative terms, the 10th and 90th percentiles of the $IPW1$ distribution are 101,5 and only 18,9 percent higher than the corresponding moments in the RE distribution. These differences are larger when the RE distribution is compared to the $IPW2$ one. The difference between the zth percentiles of the RE and $IPW2$ distributions of values are -0,320 (202,5 percent higher than the RE value) and -0,222 (31 percent higher than the RE value) for $z = 10$ and 90, respectively.

[33] This explanation solves the puzzle pointed out in Kharroubi *et al.* (2007). They find that the age of the respondent is a major determinant of health state values but also that adjusting to the age and sex distribution of the UK population has a little impact on the estimates.

As observed, the positive correlation between, on the one hand, the magnitude of the difference between the standard and the semiparametric values and, on the other hand, the severity of the health state is even higher when this correlation is analyzed using the $IPW2$ estimates instead of the $IPW1$ estimates.

A similar picture emerges when looking at the proportion of predicted negative values. The numbers for the OLS and RE values are 2,6 and 2,5, respectively, very far from those for the $IPW1$ and, in particular, for the $IPW2$ estimates, where 10,2 and 25,8 percent of the values are negative, respectively.[34] As before, we find that the valuation impact of moving from OLS estimates to a semiparametric estimator with a different weighting scheme like $IPW1$ is substantially higher than that of accounting for individual heterogeneity in the variance with the RE estimator. Additionally, while standard corrective methods suggest that adjusting to the age and sex distribution in the population has almost no effect on the predicted values, the difference in the proportion of negative numbers in the $IPW1$ and $IPW2$ values makes it clear that standard methods fail to correct for sample selection biases.

Finally, the lower bottom of Table 6 analyzes the within sample predictive ability of the models. According to the mean absolute error, the standard model estimates fit better to the values directly provided by respondents than the semiparametric estimators do. Indeed, the mean absolute error for the $IPW1$ and $IPW2$ models are 37 and 72 percent higher than that for the OLS estimates, respectively. When looking at the percentage of predictions that are within $\pm k$ units of the actual value, we find that the standard and $IPW1$ models perform similarly when dealing with high precision predictions ($k = 0.01$). On the contrary, the standard models clearly outperform the $IPW1$ model for less demanding precision predictions. The $IPW2$ model estimates perform worse than the standard ones in any circumstances.

Finding that the standard model fits closer to the original data than the semiparametric one is not a surprising result. As previously discussed, the semiparametric estimators modify the sample distribution of individuals characteristics and responses in order to estimate β. In particular, the $IPW1$ estimator weights-down (weights-up) the distribution of health state values for respondents of a given treatment status for those values of the covariates which are over-represented (under-represented) among

[34] The proportion of negative values when using the EQ-5D estimates for the UK general population in Dolan (1997) to predict the utilities of the 243 health states that can be defined in the EQ-5D classification system is 34.16 percent. The corresponding proportions for the studies that use the EQ-5D system in Spanish (Badia et al., 2001), Dutch (Lamers et al., 2006) and Japanese (Tsuchiya et al., 2002) samples are 37.45, 15.23 and 2.88 percent.

respondents with that treatment status. Obviously, this results in deviations from the original sample distribution of the outcome and independent variables. Furthermore, the $IPW2$ estimator reweights the weighted distribution of covariates to adjust to the population distribution. This accounts for the particularly large mean absolute error calculated for this estimator.

That is, the superiority of the semiparametric model cannot be judge on the basis of within sample criterias. Indeed, its superiority follows from the following theoretical and empirical statements. First, it provides a flexible way of accomodating covariates and correcting for sample selection biases in multiple continuously measured individual characteristics that matter for health state values. On the contrary, the standard way of correcting for the distribution of the covariates in the population using corrective weights is limited in the number of covariates to include and the empirical estimates in this paper show that it fails to control for sample selection biases. Second, the semiparametric estimators make no assumption on the underlying distribution of health state values and, contrary to other proposed alternatives to the standard model like the nonparametric Bayesian approach in Kharroubi et al. (2005), it is easy to implement and interpret and it provides the user with a simple table of estimated coefficients that defines the estimated preference function, resulting in efficiency and transparency gains.[35] Additionally, bootstrapping methods provide an useful alternative for inference in this setting if the reader is not familiar with the computation of asymptotic standard errors. Third, the semiparametric estimators allow for an undetermined amount of heterogeneity in the effect of interest. This is particularly relevant since we find evidence that the valuation impact of a deviation from full health is likely to vary with the respondent's background characteristics and the levels of severity in the other dimensions of health.

5 Conclusions

This paper presents a new approach to model health state values with important advantadges over the traditional one. In particular, we emphasize the following advantadges. First, our method makes no assumption on the distribution of health state values. This is a relevant issue given the skewed, truncated, non-continuous and hierachical nature of health state valuation data. The normality assumption underly-

[35] The method in Kharroubi *et al.* (2005) estimates a value of β_{kw} for every one of the 18.000 health states that can be defined in the SF-6D system.

ing the standard regression estimates is rejected in most empirical applications that formally test it including ours.

Second, as opposed to the standard regression model, our method accomodates covariates in a flexible way, eschews parametric assumptions on the relationship between the outcome and the regressors and it allows for the valuation impact of a departure form full health in a given dimension to be heterogeneous in individual characteristics and in the severity of the departure from full health in the other dimensions of health. The latter argument is particularly relevant since estimates in this paper confirm that there is a significant amount of heterogeneity in the valuation effect of a departure from full health.

Third, while the standard model estimates are very sensitive to differences in the covariate distributions for respondents valuing different levels of severity in different dimensions, our method highlights the importance of properly selecting the health states that are valued in the sample for identification not to rely on extrapolation.

Fourth, our method produces estimates for the population of interest even if the estimation sample is not representative for that population with regard to many discrete and continuous individual characteristics that affect health state values. The traditional approach of using corrective weights to adjust for the distribution of individual characteristics in the population suffers from the course of dimensionality problem. That is, it becomes more and more cumbersome as the number of discrete variables used to construct corrective weights increases and it cannot accomodate continuous variables. The estimates in this paper show that relevant differences in the distribution of individual characteristics between the sample and the population of interest persist once we adjust the sample for the proportion of individuals of a given sex and age interval in the population. This result explains the paradox commonly found in preceeding studies that the age of the respondent affects health state valuations but the estimates remain almost unchanged when using corrective weights defined over sex and age intervals only.

Fifth, despite all these advantadges the technical complexity of the proposed estimator is only slightly higher than that of the standard regression estimator. In particular, our method requires the estimation of at most two discrete choice probit or logit models that are incorporated in any statistical software. Moreover, bootstrapping methods provides an useful aternative for users not confident with the calculation of asymptotic standard errors.

Sixth, our method is easier to implement and interpret than the nonparametric

Bayesian alternative to the standard regression model that has been recently proposed in Kharroubi et al. (2005). In particular, and contrary to the estimation method in that article, our method provides the user with a simple table of estimated coefficients that defines the estimated preference function, resulting in efficiency and transparency gains.

Regarding the results of the empirical implementation, we find relevant discrepancies between the estimates obtained using our method and those obtained using the standard regression model. In particular, the semiparametric estimates use to be higher in absolute value than the regression estimates, particularly so when adjusting for the distribution of individual characteristics in the Spanish population and when analyzing the valuation effect of small departures from full health. These results suggest that the standard method underestimates the value that the Spanish population assigns to a given departure from full health and, in particular, to small departures. In fact, when these estimates are used to predict values for the 18.000 health states that can be defined using the SF-6D classification system, we find that the percentage of negative predictions is 25.7 percent, that is, 23 percentage points higher than that obtained when using the standard model estimates to predict values.

A Appendix. Asymptotic properties

We derive the asymptotic properties of $\widehat{\beta}_{kw,IPW2}$ and present those of $\widehat{\beta}_{kw,IPW1}$ as a particular case. The subscript $IPW2$ is dropped out to reduce the notation. The properties of $\widehat{\beta}_{kw}$ are derived by viewing it as an M-estimator, that is, as the solution to a set of estimating equations.[36] In particular, $\widehat{\beta}_{kw}$ is one element of the vector $\widehat{\theta}$ that solves the vector equation

$$\sum_{i=1}^{n}\psi\left(W_i,\widehat{\theta}\right)=0$$

where $W_i=[Y_i,Z_{k'w'},X_i]$, for $k'\neq k$ and $w'=2,3,..,W_{k'}$ and $\theta=[\delta,\gamma,\beta_{kw}]$. The vector equation ψ has three equations and can be written as

$$\begin{aligned}
\sum_{i=1}^{n}\psi_1\left(W_i,\theta\right) &= \sum_{i=1}^{n}\frac{D_{si}-p_{si}\left(X_i,\delta\right)}{p_{si}\left(X_i,\delta\right)\left[1-p_{si}\left(X_i,\delta\right)\right]}\frac{\partial p_{si}\left(X_i,\delta\right)}{\partial\delta}=0\\
\sum_{i=1}^{n}\psi_2\left(W_i,\theta\right) &= \sum_{i=1}^{n}\frac{Z_{kwi}-p_{kwi}\left(H_i,\gamma\right)}{p_{kwi}\left(H_i,\gamma\right)\left[1-p_{kwi}\left(H_i,\gamma\right)\right]}\frac{\partial p_{kwi}\left(H_i,\gamma\right)}{\partial\gamma}=0\\
\sum_{i=1}^{n}\psi_3\left(W_i,\theta\right) &= \sum_{i=1}^{n}\left\{A\left(1\right)\frac{Z_{kwi}Y_i}{\widehat{p}_{kwi}\widehat{p}_{si}}-A\left(0\right)\frac{\left(1-Z_{kwi}\right)Y_i}{\left(1-\widehat{p}_{kwi}\right)\widehat{p}_{si}}-\beta_{kw}\right\}=0
\end{aligned}$$

where $p_{si}=p_{si}\left(X_i,\delta\right)$, $p_{kwi}=p_{kwi}\left(H_i,\gamma\right)$, $\widehat{p}_{kwi}=\left(H_i,\widehat{\gamma}\right)$, $\widehat{p}_{si}=\left(X_i,\widehat{\delta}\right)$, $A\left(t\right)=N\left(\sum_{i=1}^{n}\frac{Z_{kwi}^{t}\left(1-Z_{kwi}\right)^{1-t}}{\widehat{p}_{kwi}^{t}\left(1-\widehat{p}_{kwi}\right)^{1-t}\widehat{p}_{si}}\right)^{-1}$ for $t=\{0,1\}$ and N is the total number of individuals in the estimation sample. The solutions to equations $\psi_1\left(W_i,\theta\right)$ and $\psi_2\left(W_i,\theta\right)$ are the maximum likelihood estimates of δ and γ, the coefficients of the binary response models used to estimate the propensity scores p_s and p_{kw}, respectively. We estimate the propensity scores using the logistic regression model, where $p\left(Q,\varphi\right)=\left\{1+\exp\left(-Q^T\varphi\right)\right\}^{-1}$. The solution to equation $\psi_3\left(W_i,\theta\right)$ is the coefficient of interest.

By standard results on M-estimation, under the true parameter value θ

$$\sqrt{n}\left(\widehat{\theta}-\theta\right)\longrightarrow N\left(0,A\left(\theta\right)^{-1}B\left(\theta\right)\left\{A\left(\theta\right)^{-1}\right\}^{T}\right)$$

where

$$A\left(\theta\right)=E\left[-\dot{\psi}\left(W,\theta\right)\right]$$

[36] Stefanski and Boos (2002) provide an excelent review of the theory of M-estimation. Additionally, Lunceford and Davidian (2004) derive the asymptotic properties of the $IPW1$ estimator.

with $\dot{\psi}(W,\theta) = \partial\psi(W,\theta)/\partial\theta^T$ and

$$B(\theta) = E\left[\psi(W,\theta)\,\psi(W,\theta)^T\right]$$

To estimate the asymptotic variance use

$$\begin{aligned}
\widehat{A} &= \frac{1}{n}\sum_{i=1}^{n} -\frac{\partial\psi\left(W_i,\widehat{\theta}\right)}{\partial\theta^T} \\
\widehat{B} &= \frac{1}{n}\sum_{i=1}^{n}\psi\left(W_i,\widehat{\theta}\right)\psi\left(W_i,\widehat{\theta}\right)^T
\end{aligned}$$

where the derivative of ψ can be calculated as

$$\frac{\partial\psi(W,\theta)}{\partial\theta^T} = \begin{pmatrix} \frac{\partial\psi_1(W,\theta)}{\partial\delta^T} & \frac{\partial\psi_1(W,\theta)}{\partial\gamma^T} & \frac{\partial\psi_1(W,\theta)}{\partial\beta_{kw}^T} \\ \frac{\partial\psi_2(W,\theta)}{\partial\delta^T} & \frac{\partial\psi_2(W,\theta)}{\partial\gamma^T} & \frac{\partial\psi_2(W,\theta)}{\partial\beta_{kw}^T} \\ \frac{\partial\psi_3(W,\theta)}{\partial\delta^T} & \frac{\partial\psi_3(W,\theta)}{\partial\gamma^T} & \frac{\partial\psi_3(W,\theta)}{\partial\beta_{kw}^T} \end{pmatrix}$$

where

$$\begin{aligned}
A_{11i} &= \frac{\partial\psi_1(W,\theta)}{\partial\delta^T} = -\frac{1}{p_{si}(1-p_{si})}P_\delta P_\delta^T \\
A_{12i} &= \frac{\partial\psi_1(W,\theta)}{\partial\gamma^T} = A_{13i} = \frac{\partial\psi_1(W,\theta)}{\partial\beta_{kw}^T} = 0 \\
A_{21i} &= \frac{\partial\psi_2(W,\theta)}{\partial\delta^T} = A_{23i} = \frac{\partial\psi_2(W,\theta)}{\partial\beta_{kw}^T} = 0 \\
A_{22i} &= \frac{\partial\psi_2(W,\theta)}{\partial\gamma^T} = -\frac{1}{p_{kwi}(1-p_{kwi})}P_\gamma P_\gamma^T \\
A_{31i} &= \frac{\partial\psi_3(W,\theta)}{\partial\delta^T} = -\left[\frac{D_{si}D_{kwi}Y_i^*}{p_{kwi}{p_{si}}^2} - \frac{D_{si}D_{kwi}Y_i^*}{(1-p_{kwi}){p_{si}}^2}\right]P_\delta \\
A_{32i} &= \frac{\partial\psi_3(W,\theta)}{\partial\gamma^T} = -\left[\frac{D_{si}D_{kwi}Y_i^*}{p_{kwi}^2 p_{si}} + \frac{D_{si}D_{kwi}Y_i^*}{(1-p_{kwi})^2 p_{si}}\right]P_\gamma \\
A_{33i} &= \frac{\partial\psi_3(W,\theta)}{\partial\beta_{kw}^T} = -1
\end{aligned}$$

where $P_\delta = \partial/\partial\delta\{p_{si}\}$, $P_\gamma = \partial/\partial\gamma\{p_{kwi}\}$ and $Y_i^* = D_{kwi}A(1)Y_i + (1-D_{kwi})A(0)Y_i$. Equivalently, the elements of B are calculated as

$$
\begin{aligned}
B_{11i} &= \psi_1(W_i,\theta)\,\psi_1(W_i,\theta)^T = \frac{1}{p_{si}(1-p_{si})}P_\delta P_\delta^T \\
B_{12i} &= \psi_1(W_i,\theta)\,\psi_2(W_i,\theta)^T = 0 \\
B_{13i} &= \psi_1(W_i,\theta)\,\psi_3(W_i,\theta)^T = \left[\frac{D_{si}D_{kwi}Y_i^*}{p_{kwi}{p_{si}}^2} - \frac{D_{si}D_{kwi}Y_i^*}{(1-p_{kwi}){p_{si}}^2}\right]P_\delta \\
B_{21i} &= \psi_2(W_i,\theta)\,\psi_1(W_i,\theta)^T = 0 \\
B_{22i} &= \psi_2(W_i,\theta)\,\psi_2(W_i,\theta)^T = \frac{1}{p_{kwi}(1-p_{kwi})}P_\gamma P_\gamma^T \\
B_{23i} &= \psi_2(W_i,\theta)\,\psi_3(W_i,\theta)^T = \left[\frac{D_{si}D_{kwi}Y_i^*}{p_{kwi}^2 p_{si}} + \frac{D_{si}(1-D_{kwi})Y_i^*}{(1-p_{kwi})^2 p_{si}}\right]P_\gamma \\
B_{31i} &= \psi_3(W_i,\theta)\,\psi_1(W_i,\theta)^T = B_{13i}^T \\
B_{32i} &= \psi_3(W_i,\theta)\,\psi_2(W_i,\theta)^T = B_{23i}^T \\
B_{33i} &= \psi_3(W_i,\theta)\,\psi_3(W_i,\theta)^T = \left(\frac{D_{si}D_{kwi}Y_i^*}{p_{kwi}p_{si}} - \frac{D_{si}(1-D_{kwi})Y_i^*}{(1-p_{kwi})p_{si}} - \beta_{kw}\right)^2
\end{aligned}
$$

Finally, it can be shown that the large-sample variance of β_{kw} is

$$
V(\beta_{kw}) = A_{33}^{-1}\left(B_{33} - B_{23}^T B_{22}^{-1} B_{23} - B_{13}^T B_{11}^{-1} B_{13}\right)\left(A_{33}^{-1}\right)^T
$$

The expression of the large-sample variance of β_{kw} in the case where γ and δ are known is $A_{33}^{-1}B_{33}\left(A_{33}^{-1}\right)^T$. The additional two terms in the parenthesis are the adjustment in the large-sample variance of the effect of interest coming from the first step estimation of the two propensity scores. Interestingly, it results that estimation of the propensity scores leads to smaller large-sample variance for these IPW estimators than using the true values. That is, as Lunceford and Davidian (2004) point out, even if the functional form of the propensity score is known exactly, it is benefitial from an efficiency pointview to estimate it. Hirano, Imbens and Ridder (2003) explain this result in the context of the Generalized Method of Moments and the Empirical Likelihood estimators.

The expression for the variance of the $IPW1$ estimator includes only the first two terms in the parenthesis in the latter expression, where B_{23} and B_{33} are now calculated as the sample average of the following expressions evaluated at the estimated value of the elements of θ

$$
\begin{aligned}
B_{23i} &= \left[\frac{D_{kwi} Y_i^*}{p_{kwi}^2} + \frac{(1 - D_{kwi})\, Y_i^*}{(1 - p_{kwi})^2} \right] P_\gamma \\
B_{33i} &= \left(\frac{D_{kwi} Y_i^*}{p_{kwi}} - \frac{(1 - D_{kwi})\, Y_i^*}{(1 - p_{kwi})} - \beta_{kw,IPW1} \right)^2
\end{aligned}
$$

B Appendix. Variable definitions and sources

The statistics for the Spanish population are constructed using data from the European Community Household Panel (ECHP) for the year 2001, and they are provided by Eurostat. In the empirical analysis we control for the sex and age (in years) of the respondent, whether he/she is married or cohabiting ($MarStat1$) or separated, divorced or widow ($MarStat2$) and whether the respondent has attained a secondary level of education ($Mid\text{-}Educ$) or an university degree ($High\text{-}Educ$). We also classify respondents according to whether their monthly total household income is below 1500 euros, between 1500 and 2000 euros ($Income2$), between 2000 and 3000 euros ($Income3$) or above 3000 euros ($Income4$). Regarding their smoking behaviour, we distinguish between non-smokers and respondents who actually smoke less than 10 cigarettes per day ($Smoke2$), between 10 and 20 cigarettes ($Smoke3$) and more than 20 cigarettes per day ($Smoke4$). Additionally, we construct two dummy variables that indicate if the respondent thinks that his/her health is fair ($Own2$) or bad/very bad ($Own3$). The other two answers to the question of how is your health in general are good and very good. Finally, we control for the number of children in the household. Importantly, while the variable constructed using ECHP data refers to children under the age of 16 years, the corresponding variable from the collected data refers to children under the age of 12.

References

[1] Angrist, J.D., Krueger, A.B., 1998. Empirical Strategies in Labor Economics. Handbook of Labor Economics, vol. 3A, edited by Orley Ashenfelter and David Card, Amsterdam. Elsevier.

[2] Badia, X., Roset, M., Herdman, M., Kind, P., 2001. A Comparison of United Kingdom and Spanish General Population Time Trade-Off Values for EQ-5D Health States. Medical Decicion Making 21, 7-16.

[3] Barnow, B., Cain C., Goldberger, A., 1981. Selection on Observables. Evaluation Studies Review Annual 5, 43-59.

[4] Brazier, J.E., Roberts, J., 2004. The Estimation of a Preference-Based Measure of Health From the SF-12. Medical Care 42 (9), 851-859.

[5] Brazier, J.E., Roberts, J., Deverill, M., 2002. The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics 21, 271-292.

[6] Busso, M., DiNardo, J., McCrary, J., 2009. New Evidence on the Finite Sample Properties of Propensity Score Matching and Reweighting Estimators. IZA Discussion Paper 3998.

[7] Crump, R.K., Hotz, V.J., Imbens, G.W., Mitnik, O.A., 2009. Dealing with limited overlap in estimation of average treatment effects. Biometrika 96, 187-199.

[8] DiNardo, J., Fortin, N.M., Lemieux, T., 1996. Labor Market Institutions and the Distribution of Wages, 1973-1992: a Semiparametric Approach. Econometrica 64 (5), 1001-1044.

[9] Dolan, P., 1997. Modelling valuations for Euroqol health states. Medical Care 35, 351-363.

[10] Dolan, P., Gudex, C., Kind, P., Williams, A., 1996. The Time Trade-Off Method: Results From a General Population Study. Health Economics 5, 141-154.

[11] Dolan, P., Roberts, J., 2002. To what extent can we explain time trade-off values from other information about respondents?. Social Science and Medicine 54, 919-929.

[12] Gold, M.R., Siegel, J.E., Russell, L.B., Weinstein, M.C., 1996. Cost-Effectiveness in Health and Medicine. Oxford University Press, Oxford.

[13] Hirano, K., Imbens, G.W., Ridder, G., 2003. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. Econometrica 71 (4), 1161-1189.

[14] Horvitz, D.G., Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association 47, 663-685.

[15] Imbens, G.W. 2004. Nonparametric Estimation of Average Treatment Effects Under Exogeneity: a Review. The Review of Economics and Statistics 86 (1), 4-29.

[16] Johnson, J.A., Coons, S.J., Ergo, A., Szava-Kovats, G., 1998. Valuation of EuroQOL (EQ-5D) Health State in an Adult US Sample. Pharmaconomics 13 (4), 421-433.

[17] Kharroubi, S., Brazier, J.E., O'Hagan, A., 2007. Modelling covariates for the SF-6D standard gamble health state preference data using a nonparametric Bayesian method. Social Science and Medicine 64, 1242-1252.

[18] Kharroubi, S., Brazier, J.E., Roberts, J., O'Hagan, A., 2007. Modelling SF-6D health state preference data using a nonparametric Bayesian method. Journal of Health Economics 26, 597-612.

[19] Lamers, L.M., McDonnell, J., Stalmeier, F.M., Krabbe, P.F.M., Busschbach, J.J.V., 2006. The Dutch Tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Economics 15, 1121-1132.

[20] Lechner, M., 1999. Earnings and Employment Effects of Continuous On-the-Job Training Programs in East Germany. Journal of Business and Economic Statistics 17, 74-90.

[21] Lubetkin, E.I., Jia, H., Franks, P., Gold, M.R., 2005. Relationship among sociodemographic factors, clinical conditions, and health related quality of life: Examining the EQ-5D in the U.S. general population. Quality of Life Research 14, 2187-2196.

[22] Lunceford, J.K., Davidian, M., 2004. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Statistics in Medicine 23, 2937-2960.

[23] National Institute for Clinical Excellence (NICE). Guide to the Methods of Technology Appraisal. London: NICE, 2004.

[24] Robins, J.M., Rotnitzky, A., Zhao, L.P., 1994. Estimation of regression coefficients when some regressors are not always observed. Journal of the American Statistical Association 89, 846-866.

[25] Rosenbaum, P., Rubin, D.B., 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika 70, 41-55.

[26] Rubin, D.R., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychologies 66, 688-701.

[27] Stefanski, L.A., Boos, D.D., 2002. The calculus of M-estimation. The American Statistician 56, 29-38.

[28] Tsuchiya, A., Ikeda, S., Ikegami, N., Nishimura, S., Sakai, I., Fukuda, T., Hamashima, C., Hisashige, A., Tamura, M., 2002. Estimating an EQ-5D population value set: the case of Japan. Health Economics 11, 341-353.

Table 1. Characteristics of sample respondents and Spanish Population

	Sample	Population[a]
Female	50.0	52.05
Age	43.60	46.97
	(16.64)	(19.04)
MarStat1	59.84	63.59
MarStat2	6.53	11.25
Mid-Educ	34.54	17.56
High-Educ	31.02	20.55
Children (presence)	48.80	25.64
Children (number)	1.82	1.41
	(0.66)	(0.63)
Income2	28.31	17.39
Income3	29.82	21.76
Income4	18.98	11.47
Smoke2	16.57	9.19
Smoke3	8.63	13.52
Smoke4	1.71	4.74
Own2	10.44	22.25
Own3	1.20	10.72
N	4980	11515

Notes: The table reports percentages for discrete variables and means and standard errors (in brackets) for continuous variables. [a] The statistics are calculated using data from the European Community Household Panel for the year 2001.

Table 2. OLS and RE estimates.

	OLS	RE
c	1.000	1.000
PF2	-0.016	-0.025**
PF3	-0.031***	-0.054***
PF4	-0.088***	-0.118***
PF5	-0.103***	-0.106***
PF6	-0.332***	-0.333***
RL2	-0.014	0.005
RL3	-0.041***	-0.046***
RL4	-0.078***	-0.091***
SF2	-0.036***	-0.070***
SF3	-0.063***	-0.079***
SF4	-0.203***	-0.194***
SF5	-0.210***	-0.240***
PAIN2	-0.016	-0.043***
PAIN3	-0.033***	-0.048***
PAIN4	-0.202***	-0.174***
PAIN5	-0.208***	-0.232***
PAIN6	-0.318***	-0.342***
MH2	-0.064***	-0.025***
MH3	-0.080***	-0.050***
MH4	-0.096***	-0.073***
MH5	-0.226***	-0.197***
VIT2	-0.055***	-0.042***
VIT3	-0.120***	-0.094***
VIT4	-0.154***	-0.155***
VIT5	-0.197***	-0.180***

Notes: *, ** and *** denote significance at the 10%, 5% and 1% level, respectively.

Table 2 (cont). *OLS* and *RE* estimates.

	OLS	RE
Sex	0.015**	0.015
Age	-0.003***	-0.003**
Age sq.	0.003***	0.003**
MarStat1	0.059***	0.063***
MarStat2	-0.014	-0.018
Mid-Educ	0.003	-0.002
High-Educ	-0.020**	-0.022
Children [a]	-0.010**	-0.010*
Income2	0.021**	0.025*
Income3	0.036***	0.039**
Income4	0.055***	0.051***
Smoke2	-0.001	-0.011
Smoke3	0.026**	0.030
Smoke4	-0.047*	-0.054
Own2	0.007	0.011
Own3	0.038	0.041
Adj. R^2	0.856	
N	4980	4980

Notes: *, ** and *** denote significance at the 10%, 5% and 1% level, respectively. [a] Number of children in the household.

Table 3. Parametric and Semiparametric estimates.

	OLS	RE	$IPW1$	$IPW2$	$IPW1r$ [a]	$IPW2r$ [a]
c	1.000	1.000	1.000	1.000	1.000	1.000
PF2	-0.016	-0.025**	-0.069***	-0.045*	-0.068***	-0.068***
PF3	-0.031***	-0.054***	-0.091***	-0.085***	-0.108***	-0.108***
PF4	-0.088***	-0.118***	-0.134***	-0.142***	-0.141***	-0.141***
PF5	-0.103***	-0.106***	-0.171***	-0.156***	-0.199***	-0.199***
PF6	-0.332***	-0.333***	-0.300***	-0.336***	-0.290***	-0.291***
RL2	-0.014	0.005	-0.062***	-0.078***	-0.063***	-0.063***
RL3	-0.041***	-0.046***	-0.102***	-0.133***	-0.097***	-0.097***
RL4	-0.078***	-0.091***	-0.156***	-0.182***	-0.122***	-0.122***
SF2	-0.036***	-0.070***	-0.055***	-0.066***	-0.058***	-0.058***
SF3	-0.063***	-0.079***	-0.079***	-0.071**	-0.084***	-0.084***
SF4	-0.203***	-0.194***	-0.222***	-0.249***	-0.242***	-0.242***
SF5	-0.210***	-0.240***	-0.236***	-0.249***	-0.237***	-0.237***
PAIN2	-0.016	-0.043***	-0.109***	-0.137***	-0.112***	-0.113***
PAIN3	-0.033***	-0.048***	-0.074***	-0.039*	-0.084***	-0.083***
PAIN4	-0.202***	-0.174***	-0.206***	-0.241***	-0.201***	-0.201***
PAIN5	-0.208***	-0.232***	-0.284***	-0.327***	-0.280***	-0.280***
PAIN6	-0.318***	-0.342***	-0.361***	-0.403***	-0.361***	-0.361***
MH2	-0.064***	-0.025***	-0.100***	-0.063**	-0.100***	-0.100***
MH3	-0.080***	-0.050***	-0.164***	-0.184***	-0.176***	-0.176***
MH4	-0.096***	-0.073***	-0.141***	-0.071**	-0.149***	-0.149***
MH5	-0.226***	-0.197***	-0.315***	-0.350***	-0.317***	-0.317***
VIT2	-0.055***	-0.042***	-0.097***	-0.121***	-0.107***	-0.107***
VIT3	-0.120***	-0.094***	-0.188***	-0.207***	-0.199***	-0.200***
VIT4	-0.154***	-0.155***	-0.268***	-0.285***	-0.280***	-0.280***
VIT5	-0.197***	-0.180***	-0.239***	-0.281***	-0.240***	-0.240***

Notes: *, ** and *** denote significance at the 10%, 5% and 1% level, respectively. [a] The elements of X are restricted to the sex and age interval the respondent belongs to.

Table 4. Estimation results for $p_s(x)$. First step of the estimation of $PF2$.

Variable	Coefficient
Constant	-2.708***
Sex	-0.038
Age	0.052***
Age sq.	-0.013
Marstat1	-1.626***
Marstat2	-1.620***
Mid-Educ	0.870***
High-Educ	0.375***
Children [a]	1.074***
Income2	1.235***
Income3	0.952***
Income4	1.131***
Smoke1	0.461***
Smoke2	-0.884***
Smoke3	-1.629***
Own2	-0.804***
Own3	-1.950***
Psedo R^2	0.236
N	13509

Notes: *, ** and *** denote significance at the 10%, 5% and 1% level, respectively. [a] Number of children in the household.

Table 5. Parametric and Semiparametric consistent estimates.

	OLS	RE	$IPW2$
c	1.000	1.000	1.000
PF2	-0.011	-0.027***	-0.045*
PF3	-0.032***	-0.058***	-0.085***
PF4	-0.079***	-0.110***	-0.142***
PF5	-0.106***	-0.111***	-0.156***
PF6	-0.338***	-0.336***	-0.336***
RL2	-0.014	0.013	-0.078***
RL3	-0.032***	-0.034***	-0.133***
RL4	-0.082***	-0.087***	-0.192***
SF2	-0.040***	-0.087***	-0.066***
SF3	-0.062***	-0.088***	-0.071**
SF45	-0.205***	-0.227***	-0.253***
PAIN23	-0.015*	-0.027***	-0.076***
PAIN4	-0.206***	-0.169***	-0.241***
PAIN5	-0.194***	-0.224***	-0.327***
PAIN6	-0.335***	-0.361***	-0.403***
MH2	-0.064***	-0.023**	-0.063**
MH3	-0.082***	-0.049***	-0.184***
MH45	-0.158***	-0.136***	-0.204***
VIT2	-0.053***	-0.035***	-0.121***
VIT3	-0.114***	-0.081***	-0.207***
VIT45	-0.183***	-0.165***	-0.228***

Notes: *, ** and *** denote significance at the 10%, 5% and 1% level, respectively.

Table 6. Estimated tariffs.

	OLS	RE	$IPW1$	$IPW2$
Predictive description				
Mean	0.445	0.444	0.293	0.168
St. Dev.	0.215	0.214	0.223	0.252
Percentiles				
10	0.158	0.158	-0.002	-0.162
25	0.302	0.302	0.140	-0.006
50	0.456	0.461	0.297	0.171
75	0.601	0.604	0.449	0.345
90	0.716	0.717	0.582	0.495
Negative values (%)	2.63	2.53	10.19	25.76
Predictive ability				
MAE	0.174	0.176	0.238	0.299
$\mid$ pred. error $\mid < k$				
$k = 0.01$	4.48	3.86	3.76	1.83
$k = 0.05$	20.56	20.16	13.86	10.60
$k = 0.10$	39.82	38.10	26.91	18.39