Specification and Estimation of Rating Scale Models – with an Application to the Determinants of Life Satisfaction*

Raphael Studer [a] and Rainer Winkelmann [a,b]

[a] *Department of Economics, University of Zurich*
[b] *CESifo, Munich, and IZA, Bonn*

January 2011

Abstract: This article proposes a new class of rating scale models, which merges advantages and overcomes shortcomings of the traditional linear and ordered latent regression models. Both parametric and semi-parametric estimation is considered. The insights of an empirical application to satisfaction data are threefold. First, the methods are easily implementable in standard statistical software. Second, the non-linear model allows for flexible marginal effects, and predicted means respect the boundaries of the dependent variable. Third, average marginal effects are similar to ordinary least squares estimates.

Keywords: rating variables, non-linear least squares, quasi-maximum likelihood, semiparametric least squares, subjective well-being

JEL Codes: C21, I00.

*Address for correspondence: University of Zurich, Department of Economics, Zürichbergstr. 14, CH-8032 Zürich, Switzerland, ☎ +41 44 634 22 97 and +41 44 634 22 92, ✉ raphael.studer@econ.uzh.ch and rainer.winkelmann@econ.uzh.ch

1 Introduction

Research using rating data has burgeoned in recent years. A rating variable represents the extent to which a quality (e.g., health, risk aversion, approval with a policy or party) is present, or absent, in a study unit. The rating is often, but not necessarily, coded on an integer-valued scale. The smallest value (often a zero) represents the complete absence of the quality, whereas the largest value represents its complete presence.

The objective of this paper is to introduce a new approach for estimating the effect of explanatory variables on a rating by specifying and estimating a non-linear single index regression model. As a key advantage, the new approach introduces an explicit and flexible cardinalization, in contrast to so-called "ordered response" models, while avoiding the shortcomings of the linear regression model, namely constant marginal effects, and possible predictions outside the range of the dependent variable. The model is easy to implement; extensions to panel data and instrumental variable estimation are feasible.

While the arguments developed in this paper apply to any regression with a rating dependent variable, we concentrate on a specific application, namely that of the economic determinants of self-rated well-being. Many household (panel) surveys include a single-item 7-point or 11-point question on general life satisfaction, as well as on satisfaction with various life domains (health, family, work etc.). To estimate the relationship between such rating variables and their determinants, almost all of the existing literature has used either the linear regression model or ordered latent models. Applications in happiness research often report results from both type of models (e.g., Clark and Oswald, 1996; Ferrer-i-Carbonell and Frijters, 2002; Frey and Stutzer, 2005).

While Kristoffersen (2010) offered a theoretical discussion of the modeling options in general and cardinality respectively ordinality in particular, there remain some unresolved methodological issues when these two estimation methods are applied to rating data. These are presented in the next section of the paper. Section 3 discusses our theoretical framework and introduces a new class of rating scale models (RSM). The new methodology is illustrated in an application to the effect of time spent commuting to work on life satisfaction in Section 4. Section 5 concludes.

2 Motivation

Textbook treatments of rating variables recommend the ordered probit and the ordered logit models (e.g., Cameron and Trivedi, 2005). These can be derived from a latent linear model with standard normally or logistically distributed errors, respectively, where a partition of the real line is used to generate the observed discrete distribution of ordered outcomes. The main advantage of ordered latent models is the implied conformity to the scaling of the dependent rating variable. In particular, rating variables are bounded from below and from above (and thus limited dependent variables). Ordered latent models preclude nonsense predictions outside these boundaries. Moreover, latent models do not impose an equidistance between answer categories of the discrete scale.

However, although the name "ordered latent model" suggests otherwise, the estimation method has a cardinal foundation (van Praag and Ferrer-i-Carbonell, 2004). In particular, the ordinalization of the cardinal latent model hinges on an arbitrary assumption, such as that of a standard normally distributed error term in the latent model equation. The cardinality of ordered models also shows up when the model is interpreted. For instance, marginal probability effects are computed, or necessary changes in explanatory variables in order to attain a different response category are quantified. These are cardinal effects, which would not exist in a truly ordinal model. This raises the question, why a model with an implicit cardinalization should be preferred over a model which makes the cardinalization explicit.

In practice, these textbook models are therefore often abandoned in favor of the simpler linear regression model. It offers a convenient interpretation of estimated coefficients as marginal effects (although this interpretation is admittedly implausible since constant marginal effects stand in contradiction to the boundedness of the dependent variable). Indeed, researchers on life satisfaction seem to have little discomfort in reporting mean satisfaction levels (for instance by country, or group; see e.g., Stone et al., 2010 and Sacks et al., 2010). If one follows these practitioners and accords plausibility to reported (conditional) mean rating values, the only factors speaking against the use of the linear regression model is indeed that it imposes constant marginal effects and can predict rating scores outside the range of the rating scale.

The obvious remedy is to use a non-linear regression model that imposes bounds on the predicted values. If the attention is restricted to the class of single index models, the problem then becomes one of modeling the conditional expectation function $\mathrm{E}(y|x) = G(x'\beta)$, where G is a twice differentiable monotonic function such that $y^{min} \leq G(x'\beta) \leq y^{max}$ for all values of x and β. If $y \in \{0, 1\}$ (the rating takes only two values), this model has the form of standard binary response models. This similarity is deceiving, though, because it is only in the binary response model that probability function and conditional expectation function coincide. For more than two-valued rating scales, using the non-linear CEF approach is truly different from that of ordered response models.

We discuss a number of such rating scale models that differ in the assumptions they make regarding the G function. If a given parametric form is selected, estimation can proceed by non-linear least squares or quasi-maximum likelihood (see Papke and Wooldridge, 1996, for a closely related approach to fractional data). On the other hand, semiparametric least squares (Ichimura, 1993) can be used in order to estimate the RSM without making functional form assumptions.

3 Econometric Rating Scale Models

3.1 Specification

A rating variable y has domain $y \in [0, y^{max}]$ where we have normalized the lower bound $y^{min} = 0$ for convenience. Thus the value “0” represents the complete absence of the quality, whereas y^{max} represents its complete presence. Suppose that there are N observations, and that y_i, $i = 1, \ldots, N$, is the rating for observation unit i.

The RSM is specified in terms of a conditional expectation as a non-linear mapping of a single index:

$$\mathrm{E}(y_i|x_i) = G(x_i'\beta) \tag{1}$$

The vector x_i is of dimension $(k \times 1)$ and β is a conformable parameter vector. $G(.)$ specifies the non-linear relationship between the additive linear index $x_i'\beta$ and the rating variable y_i. $G(.)$ can take an arbitrary functional form. It is twice differentiable and accounts for the boundedness of the rating scale by satisfying $0 \leq G(.) \leq y^{max}$.

In a parametric rating scale model, a specific functional form is assumed for $G(.)$. Two possible specifications are a logit type model

$$G(x_i'\beta) = y^{max} \frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)} \tag{2}$$

and a probit type model

$$G(x_i'\beta) = y^{max} \Phi(x_i'\beta) \tag{3}$$

where $\Phi(.)$ is the cumulative density function of the standard normal distribution. These models imply that the transformed rating scale $z_i = y_i/y^{max}$ has a logit- or probit-type CEF, respectively.

Rating scale models (2) and (3) respect the boundaries of the dependent variable. They also imply non-constant marginal effects. In the logit RSM

$$\frac{\partial E(y_i|x_i)}{\partial x_{il}} = y^{max} \frac{\exp(x_i'\beta)}{(1 + \exp(x_i'\beta))^2} \beta_l \tag{4}$$

In the probit RSM

$$\frac{\partial E(y_i|x_i)}{\partial x_{il}} = y^{max}\phi(x_i'\beta)\beta_l \tag{5}$$

The parameters of the model can be estimated by non-linear least squares or by quasi-maximum likelihood, as explained in the next section. Alternatively, one can refrain from specifying the functional form of $G(.)$ and rather estimate it from data, together with the parameters β. This is a standard semiparametric estimation problem, and we propose to use the method of Ichimura (1993) for estimation.

3.2 Estimation

There are a number of possible ways for estimating the parameters of the parametric RSMs defined by (2) and (3). We start with one that actually is to be avoided, namely transforming the dependent variable in order to make the model linear in the parameters, in which case a linear regression model can be used. This method has been proposed, in the context of a scale bounded between 0 and 1, by Aitchison (1986). For a general rating scale, we can write

$$\log\left(\frac{y_i/y^{max}}{1-y_i/y^{max}}\right) = x_i'\beta + \eta_i \text{ where } \mathrm{E}(\eta_i|x_i) = 0 \tag{6}$$

At first glance, the re-specification is appealing because the logratio can take any real value. The unknown parameters β can be estimated consistently by ordinary least squares. However, there are two problems with this approach. First, y_i cannot take the extreme values of 0 or y^{max}. Second, it is impossible to recover the grandeurs of interest, especially the conditional expectation of the dependent variable y_i, since

$$\mathrm{E}(y_i|x_i) = \mathrm{E}\left(y^{max}\frac{\exp(x_i'\beta)\cdot\exp(\eta_i)}{1+\exp(x_i'\beta)\cdot\exp(\eta_i)}\middle| x_i\right) \neq y^{max}\frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}$$

Thus, the model is substantially different from (2) and as a consequence, it is hard to interpret β, the estimand in the linear regression model, other than saying that β measures the effect of x on the logratios. Transforming the rating scale to a real number is therefore not attractive. A method is needed that directly specifies the conditional expectation of the (untransformed) dependent rating variable.

3.2.1 Non-linear least squares

Non-linear least squares minimizes the sum of squared residuals of model (1). This is equivalent to choosing $\hat{\beta}$, which solves the following first order condition:

$$s(\beta; y, x) = \sum_{i=1}^{N} (y_i - G(x_i'\beta))g(x_i'\beta)x_i = 0 \text{ where } g(x_i'\beta) = \frac{\partial G(x_i'\beta)}{\partial x_i'\beta} \tag{7}$$

As a member of the family of extremum estimators, the NLS estimator is consistent, if the sample is independent and identically distributed and if $G(.)$ fulfills some regularity conditions (e.g., Hayashi, 2000).

Asymptotic theory enables the computation of standard errors. Default options in statistical software packages assume a spherical error variance. However, due to the boundedness the variance of rating variables is heteroscedastic. Intuitively, the closer the rating score moves to the boundaries the less dispersion is possible. The error term defined as $\varepsilon_i = y_i - E(y_i|x_i'\beta)$ inherits the heteroscedasticity of the rating variable. Therefore, a heteroscedastic consistent variance-covariance estimator for $\hat{\beta}$, as proposed by Huber (1967) and White (1980) is employed:

$$\text{AVar}(\hat{\beta}) = n^{-1}I^{-1}(\beta)J(\beta)I^{-1}(\beta)$$

where

$$I(\beta) = \text{E}[H(\beta; y, x)] = \text{E}\left[-g(x_i'\beta)^2 x_i x_i'\right]$$

and

$$J(\beta) = \text{Var}(s(\beta; y, x)) = E\left[(y_i - G(x_i'\beta))^2 g(x_i'\beta)^2 x_i x_i'\right]$$

Replacing the population moments reported above by their sample analogs leads to a consistent estimator of the heteroscedastic consistent variance-covariance matrix of $\hat{\beta}$.

3.2.2 Quasi-Maximum Likelihood Estimation

The parameters of the RSM (1) can be estimated consistently by embedding it in any member distribution of the linear exponential family and using maximum likelihood. Available distributions include, among others, the normal distribution, the Poisson distribution and the Bernoulli distribution (Gourieroux et al., 1984). The only requirement for consistency is that the CEF of the response scale model is correctly specified. This approach is referred to as quasi-maximum likelihood estimation (QML).

For example, if the normal distribution is used, QML is equivalent to non-linear least squares. If the Bernoulli distribution $B(1,p)$ is used as a basis for estimation, one needs to observe that $0 \leq p \leq 1$, whereas the CEF of the RSM is bounded from above at y^{max}. This problem can be solved by dividing both sides of equation (1) by y^{max}. The Bernoulli QML estimator is obtained by setting $p_i = G(x_i'\beta)/y^{max}$, and the first order conditions are:

$$\sum_{i=1}^{N} \frac{y_i - G(x_i'\beta)}{y^{max}} \frac{g(x_i'\beta)}{(1 - G(x_i'\beta)/y^{max})G(x_i'\beta)} x_i = 0 \text{ where } g(x_i'\beta) = \frac{\partial G(x_i'\beta)}{\partial x_i'\beta} \quad (8)$$

The QML framework does not impose any restrictions on the second or any higher moment of the dependent variable. In fact, the second moment is misspecified in the Bernoulli QML framework. Hence, the maximum likelihood variance estimation, which equals the inverse of the Hessian's expectation, has to be replaced by the robust sandwich variance estimator (Gourieroux et al. 1984).

3.2.3 Which estimator to choose

For a correctly specified CEF $G(.)$, both NLS and Bernoulli QML are consistent estimators. In small samples they may differ, since they use different weights w_i for the sample analog of the set of orthogonality conditions:

$$\sum_{i=1}^{N} (y_i - G(x_i'\beta) x_i w_i) = 0 \quad (9)$$

On one hand, NLS weighs the orthogonality conditions with the standard normal or the logistic probability density functions, respectively. On the other hand, the Bernoulli

QML estimator weighs observations with the probability density divided by the variance of a Bernoulli distributed variable. For the logistic model, these terms cancel and all elements of the score vector are weighted equally. The optimal weighting scheme depends on the true data generating process and its higher order moments. Since no such assumptions were made in our rating scale model, estimation with equal weights appears like a good starting point.

Both estimation methods are easy to implement in standard statistical software packages. In Stata (StataCorp., 2003), for example, the relevant model environment is given by the generalized linear model (glm) command. It allows to define distribution as well as link function. Choosing the normal distribution in conjunction with the logit link gives, for example, the non-linear least squares estimators of the logit-type RSM. Choosing the Bernoulli distribution instead results in the corresponding QML estimator. In either case, all ratings have first to be divided by the upper bound y^{max}, and robust standard errors need to be computed.

3.2.4 Semiparametric Least Squares

NLS and Bernoulli QML provide consistent parameter estimates for model (1) if the conditional expectation is correctly specified. An alternative to assuming a specific functional form for $G(.)$ is to estimate its conditional expectation. This approach remains consistent for β as long as the single index structure holds, regardless of the true $G(.)$. Different semiparametric estimators can be used. This paper employs the one that is the most simple to implement, namely semiparametric least squares (SLS) proposed by Ichimura (1993). SLS minimizes the sum of squared residuals of model (1).

$$\min_\beta \sum_{i=1}^{N} (y_i - \hat{\mathrm{E}}(G(x_i'\beta)|x_i'\beta))^2 \tag{10}$$

Iterative methods with an initial guess on $\hat{\beta}$ have to be applied in order to estimate both β and $\mathrm{E}(G(x_i'\beta)|x_i'\beta)$. For the latter, the local constant estimator proposed by Watson (1964) and Nadaraya (1965) is used. The local constant estimator depends on a kernel function and a bandwidth. If the choice of the kernel does not matter much, the bandwidth selection is important. The most appropriate way to choose the optimal bandwidth used in kernel regression is to apply cross validation (see e.g., Cameron and Trivedi, 2005).

Using an independent and identically distributed sample, a bandwidth sequence which converges towards 0 as N increases and with some technical requirements on the parameter space and the kernel, the properties of consistency and asymptotic normality can be established (Ichimura, 1993). Parameters are identified only up to location and scale. In other words any additive and multiplicative shifts in the regressors are incorporated by $G(.)$. Therefore, x_i does not include a constant term, and all remaining parameters are normalized with respect to a continuous variable's parameter. Marginal probability effects can be recovered for all explanatory variables.

Ichimura (1993) proposed to use the properties of asymptotic normality in order to compute standard errors of the parameters. However, the analytical derivation and implementation is very cumbersome. Therefore, bootstrapping standard errors is in general preferred.

The semiparametric RSM can be implemented conveniently using the non-parametric package in R (Hayfield and Racine, 2008). The program routine chooses the optimal bandwidth using cross validation and proposes as outputs estimates of the parameter vector, marginal effects and bootstrapped standard errors for those estimates.

4 Empirical Application to Happiness Data

Interest in measures of subjective well-being is increasing. Beginning with Easterlin's (1974) seminal publication on the relationship between economic growth and happiness, economists have been paying increasing interest to subjective well-being data. For a detailed review of the literature, see Frey and Stutzer (2002). More recently, policy makers have become interested as well. In particular, the prime ministers of France and Great Britain, Nicolas Sarkozy and David Cameron, promote new target functions, different from the Gross Domestic Product, which measure the population's well-being.[1] There will be more data available and more attention will be paid to subjective well-being measures in the future. Therefore, subjective well-being data is important area of investigation for an empirical application of the proposed estimation methods.

4.1 Replication of Stutzer and Frey (2008)

Stutzer and Frey (2008) analyzed in their paper "Stress that doesn't pay: The commuting paradox" the effect of commuting time and distance on satisfaction. One of the regression analysis of Stutzer and Frey will be replicated and re-estimated using the RSMs. In this particular regression, the rating dependent variable *"overall live satisfaction"*, measured on a discrete scale ranging from 0 to 10, was modeled. The authors were interested in the effect of commuting time, which was reported in minutes to work for one way by survey participants.[2] Data from eight waves of the German Socio Economic Panel (GSOEP) (1985, 1990, 1991, 1992, 1993, 1995, 1998, 2003) were used. The sample excluded people with irregular commuting patterns. Commuting times for people working from home were set to zero. The authors pooled all eight waves and estimated the model by OLS. Even though the number of observations in the samples used in this replication and in the paper of Stutzer and Frey differ by 707 observations, summary statistics, which are reported in Table 3 in the appendix, and estimates are virtually the same.[3]

[1] The France commission's homepage is: http://www.stiglitz-sen-fitoussi.fr/en/index.htm
A commentary about David Cameron's decision can be found at:
http://www.economist.com/node/17578888

[2] Table 1 Column 3 in Stutzer and Frey (2008).

[3] We thank the authors for providing us with assistance in the replication of the paper.

4.2 Application of Parametric RSMs

Table 1 shows estimates of the marginal effect of a one-minute increase in commuting time on general life satisfaction scores.[4] Standard errors of the average effects are in parentheses. Every column shows a different set of estimates. Column 1 replicates the ordinary least squares estimates found by Stutzer and Frey (2008). Column 2 to 5 report average marginal effects of the parametric RSMs. In Columns 2 and 3 the Bernoulli QML estimates are shown. NLS estimates are given in Columns 4 and 5, respectively.

OLS predicts the highest average reduction in happiness scores. A person commuting 60 minutes one-way is expected to have a 0.275 point lower satisfaction score than a comparable person, who does not commute (Stutzer and Frey 2008). The average marginal probability effect estimated by the logit-type Bernoulli QML estimates the lowest difference between these two individuals, namely 0.269 point. Overall, average marginal effects vary only slightly among regression models. The different weighting schemes of NLS and Bernoulli QML discussed earlier appear not to matter much in this empirical application.

Table 1: Effect of Commuting Time on Satisfaction - Parametric RSMs

	(1) OLS	(2) QML-Logit	(3) QML-Probit	(4) NLS-Logit	(5) NLS-Probit
Commuting Time	-0.00459	-0.00449	-0.00453	-0.00451	-0.00453
	(0.00046)	(0.00047)	(0.00047)	(0.00047)	0.00048
Individual characteristics	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes
Robust Standard Errors	No	Yes	Yes	Yes	Yes
Observations	39747	39747	39747	39747	39747

· Standard errors reported in parentheses.
· Individual characteristics: sex, age, age^2, 6 categories of years of education, 2 variables for the relationship to the household head, 9 variables for marital status, 4 variables for number of children in the hh, square root of the number of hh members, East German, foreigners with EU nationality, foreigners without EU nationality, self-employment
· Effects in columns (2), (3) and (4) represent average marginal probability effects.
· Column (1) corresponds to Column (3) of Table1 in Stutzer and Frey (2008).

The estimates of the effect of commuting time are found to be statistically significant at common confidence levels in all regressions. In contrast to the OLS model, where

[4] The estimated parameter vectors for all variables are reported in Table 4 in the appendix.

homoscedasticity of the error variance is assumed, the robust sandwich variance estimator is used for computation of the standard errors for regression model 2 to 5.

In the light of the wide utilization and acceptance of OLS in the rating variable literature, it is appealing to find the parametric RSMs estimating very similar average marginal effects. However, the non-linear specification of the conditional expectation differs from OLS in two points. First, RSMs' mean predictions are bounded. Second, RSMs' marginal effects are not constant. Two graphical illustrations make these differences apparent.

Figure 1: Predicted Satisfaction for Sample Members

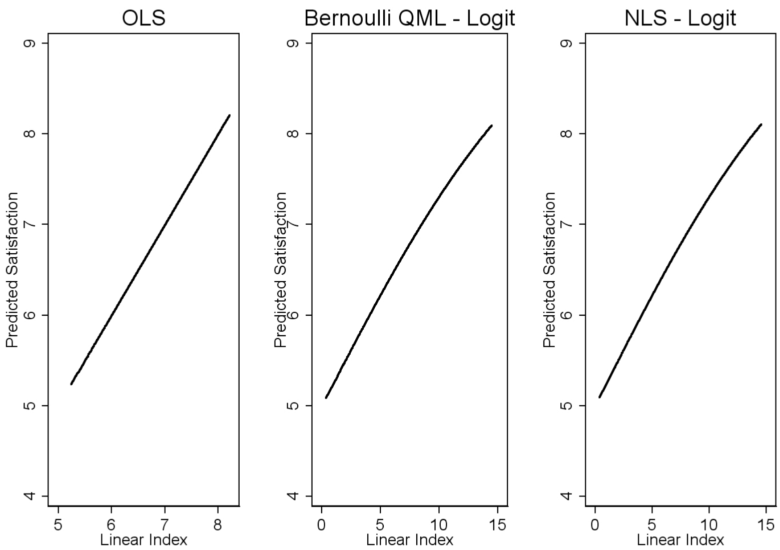

Figure 1 plots the estimated conditional expectation, i.e. predicted satisfaction scores for all sample members. The three graphs report the predictions obtained by OLS, Bernoulli QML and NLS (from left to right). For the later two models the function $G(.)$ is specified using the logistic cumulative distribution function. For OLS the mean predictions equal the linear index. In this application, OLS predictions for sample members do not hurt the bounds of zero and ten. But, this need not hold in general. Moreover, it is always possible to make OLS out-of-sample predictions that violate the bounds.

The Bernoulli QML and the NLS predictions are very similar. Both predict a locally concave relationship between linear indexes and life satisfaction scores. Remembering the shape of the cumulative distribution function of the logistic distribution, it follows that the sample predictions are centered around the upper flection. In this application the lower flection would only play a role for out-of-sample predictions with a negative linear index. In other words, the parametric assumption ensures the boundedness of (out-of-sample) predictions.

Figure 2: Marginal Effect of Commuting Time on Life Satisfaction

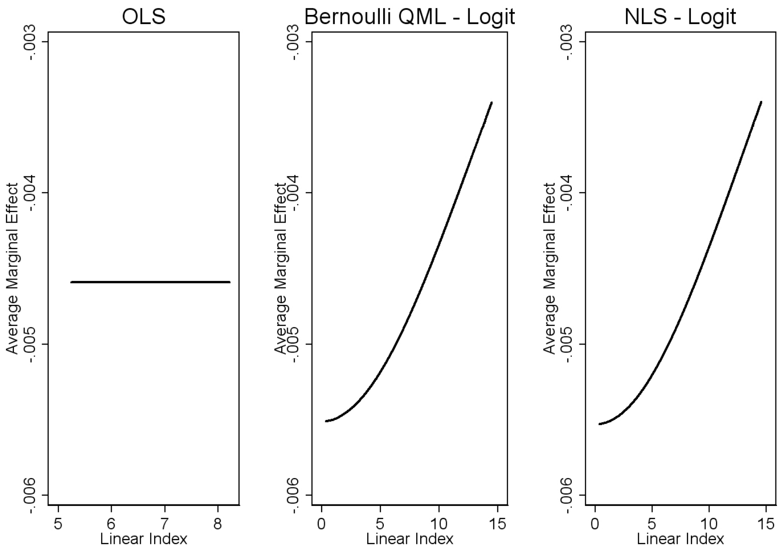

Figure 2 plots the estimated marginal effects. For OLS the marginal effect is constant among all individuals and represents an average effect. However, the boundedness of rating variables preclude constant marginal effects. The individual specific marginal effects are shown for the logit type Bernoulli QML and the NLS models in the second and third graph from the left of Figure 2. The marginal effects depend on the local shape of the cumulative distribution function. The shape of the predicted function $G(x_i'\beta)$ therefore suggests, that with an increasing linear index commuting time affects individuals less. This is plausible. Very satisfied people, who feel themselves fully blessed with luck, weigh a one-minute increase in commuting time less than people, who perceive their life as unsatisfactory. The non-constant marginal effects provide therefore useful information, for instance for policy makers, who want to focus only on certain subgroups of observations where the effects of an intervention can be expected to be particularly large.

4.3 Application of the Semiparametric RSM

We choose to implement the SLS estimator using a plug-in bandwidth for two reasons. First, the huge sample and the big number of parameters make cross validation computationally intensive. Second, cross validation resulted in a too small bandwidth, i.e in an under smoothed estimate of the conditional expectation of the functional form. On one hand, this might be due to the lack of independence among observations, as the sample is pooled over time periods. On the other hand, the parameter of normalization (commuting time) is relatively small. Hence, the range of linear indexes is wide. Different essays identified a bandwidth of 10 to provide appropriate smoothing.[5]

Table 2: Effect of Commuting Time on Satisfaction - Sempirametric RSM

	(1) OLS	(2) SLS
Commuting Time	-0.00459	-0.00479
Individual characteristics	Yes	Yes
Time fixed effects	Yes	Yes
Observations	39747	39747

· Individual characteristics: sex, age, age^2, 6 categories of years of education, 2 variables for the relationship to the household head, 9 variables for marital status, 4 variables for number of children in the hh, square root of the number of hh members, East German, foreigners with EU nationality, foreigners without EU nationality, self-employment
· Effects in columns (2) are marginal probability effects evaluated at the mean characteristics.
· Column (1) corresponds to Column (3) of Table 1 in Stutzer and Frey (2008).

Table 2 reports ordinary least squares and semiparametric least squares estimates of the marginal effect of commuting time on life satisfaction.[6] The model is estimated with

[5]A Matlab (The MathWorks Inc., 2008) code implementing the SLS estimator can be found at http://www.sts.uzh.ch/members/Studer/slsmatlabcode.pdf.

[6]SLS marginal effects are computed at the mean characteristics and not for each individual. This is preferable as for individuals with a linear index closed to the bounds, marginal effects might be severely biased when the local constant estimator is used for estimating the conditional expectation of $G(.)$.

the same set of explanatory variables as those listed in Table 1. SLS and OLS estimates are very similar. A 30 minutes increase in one-way commuting time lowers life satisfaction approximately by 0.144 respectively 0.138 point ceteris paribus. Table 5 in the appendix shows that this finding generalizes to other variables. The marginal effects of all regressors are similar between OLS and SLS.

Figure 3: Predicted Satisfaction for Sample Members

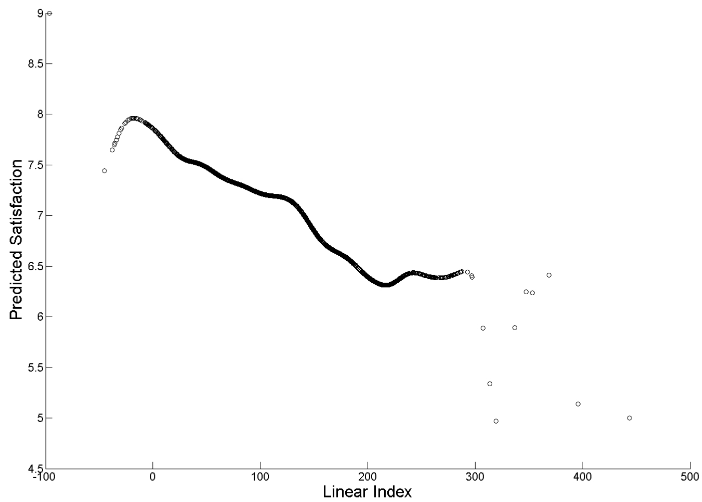

Figure 3 plots the SLS mean predictions. Two peculiarities deserve to be mentioned. First, the estimated conditional expectation is decreasing in the linear index. This stands in contradiction to the plotted predictions in Figure 2 and is due to the negative sign of the normalization coefficient (the effect of commuting time). Hence, relative coefficients take the opposite sign of the coefficients estimated in OLS or the parametric RSMs. Second, as boundaries of the support data ($x_i'\beta$) are approached predictions are widespread. This is an artifact of the local constant estimator, which suffers from an edge bias. In particular, kernel estimates at the boundaries of $x_i'\beta$ will be based on one sided observations only. $\mathrm{E}(G(x_i'\beta)|x_i'\beta)$ estimated for an observation at the upper bound of the training data will be underestimated, as $G(.)$ is decreasing in the linear index. The local linear regression estimator would overcome this source of bias. However, it is computationally more demanding and can be instable (Racine, 2008). Moreover, consistency of the parameter estimates can still be established when using the local constant estimator for $\mathrm{E}(G(x_i'\beta)|x_i'\beta)$ (Ichimura, 1993).

Several remarks apply when comparing SLS to OLS and the fully parametric models. First, the SLS estimates of marginal effects are very closed to OLS (and therefore to the average marginal effects of the parametric RSMs). Second, SLS respects the boundaries of the rating dependent variable as data points are used to estimate the conditional expectation of $G(.)$. Finally, researchers should be aware that SLS does not allow for out-of-sample predictions.

5 Conclusion

This paper focused on econometric models for rating data. We established that existing models, such as ordered latent models or the linear regression model, have a number of shortcomings. A new general framework for a cardinal rating scale model addresses these issues. Depending on the specific assumptions, model parameters can be estimated by non-linear least squares, by quasi-maximum likelihood or by semi-parametric least squares.

Predicted means of the new model automatically satisfy the logical constraints provided by the upper and lower bounds of the rating scale. It works equally well for discrete ratings, as it does for continuous ones. Continuous, or near continuous, ratings scales are empirically relevant. For example, the Standard & Poors rating of investment grades distinguishes 25 values. Also, subjective ratings in a survey can be performed by a visual mark on a ruler, a method that has been employed occasionally in psychometrics, and is likely to become more widespread in the future. In these cases, ordered latent models clearly are impractical, and the proposed RSM is a superior alternative to the linear regression model that ignores the boundary condition of such scales.

In an empirical application to discrete life satisfaction scores, we showed that the methods are easy to implement. It turned out that the average marginal effects of the nonlinear

RSMs were similar to ordinary least squares estimates. However, substantial differences in predicted individual specific marginal effects could be found for observations in the tails of the distribution of predicted satisfaction scores.

References

Aitchison, J., 1986, "The Statistical Analysis of Compositional Data", *Chapman and Hall*, New York

Cameron, C.A. and P.K. Trivedi, 2005, "Microeconometrics, Methods and Applications", *Cambridge University Press*, New York

Clark, A.E. and A.J. Oswald, 1996, "Satisfaction and Comparison Income', *Journal of Public Economics*, Vol. 61, Issue 3, 95-144

Easterlin, R. A., 1974, "Does Economic Growth Improve the Human Lot?", *David, P.A. and M. W. Reder, eds., Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*, New York: Academic Press, Inc., 98-125

Ferrer-i-Carbonell, A. and P. Frijters, 2004, "How important is Methodology for the Estimates of the Determinants of Happiness?", *The Economic Journal*, Vol. 114, Issue 497, 641-659

Frey, B.S. and A. Stutzer, 2002, "Happiness and Economics", *Princeton University Press*, Princeton

Frey, B.S. and A. Stutzer, 2005, "Happiness Research: State and Prospects", *Review of Social Economy*, Vol. 62, 207-228

Gourieroux C., A. Monfort and A. Trognon, 1984, "Pseudo Maximum Likelihood Methods: Theory", *Econometrica*, Vol. 52 No.3, 681-700

Hayashi, F. 2000, "Econometrics", *Princeton University Press*, New Jersey

Hayfield, T. and J.S. Racine, 2008, " Nonparametric Econometrics: The np Package", Journal of Statistical Software 27(5). URL http://www.jstatsoft.org/v27/i05/

Huber, P. J., 1967, "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions", *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. I, pp. 22133.

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

Kristoffersen, I., 2010, "The Metrics of Subjective Wellbeing: Cardinality, Neutrality and Additivity", *The Economic Record*, Vol. 86, No. 272, 98-123

Nadaraya, E.A., 1965, "On nonparametric estimates of density functions and regression curves", *Theory of Applied Probability*, 10, 186-190

Papke, L.E. and J.M. Wooldridge, 1996, "Econometric methods for fractional response variables with an application to 401(k) plan participation rates", *Journal of Applied Econometrics*, Vol. 11, Issue 6, 619-632

Racine, J.S, 2008, "Nonparametric Econometrics: A Primer", *Foundations and Trends in Econometrics*, 3(1), 1-88

Sacks, D.W., B. Stevenson and J. Wolfers, 2010, "Subjective Well-being, Income, Economic Development and Growth", *NBER Working Paper Series*, No. 16441

StataCorp, 2003, *Stata Statistical Software: Release 10*, College Station, Texas

Stone, A.A., J.E., Schwartz, J.E., Brodericka and A. Deaton, 2010, "A Snapshot of the Age Distribution of Psychological Well-being in the United States", PNAS Paper

Stutzer, A. and B.S. Frey, 2008, "Stress that Doesn't Pay: The Commuting Paradox", *Scandinavian Journal of Economics*, Vol. 110, Issue 2, 339-366

The MathWorks Inc., 2008, *Matlab, The Language of Technical Computing: Version 7.6.0.324 (R2008a)*

van Praag, B.M.S.and A. Ferrer-i-carbonell, 2004, "Happiness Quantified: A Satisfaction Calculus Approach", *Oxford University Press , New York*

Watson, G.S., 1964, "Smooth regression analysis", *Sankhya*, 26(15), 359-372

White, H., 1980, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica*, 48, 81738

Appendix

Table 3: Replication of Summary Statistics

	Mean	Std.Dev.	Min	Max
Age	38.84	11.60	14	86
Years of Education	11.41	3.08	7	18
Children in hh	0.75	0.98	0	9
Persons in hh	3.12	1.35	1	14
Female	0.44	0.49	0	1
Child of hh-head	0.13	0.33	0	1
No hh-head	0.01	0.1	0	1
Single-wp	0.25	0.43	0	1
Married	0.65	0.48	0	1
Seperated-wp	0.02	0.13	0	1
Seperated-np	0.002	0.04	0	1
Divorced-wp	0.06	0.24	0	1
Divorced-np	0.004	0.07	0	1
Widowed-wp	0.01	0.12	0	1
Widowed-np	0.001	0.04	0	1
Spouse abroad	0.002	0.04	0	1
Selfemployed	0.15	0.36	0	1
East-German	0.2	0.40	0	1
EU-citizen	0.07	0.26	0	1
Foreigner Non-EU	0.1	0.3	0	1

· N=39747
· Abbreviations: hh: household, np: no partner, wp: with partner
· This table replicates the summary statistics provided in the appendix of Stutzer and Frey (2008).

Table 4: Raw Regression Output - Parametric RSMs

	OLS	QML-Logit	QML-Probit	NLS-Logit	NLS-Probit
Commuting Time	-4.59E-03	-2.20E-03	-1.33E-03	-2.21E-03	-1.33E-03
	(4.57E-04)	(2.31E-04)	(1.40E-04)	(2.31E-04)	(1.40E-04)
Age	-4.90E-02	-2.44E-02	-1.45E-02	-2.50E-02	-1.49E-02
	(5.76E-03)	(3.00E-03)	(1.79E-03)	(3.02E-03)	(1.80E-03)
Age^2	5.06E-04	2.53E-04	1.50E-04	2.60E-04	1.55E-04
	(6.83E-05)	(3.56E-05)	(2.13E-05)	(3.57E-05)	(2.13E-05)
Female	-1.77E-02	-8.53E-03	-5.19E-03	-8.15E-03	-4.96E-03
	(1.71E-02)	(8.40E-03)	(5.04E-03)	(8.44E-03)	(5.07E-03)
Education = 7y.	-4.22E-02	-2.10E-02	-1.29E-02	-1.84E-02	-1.14E-02
	(4.56E-02)	(2.52E-02)	(1.51E-02)	(2.53E-02)	(1.52E-02)
Education = 10y.	1.55E-01	7.73E-02	4.63E-02	7.68E-02	4.60E-02
	(2.62E-02)	(1.37E-02)	(8.21E-03)	(1.37E-02)	(8.24E-03)
Education = 12y.	1.94E-01	9.62E-02	5.77E-02	9.55E-02	5.73E-02
	(3.27E-02)	(1.63E-02)	(9.81E-03)	(1.64E-02)	(9.83E-03)
Education = 14y.	2.47E-01	1.22E-01	7.30E-02	1.24E-01	7.43E-02
	(3.69E-02)	(1.80E-02)	(1.08E-02)	(1.81E-02)	(1.09E-02)
Education = 18y.	3.94E-01	1.95E-01	1.17E-01	1.95E-01	1.17E-01
	(3.85E-02)	(1.87E-02)	(1.12E-02)	(1.88E-02)	(1.13E-02)
Child of hh-head	8.55E-02	4.30E-02	2.65E-02	3.67E-02	2.27E-02
	(4.25E-02)	(2.11E-02)	(1.27E-02)	(2.12E-02)	(1.27E-02)
No hh-head	-1.68E-01	-8.12E-02	-4.95E-02	-7.70E-02	-4.70E-02
	(8.42E-02)	(4.25E-02)	(2.56E-02)	(4.27E-02)	(2.58E-02)
Single-wp	9.26E-01	4.15E-01	2.54E-01	4.15E-01	2.54E-01
	(2.07E-01)	(1.09E-01)	(6.76E-02)	(1.10E-01)	(6.81E-02)
Married	1.14	5.18E-01	3.16E-01	5.14E-01	3.14E-01
	(2.06E-01)	(1.09E-01)	(6.75E-02)	(1.10E-01)	(6.79E-02)
Separated-wp	5.04E-01	2.24E-01	1.37E-01	2.24E-01	1.37E-01
	(2.16E-01)	(1.14E-01)	(7.06E-02)	(1.15E-01)	(7.10E-02)
Separated-np	-5.08E-01	-2.21E-01	-1.37E-01	-2.17E-01	-1.34E-01
	(2.20E-01)	(1.28E-01)	(7.87E-02)	(1.29E-01)	(7.95E-02)
Divorced-wp	7.69E-01	3.45E-01	2.11E-01	3.42E-01	2.10E-01
	(2.09E-01)	(1.10E-01)	(6.81E-02)	(1.11E-01)	(6.85E-02)
Divorced-np	-2.33E-03	-6.55E-03	-3.53E-03	-4.53E-03	-2.30E-03
	(1.30E-01)	(6.95E-02)	(4.20E-02)	(6.99E-02)	(4.22E-02)
Widow-wp	8.09E-01	3.64E-01	2.22E-01	3.61E-01	2.21E-01
	(2.17E-01)	(1.14E-01)	(7.06E-02)	(1.15E-01)	(7.10E-02)
Widow-np	-4.53E-01	-2.03E-01	-1.24E-01	-2.01E-01	-1.23E-01
	(2.38E-01)	(1.51E-01)	(9.28E-02)	(1.50E-01)	(9.25E-02)
Child-hh=1	-6.36E-02	-3.12E-02	-1.88E-02	-3.06E-02	-1.85E-02
	(2.52E-02)	(1.24E-02)	(7.43E-03)	(1.23E-02)	(7.39E-03)
Child-hh=2	-7.67E-02	-3.80E-02	-2.31E-02	-3.55E-02	-2.16E-02
	(3.32E-02)	(1.60E-02)	(9.63E-03)	(1.60E-02)	(9.62E-03)
Child-hh>3	-2.22E-01	-1.09E-01	-6.54E-02	-1.10E-01	-6.60E-02
	(5.09E-02)	(2.53E-02)	(1.52E-02)	(2.53E-02)	(1.52E-02)
Squareroot Persons in hh	1.11E-01	5.45E-02	3.26E-02	5.57E-02	3.33E-02
	(4.00E-02)	(2.00E-02)	(1.20E-02)	(2.02E-02)	(1.21E-02)
Selfemployed	-9.04E-02	-4.40E-02	-2.65E-02	-4.36E-02	-2.63E-02
	(2.33E-02)	(1.14E-02)	(6.87E-03)	(1.14E-02)	(6.88E-03)
East-German	-7.13E-01	-3.36E-01	-2.04E-01	-3.36E-01	-2.03E-01
	(2.24E-02)	(1.01E-02)	(6.14E-03)	(1.02E-02)	(6.17E-03)
EU-citizen	1.26E-01	6.47E-02	3.88E-02	6.27E-02	3.76E-02
	(3.53E-02)	(1.89E-02)	(1.12E-02)	(1.90E-02)	(1.13E-02)
Foreigner Non-EU	-1.19E-01	-5.88E-02	-3.53E-02	-5.86E-02	-3.52E-02
	(3.03E-02)	(1.62E-02)	(9.67E-03)	(1.63E-02)	(9.75E-03)
First interview	2.54E-01	1.31E-01	7.77E-02	1.31E-01	7.81E-02
	(3.66E-02)	(1.89E-02)	(1.12E-02)	(1.89E-02)	(1.12E-02)
Year 90	8.37E-02	4.34E-02	2.56E-02	4.51E-02	2.66E-02
	(2.83E-02)	(1.46E-02)	(8.70E-03)	(1.47E-02)	(8.74E-03)
Year 92	-4.12E-01	-1.81E-01	-1.11E-01	-1.79E-01	-1.10E-01
	(6.70E-02)	(2.87E-02)	(1.77E-02)	(2.88E-02)	(1.77E-02)
Year 95	-5.99E-02	-2.95E-02	-1.81E-02	-2.63E-02	-1.62E-02
	(2.51E-02)	(1.24E-02)	(7.47E-03)	(1.25E-02)	(7.49E-03)
Year 98	-1.14E-02	-6.44E-03	-4.01E-03	-4.35E-03	-2.79E-03
	(2.58E-02)	(1.26E-02)	(7.54E-03)	(1.26E-02)	(7.56E-03)
Year 03	-7.20E-02	-3.55E-02	-2.17E-02	-3.20E-02	-1.96E-02
	(2.33E-02)	(1.15E-02)	(6.91E-03)	(1.16E-02)	(6.94E-03)
Constant	7.11	9.47E-01	5.78E-01	9.59E-01	5.85E-01
	(2.41E-01)	(1.26E-01)	(7.75E-02)	(1.27E-01)	(7.79E-02)

· Standard errors reported in parentheses.
· Estimated coefficients correspond to the parameter vector β in models (**??**).
· First line of column (1) corresponds to column (3) of table 1 in Stutzer and Frey (2008).
· N=39747

Table 5: Marginal Effects - Semiparametric RSM

	OLS	SLS
Commuting Time	-4.59E-03	-4.79E-03
Age	-4.90E-02	-5.08E-02
Age^2	5.06E-04	5.31E-04
Female	-1.77E-02	-1.85E-02
Education = 7y.	-4.22E-02	-4.42E-02
Education = 10y.	1.55E-01	1.62E-01
Education = 12y.	1.94E-01	2.04E-01
Education = 14y.	2.47E-01	2.61E-01
Education = 18y.	3.94E-01	4.06E-01
Child of hh-head	8.55E-02	8.93E-02
No hh-head	-1.68E-01	-1.77E-01
Single-wp	9.26E-01	9.79E-01
Married	1.14E+00	1.18E+00
Separated-wp	5.04E-01	5.26E-01
Separated-np	-5.08E-01	-5.36E-01
Divorced-wp	7.69E-01	7.85E-01
Divorced-np	-2.33E-03	-2.44E-03
Widow-wp	-4.53E-01	-4.74E-01
Widow-np	8.09E-01	8.57E-01
Child-hh=1	-6.36E-02	-6.63E-02
Child-hh=2	-7.67E-02	-8.08E-02
Child-hh>3	-2.22E-01	-2.35E-01
Squareroot Persons in hh	1.11E-01	1.13E-01
Selfemployed	-9.04E-02	-9.54E-02
East-German	-7.13E-01	-7.43E-01
EU-citizen	1.26E-01	1.33E-01
Foreigner Non-EU	-1.19E-01	-1.25E-01
First interview	2.54E-01	2.66E-01
Year 90	8.37E-02	8.29E-02
Year 92	-4.12E-01	-4.29E-01
Year 95	-5.99E-02	-6.26E-02
Year 98	-1.14E-02	-1.21E-02
Year 03	-7.20E-02	-7.49E-02
Constant	7.11	

· Reported coefficients correspond to marginal effects. Marginal effects in Column 2 are evaluated at the mean characteristics.
· First line of column (1) corresponds to column (3) of table 1 in Stutzer and Frey (2008).
· The life satisfaction score is modeled as dependent variable.
· N=39747