Econometric Analysis of Ratings – with an Application to Health and Wellbeing[a]

Raphael Studer[b] and Rainer Winkelmann[c]

SUMMARY

We propose a new non-linear regression model for rating dependent variables. The rating scale model accounts for the upper and lower bounds of ratings. Parametric and semi-parametric estimation is discussed. An application investigates the relationship between stated health satisfaction and physical and mental health scores derived from self-reports of various health impairments, using data from the German Socio-Economic Panel. We compare our new approach to modeling ratings with ordinary least squares (OLS). In one specification, OLS average effects exceed that from our rating scale model by up to 50 percent. Also, OLS in-sample mean predictions violate the upper bound of the dependent variable in a number of cases.

1. Introduction

Empirical economic 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 (commonly a zero) represents the complete absence of the quality, whereas the largest value represents its complete presence. The example considered in this paper is satisfaction with health, measured on a 0 to 10 scale, where 0 means "completely dissatisfied" and 10 means "completely satisfied". Other examples are self-assessed health (1 = "very bad", 5 = "very good") or subjective risk aversion.

So far, regression analyses for rating dependent variables have followed one of two approaches: Either, the rating is treated as an ordinal variable, indicating the use of ordered probit or ordered logit models (see, e. g., CLARK and OSWALD, 1996). Or else, the rating is treated as cardinal and categories as equidistant, and simple linear regression models are employed (see, e. g., FERRER-I-CARBONELL and FRIJTERS, 2004).

Both approaches have shortcomings. For example, the interpretation of the ordered probit model becomes cumbersome as the number of categories increases. Relatedly, it does not directly provide the key object of interest in most applications, the effect of a regressor on the expected outcome. The linear regression model, on the other hand, ignores that the dependent variable is bounded, and that marginal effects cannot be constant. It can lead to predictions outside of the admissable range.

In this paper, we therefore advocate a third way, an alternative approach for estimating the effects of explanatory variables on a rating, based on a class of non-linear single index regression models and building on work by PAPKE and WOOLDRIDGE (1996) on fractional responses. As in linear regression, we focus on the conditional expectation. However, in order to maintain model consistency, we require that the conditional expectation respects the upper and lower bounds implied by the rating scale. As a consequence, model predictions outside the range of the dependent variable are impossible and marginal effects are not constant. The model is easy to implement. It works for any number of categories, and 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 satisfaction with health. Many household (panel) surveys contain a single-item 7-point or 11-point question on general life satisfaction, as well as satisfaction in a number of specific domains, including health.

Not surprisingly, therefore, there is already a sizeable literature on the determinants of health satisfaction (e. g., Frijters et al., 2005, Romeu Gordo, 2006, Jones and Schurer, 2011). We are the first to present estimation results from a rating scale model in this context. Substantively, we find that health satisfaction is strongly related to physical and mental health scores derived from a list of health impairment questions. These indicators can explain more than half of the variation in health satisfaction. Moreover, in one of our specifications, average effects obtained from OLS and from a rating scale model differ by up to 50 percent. The next section provides a formal exposition of rating scale models. Section 3 reports on the application to the relationship between satisfaction with health and physical and mental health scores. Section 4 concludes.

2. Econometric Models for Rating Scale Variables

2.1 Specification

A rating scale variable y lies in an interval $[0, y^{max}]$ with probability one. If the initial scale of y does not start at zero, subtracting y^{min} first ensures that the transformed scale does. Rating variables thus fall within the class of limited dependent variables (Maddala, 1983). Unless an analog scale is used, such as in Studer and Winkelmann (2014), ratings are also discrete dependent variables, but this is unimportant for the current argument, as we focus on the (continuous) mean response only.

Consider the following general non-linear specification for the average rating of observation unit i:

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

Here, x_i is a $(k \times 1)$ vector of explanatory variables, and β is a conformable parameter vector. To give meaning to (1), we need to find a function G such that $0 \leq G(.) \leq y^{max}$. The most straightforward way of doing this is by means of scaled versions of the probit and logit models, such that

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

or

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

respectively. We also explore an alternative approach, where G is left unspecified and estimated jointly with β using semiparametric non-linear least squares. Note that (1) is equivalent to a non-linear regression model

$$y_i = G(x_i'\beta) + \varepsilon_i \qquad (4)$$

where $\mathrm{E}(\varepsilon_i | x_i) = 0$ and ε_i is necessarily heteroskedastic, since

$$-G(x_i'\beta) \leq \varepsilon_i \leq y^{max} - G(x_i'\beta).$$

In particular, $\mathrm{Var}(\varepsilon_i | x_i) \to 0$ as $G(x_i'\beta) \to 0$ or $G(x_i'\beta) \to y^{max}$.

2.2 Estimation

The model parameters of the rating scale model (RSM) can be estimated by non-linear least squares. However, due to the inherent heteroskedasticity, more efficient estimators are available. In particular, in the spirit of PAPKE and WOOLDRIDGE (1996) for fractional responses, we propose to use quasi-maximum likelihood estimation (GOURIEROUX et al., 1984) based on the Bernoulli distribution. For implementation, note that the mean of a Bernoulli variable is bounded between 0 and 1, whereas a rating scale variable is bounded from above by y^{max}. For an appropriate modification, divide 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 using as dependent variable the "fractional response" y_i / y^{max}. The Bernoulli quasi likelihood function for a sample of n independent observations is then

$$L(\beta; x, y) = \prod_{i=1}^{n} \left(\frac{G(x_i'\beta)}{y^{max}} \right)^{y_i / y^{max}} \left(\frac{y^{max} - G(x_i'\beta)}{y^{max}} \right)^{1 - y_i / y^{max}}, \qquad (5)$$

with first order conditions:

$$\sum_{i=1}^{N} \frac{y_i - G(x_i'\beta)}{y^{max}} \frac{\dfrac{g(x_i'\beta)}{y^{max}}}{\left(1 - \dfrac{G(x_i'\beta)}{y^{max}}\right) \dfrac{G(x_i'\beta)}{y^{max}}} x_i = \sum_{i=1}^{N} \frac{[y_i - G(x_i'\beta)]g(x_i'\beta)x_i}{(y^{max} - G(x_i'\beta))G(x_i'\beta)} = 0 \qquad (6)$$

$$\text{where } g(x_i'\beta) = \frac{\partial G(x_i'\beta)}{\partial x_i'\beta}.$$

Under the assumption of a correctly specified conditional expectation function, the estimator $\hat{\beta}$ that solves these first order conditions is consistent for β and asymptotically normally distributed. Since the dependent variable is not really binary but rather a rating, the model is misspecified, and the robust sandwich variance estimator (White 1980, Gourieroux et al. 1984) needs to be used:

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

where

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n}\frac{g(x_i'\hat{\beta})^2 x_i x_i'}{G(x_i'\hat{\beta})[y^{max} - G(x_i'\hat{\beta})]} \tag{7}$$

and

$$\hat{J} = \frac{1}{n}\sum_{i=1}^{n}\frac{[y_i - G(x_i'\hat{\beta})]^2 g(x_i'\hat{\beta})^2 x_i x_{i'}}{\{G(x_i'\hat{\beta})[y^{max} - G(x_i'\hat{\beta})]\}^2} \tag{8}$$

Estimation can be performed with standard software packages for binary response models using the following steps: first, divide the rating variable by y^{max}. The transformed "fractional" variable takes values between 0 and 1. Second, estimate a probit or logit model with y_i/y^{max} as dependent variable. Compute robust standard errors. Third, compute average partial effects (see Section 2.3 below) from the binary response model and multiply them (and their standard errors) by y^{max}.

For a semiparametric RSM, G and β can be estimated simultaneously by semiparametric least squares (SLS, Ichimura, 1993). SLS minimizes the sum of squared residuals of model (1). Iterative methods with an initial guess on $\hat{\beta}$ are applied. Parameters of the linear index are identified up to location and scale. Therefore, x_i does not include a constant term, and all remaining parameters are normalized with respect to the parameter of a continuous regressor. Marginal effects can be recovered for all explanatory variables and standard errors are obtained by bootstrap.

2.3 Interpretation and Relation to Ordered Probit

Specifications (2) and (3) guarantee that predicted averages always fall within the admissable $[0, y^{max}]$ range. Marginal effects decrease as predictions approach the bound, and become zero in the limit (for $|x_i'\beta| \to \infty$). Specifically, for the logit RSM

$$\frac{\partial E(y_i \mid x_i)}{\partial x_{il}} = y^{max} \frac{\exp(x_i'\beta)}{(1 + \exp(x_i'\beta))^2} \beta_l \tag{9}$$

whereas for the probit RSM

$$\frac{\partial E(y_i \mid x_i)}{\partial x_{il}} = y^{max} \phi(x_i'\beta)\beta_l \tag{10}$$

where $\phi(.)$ denotes the density function of the standard normal distribution. In the application below, average marginal effects are reported, i.e., we compute marginal effects for each observation $i = 1, \ldots, n$ and then take the arithmetic mean.

Suppose that the true data generating process is an ordered response model instead. In this case the conditional expectation of the latent dependent variable is linear. But the latent variable is not what interests us. The observed responses are obtained via a partitioning of the real line using y^{max} thresholds $\kappa_1 < \kappa_2 < \ldots < \kappa_{y^{max}}$ (e.g., CAMERON and TRIVEDI, 2005). The conditional expectation of the observed response depends on the category labels (and is not invariant to reparameterization). If we let as before $y = 0, 1, \ldots, y^{max}$, then it is easy to show that

$$E(y_i \mid x_i) = G(\kappa_1 - x_i'\beta) + \ldots + G(\kappa_{y^{max}} - x_i'\beta) \tag{11}$$

In principle, parameters of this model could be estimates by NLS or QML as well, although estimation by maximum likelihood is of course more efficient.

The above specification fulfills the basic requirement of a RSM, since $0 \le E(y|x) \le y^{max}$ for sure. However, the RSM (1) is more parsimonious and also easier to interpret. Hence, if one is primarily interested in the effect on average outcomes, rather than probabilities of single events, it has considerable advantages. If the ordered probit model is correct, and if the rating scale model is estimated by NLS, it gives us the best (in a minimum mean squared error sense) probit- or logit approximation to the true conditional expectation.

3. Application to Health Satisfaction

Our application is concerned with the relationship between stated health satisfaction (SHS) and health status based on detailed self-reports of physical and mental health impairments (see also Cutler et al., 1997, Sullivan and Karlsson, 1998, Jones et al., 2012). Determining the strength of such a relationship is relevant for at least two reasons. First, it is quite common in applied socio-economic regression analyses that a researcher would like to include detailed controls for health status but the data only provide information on SHS or self-assessed health (SAH). Hence, evidence on the extent to which SHS or SAH proxy for underlying health is highly relevant. Second, knowing how different health impairments affect satisfaction with health can be used in cost-benefit analyses. For example, the benefit, and effectiveness, of different treatment options can be compared by their impact on subjectively perceived SHS.

Specifically, we relate SHS, measured on a Likert scale ranging from 0 to 10, to a 12-item short form survey on health conditions, using data from the 2008 wave of the German Socio Economic Panel (SOEP). The 12-item survey was derived from the so-called SF-36, a 36-item short form survey developed as part of the RAND Medical Outcomes Study (Andersen et al., 2007). Questions include for example the presence, or absence of vitality and frequency of bodily pain during the previous 4 weeks. A factor analysis aggregates the twelve sub-items into two separate scales, one related to physical and one related to mental health. Scores are standardized. They range from 0 to 100, with mean 50 and standard deviation 10.

Results are shown in Table 1. The first specification, in the left panel of the table, includes the scores for physical health and for mental health, as well as the logarithm of household income, gender and age. The first column shows the OLS results, the second the rating scale model estimated by Bernoulli quasi maximum likelihood, and the third column the rating scale model estimated by non-linear least squares. The logit RSM is used throughout. In order to provide comparable results, all parameters are converted to average marginal effects. In this specification, the marginal effects are quite similar. For instance, a one standard deviation increase in the physical health score increases SHS on average by about 1.4 points on the 0 to 10 scale, regardless of specification. The standard deviation of health satisfaction in the sample is about 2.2, resulting in an elasticity of SHS with respect to the physical health score of 0.65; The effect of the mental health score is half as big, which is an interesting, perhaps unexpected finding of our analysis.

Table 1: Determinants of Health Satisfaction ($N = 18{,}030$)

		Rating scale				Rating scale	
	OLS	QML	NLS	OLS	QML	NLS
Physical score $\times 10^{-1}$	1.412	1.424	1.405	0.935	0.647	0.714
	(0.013)	(0.015)	(0.015)	(0.075)	(0.083)	(0.084)
Relative score $\times 10^{-1}$				0.433	0.692	0.639
				(0.065)	(0.072)	(0.073)
Mental score $\times 10^{-1}$	0.722	0.744	0.727			
	(0.012)	(0.013)	(0.013)			
Female	0.168	0.186	0.185	0.003	−0.015	−0.006
	(0.021)	(0.024)	(0.024)	(0.024)	(0.027)	(0.027)
Age	−0.001	−0.001	−0.001	−0.010	−0.020	−0.017
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Log (income)	0.077	0.080	0.080	0.233	0.249	0.248
	(0.019)	(0.022)	(0.022)	(0.021)	(0.023)	(0.024)

Notes: Average marginal effects are reported. Robust standard errors in parentheses. Parametric RSM assumes a logistic cumulative distribution function. QML stands for quasi maximum likelihood estimation, NLS for non-linear least squares. The relative score is the normalized difference between the individual physical score and an average score in a reference group defined by age and gender.

Women tend to report a higher satisfaction with health than men, *ceteris paribus*. While there is an income gradient, it is rather flat. A 10 percent increase in income is predicted to increase health satisfaction by 0.01, a minimal amount indeed. The effect of age is insignificant. Of course, this does not mean that age and income are unimportant for health. Rather, it means that once we account for physical and mental health scores, income and age have no additional predictive power for SHS. Indeed, the R^2 drops only slightly, from 0.548 to 0.546, when we omit all socio-economic variables. In other words, health satisfaction can be seen as a good proxy for health, as it mostly captures the information in the underlying health scores, rather than that of other socio-economic dimensions.

Although the average responses are similar in this case, this does not hold for subgroups. For example, in the group with the highest predicted health satisfaction, say the upper decile, the predicted OLS effect stays constant at +1.4, whereas the average effect computed from the RSM is reduced to +0.96. This attenuation is a reflection of the effect heterogeneity built into the non-linear RSM.

The next specification, in the right panel of Table 1, illustrates that overall average effects can differ quite a bit between OLS and RSM as well. Here, we distinguish between two effects of physical health on health satisfaction, an absolute and a relative one. A relative channel exists if the respondents' satisfaction level is affected by their health status in comparison to the average health status of a reference group. Here, the reference group is defined by gender and age categories (18–24, 25–34, 35–44, …). The relative score takes first the difference between individual scores and average reference group scores, and is then standardized to an average of 5 and a standard deviation of 1.

From Table 1, we see that the sum of absolute and relative effects of a one standard deviation increase is not far from the +1.4 effect of the physical score found above. However, the models attribute different importance to the two channels. While both RSM estimates weigh the two channels about evenly, OLS puts substantially more weight, in fact using a 2:1 ratio, on the absolute, direct effect. This is a real concern, since there is some folklore in applied research using ratings that OLS and other single-index non-linear methods, such as the ordered probit model, get at least the "trade-off ratios" (the ratio of two slope parameters) right, if not their absolute effects. Here, we provide a counterexample.

The crucial differences between the linear model and the RSM become apparent in Figure 1. Mean predictions are displayed in the upper panel. In the lower panel marginal effects are plotted against the predicted values of the linear index. The left graphs show estimates for the linear regression model, the middle graphs the RSM quasi likelihood results and the right graphs result from a semiparametric RSM.

In the top left graph in Figure 1, it can be seen that some OLS in-sample mean predictions violate the upper bound of the dependent variable. In the middle graph the concave flection of the logistic cumulative distribution function constrains mean predictions to respect the upper bound. Also the semiparametrically estimated RSM flattens out as predictions approach the bounds of zero and ten, respectively.

The differences in direct and relative physical health gradients are apparent in the lower panel of Figure 1. Whereas the linear model estimates a constant effect for the entire sample, the RSMs allow for heterogeneous effects in the population. The RSMs suggest that the impact of a change in physical health status on health satisfaction is substantially reduced if an individual is predicted to have either very low or very high health satisfaction. The same effect is found for the semiparametric RSM, where it is data- and not model-driven.

Whether average effects of the different models coincide depends on the distribution of the regressors. Although OLS is known to provide the best linear approximation, in a minimum mean squared error sense, to any non-linear

Figure 1: Predicted Health Satisfaction
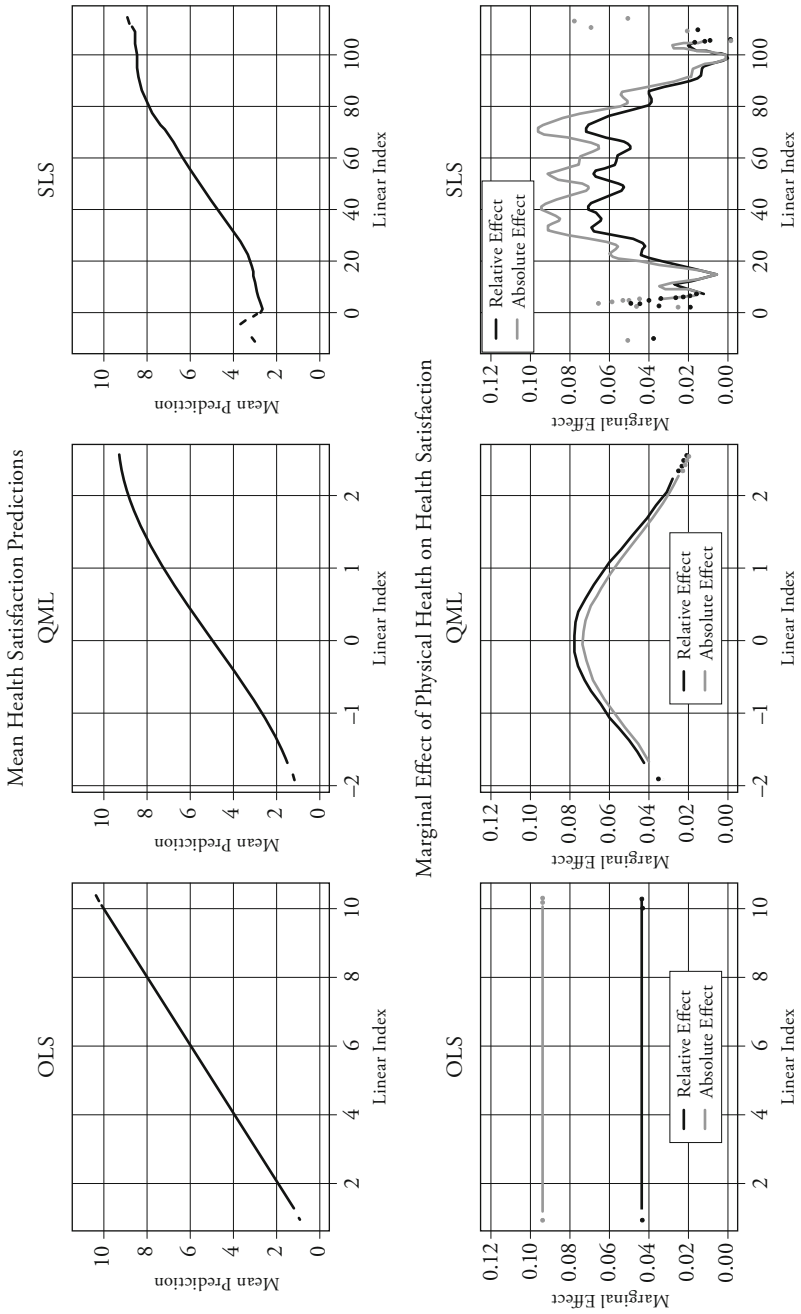
conditional expectation, it does not identify the average effect in general. As shown by YITZHAKI (1996), OLS overweighs effects for regions where x has a large variance. For example, in the case of an RSM, OLS will underpredict the true average effect, if there are few people with very large or very low linear index values (and thus a predicted health satisfaction near 0 or 10), as their close-to-zero marginal effects enter the OLS computation with disproportionate weights.

4. Conclusions

In this paper, we presented a new method for estimating regression models when the dependent variable is a rating. The application to the determinants of stated health satisfaction illustrated the unwanted consequences that can arise when a linear regression model is used for a bounded dependent variable such as a rating. There are inadmissible predictions, and imposing constant marginal effects is implausible. The rating scale models introduced in this paper overcome these shortcomings of a linear model. They can be estimated in a number of ways, including Bernoulli quasi maximum likelihood, nonlinear least squares or even semi-parametric least squares. Among the two parametric approaches, the Bernoulli approach has theoretical advantages, as it is efficient under certain kinds of heteroskedasticity, and it is easy to implement using standard statistical software.

On a substantive level, we found, using a sample of 18 030 respondents from the German Socio-Economic Panel, that stated satisfaction with health is strongly related to two physical and mental health scores that were derived from a well-known health question inventory (a subset of the "SF-36"). Our analysis suggests that OLS results can be quite misleading. In one specification, average effects differed by as much as 50 percent between the OLS and the RSM results. We conclude with two caveats. First, the left-hand side variable in our analysis was "stated" health satisfaction, and results should be interpreted accordingly. For instance, systematic response bias, such as aversion to the extreme responses 0 or 10, would invalidate direct extrapolation to statements about satisfaction per se.

Second, the results in this paper hold regardless of whether one considers the relation between the variables as "mere" associations or as causal. Indeed, any relation between health scores and stated health satisfaction is likely driven by an underlying common factor, "true" health, and thus not causal. However, the same general point, that OLS may lead to non-sense predictions and poor approximations of average effects, applies in a causal analysis, where a rating is regressed on a truly exogenous regressors, for example in the context of a randomized controlled trial.

References

Andersen, Hanfried, Axel Mühlbacher, Mathias Nübling, Jürgen Schupp, and Gerd Wagner (2007), "Computation of Standard Values for Physical and Mental Health Scale Scores Using the SOEP Version of SF-12v2", *Schmollers Jahrbuch*, 127, pp. 171–182.

Cameron, Colin, and Pravin Trivedi (2005), *Microeconometrics, Methods and Applications*, Cambridge University Press, New York.

Clark, Andrew and Andrew Oswald (1996), "Satisfaction and Comparison Income", *Journal of Public Economics*, 61(3), pp. 359–381.

Cutler, David, Elizabeth Richardson, Theodore Keeler and Douglas Staiger (1997), "Measuring the Health of the U. S. Population", *Brookings Papers on Economic Activity. Microeconomics*, pp. 217–282.

Ferrer-i-Carbonell, Ada, and Paul Frijters (2004), "How Important is Methodology for the Estimates of the Determinants of Happiness?", *Economic Journal*, 114, pp. 641–659.

Frijters, Paul, John Haisken-DeNew, and Michael Shields (2005), "The Causal Effect of Income on Health: Evidence from German Reunification", *Journal of Health Economics* 24, pp. 997–1017.

Gourieroux Christian, Alain Monfort, and Alain Trognon (1984), "Pseudo Maximum Likelihood Methods: Theory", *Econometrica*, 52, pp. 681–700.

Ichimura, Hidehiko (1993), "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models", *Journal of Econometrics*, 58, pp. 71–120.

Jones, Andrew, Nigel Rice, Teresa Bago D'Uva, and Silvia Balia (2012), *Applied Health Economics*, Second Edition, Routledge.

Jones, Andrew, and Stefanie Schurer (2011), "How Does Heterogeneity Shape the Socioeconomic Gradient in Health Satisfaction?", *Journal of Applied Econometrics*, 26, pp. 549–579.

Maddala, Gangadharrao (1983), *Limited-Dependent and Qualitative Variables in Economics*, Cambridge University Press.

Papke, Leslie, and Jeffrey Wooldridge (1996), "Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates", *Journal of Applied Econometrics*, 11, pp. 619–632.

Romeu Gordo, Laura (2006), "Effects of Short- and Long-Term Unemployment on Health Satisfaction: Evidence from German Data", *Applied Economics*, 38, pp. 2335–2350.

Studer, Raphael, and Rainer Winkelmann (2014), "Reported Happiness, Fast and Slow", *Social Indicators Research*, 117, pp. 1055–1067.

SULLIVAN, MARIANNE, and JAN KARLSSON (1998), "The Swedish SF-36 Health Survey III: Evaluation of Criterion-Based Validity: Results from Normative Population", *Journal of Clinical Epidemiology*, 51, pp. 1105–1113.

WHITE, HALBERT (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica*, 48, pp. 817–838.

YITZHAKI, SHLOMO (1996), "On Using Linear Regressions in Welfare Economics", *Journal of Business & Economic Statistics*, 14, pp. 478–486.