HEDG
HEALTH, ECONOMETRICS AND DATA GROUP
THE UNIVERSITY of York

WP 16/19

Generalized Difference in Differences for Ordinal Responses with a Varying Number of Categories

Young-sook Kim & Myoung-jae Lee

August 2016

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

Generalized Difference in Differences for Ordinal Responses with a Varying Number of Categories

(July 5, 2016)

Young-sook Kim
Korean Women's
Development Institute
Seoul 03367, South Korea
youngkim@kwdimail.re.kr

Myoung-jae Lee (corresponding author)
Dept. Economics, Korea University
Seoul 02841, South Korea
phone/fax: +82-2-3290-2229
myoungjae@korea.ac.kr

Ordinal responses often represent verbal descriptions (e.g., happy, neutral, and unhappy). Sometimes, ordinal responses with different numbers of categories have to be used together, as in comparing happiness across different periods/countries. One way to proceed is collapsing some of them such that the same ordinal scale holds for all responses. But this loses information, and how to exactly collapse the ordinal responses is not obvious. We show how to use multiple ordinal responses without equalizing the categories, and apply 'generalized difference in differences' (GDD) to them; GDD allows nonparallel untreated trajectories across the treatment and control groups, unlike the popular difference in differences (DD). With GDD, Korean data are used to assess the effects on self-assessed health of an aid program for the severely disabled, where four-wave repeated cross-sections appear, with five categories in the first two years and four in the last two; we find a significant effect with DD, but not with GDD. We also apply our method of dealing with different ordinal responses to the European health data 'SHARE' where two ordinal scales are worded differently.

JEL Classification Numbers: C31, C35, I10.

Key Words: Ordinal response, self-assessed health, difference in differences, generalized difference in differences, minimum distance estimation.

1 Introduction

Ordinal scales are often used to represent health status in health economics (e.g., very healthy, healthy and so on), product satisfaction in marketing (very satisfied, satisfied, ...), credit rating in finance (AAA, AA, ...), just to name a few. From a survey design view point, there are at least three critical decisions to make: (i) how many categories to use (typically, 5 to 10), (ii) whether or not to put a neutral category between the positive and negative categories, and (iii) whether or not to add verbal descriptions for middle categories; e.g., with five being the best and one being the worst, the middle categories 2-4 may not be described or described as good, neutral and bad.

It is well documented that ordinal responses in surveys are affected by how the questions are designed/framed, where in the survey the questions appear, how many times they are asked, and so on (see, e.g., Weijters et al. 2010, Lumsdaine and Exterkate 2013, Moors et al. 2014, and references therein). In health economics, a popular ordinal response is self-assessed health, and its problems have been studied extensively; see, e.g., Crossley and Kennedy 2002, Lindeboom and van Doorslaer 2004, Greene et al. 2014, and references therein.

Suppose we want to use two ordinal responses, one with four categories and the other with five. This can occur when we compare health in two countries, with one country's health recorded in four categories and the other in five. One way to do the task is collapsing the "long" categories, but this is not straightforward, unlike the case of four versus eight where the eight categories might be collapsed into four by merging each two adjacent categories. Also, collapsing categories entails information loss.

Even if the number of the categories is the same, still the verbal descriptions may differ. For instance, the 'Survey of Health, Ageing and Retirement in Europe (SHARE)' provides two five-category self-assessed health variables: the WHO format (very bad, bad, fair, good, and very good) and the US format (poor, fair, good, very good, and excellent). The former is balanced in the negative and positive responses with the middle category 'fair' construed as neutral, whereas the latter is not. Also, although the two formats share the three categories (fair, good, and very good), since they are ranked differently (fair is the third in the former, but the second in the latter), it is not clear whether the respondents regard them as the same or not. Since the two health variables are transformations of the same latent continuous health, we may obtain a more efficient estimator by using both variables.

One goal of this paper is to show how to use ordinal responses with different numbers of categories without collapsing any of them; we searched for the literature, but failed to find any reference. Another goal is to show how to do 'generalized difference in differences' (GDD; Lee 2016a) with ordinal responses. GDD generalizes the popular difference in differences (DD) by allowing nonparallel untreated response trajectories across the treatment and control groups.

We apply our methodology to a Korea data set to find the effects of a supportive at-home service (treatment D) for the severely disabled on their self-assessed health (response Y). The data are four-wave repeated cross-sections, where the first two waves have five categories and the last two have four, and the treatment D is the interaction between the qualification Q (severely disabled) and the time dummy for the last wave. We also apply our method of dealing with different ordinal responses to SHARE.

The remainder of this paper is organized as follows. Section 2 distinguishes three cases of different ordinal responses; to simplify exposition, we consider only two ordinal responses. Section 3 explains the methodology for GDD, and applies it to Korean self-assessed health data. Section 4 explains our method for SHARE, and presents the empirical analysis. Finally, Section 5 concludes our findings.

Putting our findings for the Korean data in advance, DD finds a significant health-enhancing effect for the severely disabled, but GDD does not. The misleading DD effect is due to nonparallel untreated health trajectories of the severely disabled (relatively improved) and the other disabled (not improved). Essentially, DD takes this untreated difference as the treatment effect. As for the SHARE data, we find that the respondents are incoherent/discordant in their answers to the two self-assessed health questions, and that the gain in combining the two answers is small.

2 Three Cases for Different Ordinal Responses

Consider two observed ordinal responses Y_1 and Y_2 generated from latent continuous responses Y_1^* and Y_2^*:

$$Y_{1i}^* = X_{1i}'\beta + U_{1i} \quad \text{and} \quad Y_{2i}^* = X_{2i}'\beta + U_{2i}, \quad i = 1, ..., N$$

where X_{1i} and X_{2i} are regressors, β is the slopes, and U_{1i} and U_{2i} are the error terms. Since we assume iid (independent and identically distributed) across $i = 1, ..., N$, we will often omit the subscript i indexing individuals, as is done here. We assume the same regression function parameter β for Y_1^* and Y_2^* although the error terms differ; if the parameters are different, there is little reason to merge the two data sets. Let

$$X \text{ consist of all elements of } (X_1, X_2).$$

Also let $1[A] = 1$ if A holds and 0 otherwise.

Specifically, suppose that Y_1 and Y_2 are generated from Y_1^* and Y_2^* as follows:

$$Y_{1i} = \sum_{j=1}^{J} 1[\xi_j \le Y_{1i}^*] = \sum_{j=1}^{J} 1[\xi_j - X_{1i}'\beta \le U_{1i}], \quad \xi_1 = 0 < \xi_2 < \cdots < \xi_J; \quad (2.1)$$

$$Y_{2i} = \sum_{l=1}^{L} 1[\zeta_l \le Y_{2i}^*] = \sum_{l=1}^{L} 1[\zeta_l - X_{2i}'\beta \le U_{2i}], \quad \zeta_1 = 0 < \zeta_2 < \cdots < \zeta_L;$$

$$(U_1, U_2) \amalg X \quad (\text{`} \amalg \text{' means independence}) \quad \text{and } SD(U_1) = SD(U_2) = 1.$$

Y_1 takes $J+1$ ordered categories $(0, 1, .., J)$ with J cutoffs, and Y_2 takes $L+1$ ordered categories $(0, 1, .., L)$ with L cutoffs. $SD(U_1) = SD(U_2) = 1$ is a scale normalization, and $\xi_1 = \zeta_1 = 0$ is a location normalization. In using different ordinal responses together, broadly viewed, there are three cases to distinguish in terms of relationship between U_1 and U_2 as follows.

First, U_1 and U_2 are independent, which happens for repeated cross-sections. For instance, Y_1 is a response in the first wave, Y_2 is a response in the second wave, and person i in the first wave is different from person i from the second wave. Given X,

$$Y_1 = q, \; Y_2 = r \iff \xi_q - X_1'\beta < U_1 < \xi_{q+1} - X_1'\beta, \quad \zeta_r - X_2'\beta < U_2 < \zeta_{r+1} - X_2'\beta.$$

Denoting the $N(0,1)$ distribution function as Φ, this renders the product likelihood function (due to the independence)

$$\{\Phi(\xi_{q+1} - X_1'\beta) - \Phi(\xi_q - X_1'\beta)\} \cdot \{\Phi(\zeta_{r+1} - X_2'\beta) - \Phi(\zeta_r - X_2'\beta)\},$$

and maximum likelihood estimator (MLE) can be used. A generalization of this setup applies to our Korean health data that are four wave repeated cross-sections with a varying number of categories.

Second, U_1 and U_2 are related although $U_1 \neq U_2$. For instance, Y_1 stands for a right eyesight, and Y_2 the left eyesight. Let $\Psi(\cdot,\cdot;\rho)$ be the bivariate standard normal distribution function with correlation ρ. Under a joint normality for (U_1, U_2), the event $(Y_1 = q, Y_2 = r)$ renders a likelihood component involving Φ and Ψ for MLE. As another example, Y_1 and Y_2 are responses in two-wave panel data, and

$$U_{1i} = \delta_i + V_{1i} \quad \text{and} \quad U_{2i} = \delta_i + V_{2i} \quad \text{with} \quad V_1 \amalg V_2.$$

Here U_1 and U_2 are dependent by sharing the 'unit-specific effect' δ, but otherwise independent. A product likelihood function obtains, once δ is conditioned on. Since the MLE for this 'random effect panel ordered probit' is a straightforward generalization of random effect panel probit, and can be implemented using a popular econometric package such as STATA, we will not discuss this case in this paper. The above eyesight example can be dealt with also in the same "panel data" unit-specific effect framework.

Third, $Y_1^* = Y_2^*$ (i.e., $X_1 = X_2$, and $U_1 = U_2 \equiv U$). Here, a single error term U (and $X'\beta$) generates two ordinal responses, and there arises the 'coherence' issue, because $Y_1 = good$ and $Y_2 = poor$ may be incoherent; this issue does not arise in the two previous cases. When the two observed responses are incoherent, their likelihood function cannot be constructed; e.g., $Y_1 = good \Longleftrightarrow 0 < U$, and $Y_2 = poor \Longleftrightarrow U < 0$. One way to proceed in this case is using minimum distance estimator (MDE); see Lee (2010) and references therein. For the MDE, apply ordered probit to the two equations separately, and combine only the estimates for β from the two ordered probits. In fact, this MDE approach can be applied to the two preceding cases as well, because the MDE allows any form of relationship between U_1 and U_2, although it is inefficient in general when U_1 and U_2 are related.

3 Repeated Cross-Sections with Varying Categories

3.1 Model and Likelihood Function

In the Korean data, the treatment D is a new at-home service from 2013 and on. The treatment-eligible group ($Q = 1$) is the severely disabled with severity classification 1 or 2, and the control group is the other disabled. Four periods of repeated cross-sections are available at three year intervals: 2005, 2008, 2011 and 2014; only $t = 4$ for 2014 is the treated period. Unfortunately, the number of the self-assessed health Y_{it} categories varies across $t = 1, 2, 3, 4$: five (0 to 4) at $t = 1, 2$ and four (0 to 3) at $t = 3, 4$. This results in complications in the identified parameters, as will be seen shortly.

Let S_i be the sampled period for person i, and let $S_{it} = 1[S_i = t]$, $t = 1, 2, 3, 4$. Assume S_i to be independent of the other random variables in the model. Suppose

$$Y_{it}^* = \beta_1 + \sum_{\tau=2}^{4} \Delta\beta_\tau S_{i\tau} + \beta_q Q_i + \beta_d Q_i S_{i4} + W_i^{'} \beta_w + U_{it} \tag{3.1}$$

where W_{it} are covariates; notice the time-varying intercept represented by the first period intercept β_1 and the increments relative to β_1:

$$\Delta\beta_\tau \equiv \beta_\tau - \beta_1, \qquad \tau = 2, 3, 4.$$

In this DD setup, the treatment effect is β_d for the treatment dummy

$$D_{it} = Q_i S_{i4} = Q_i 1[S_i = 4].$$

The observed ordered response is, with ξ_j^* and ζ_j^* denoting the cutoffs,

$$\begin{aligned} Y_{it} &= \sum_{j=1}^{4} 1[\xi_j^* \le Y_{it}^*], \qquad \xi_1^* < \xi_2^* < \xi_3^* < \xi_4^* \qquad \text{for } t = 1, 2; \\ &= \sum_{j=1}^{3} 1[\zeta_j^* \le Y_{it}^*], \qquad \zeta_1^* < \zeta_2^* < \zeta_3^* \qquad \text{for } t = 3, 4. \end{aligned}$$

Define the ξ_1^*- and ζ_1^*-normalized cutoffs and the intercepts:

$$\xi_j \equiv \xi_j^* - \xi_1^*, \quad \bar{\beta}_1 \equiv \beta_1 - \xi_1^* \qquad \text{and} \qquad \zeta_j \equiv \zeta_j^* - \zeta_1^*, \quad \tilde{\beta}_3 \equiv \beta_3 - \zeta_1^*.$$

Here $\tilde{\beta}_3 \equiv \beta_3 - \zeta_1^*$ appears instead of $\beta_1 - \zeta_1^*$, because $t = 3$ becomes the base period for $t = 3, 4$ when the number of categories changes at $t = 3$.

The appendix shows that the Y_{it} equation can be written as

$$Y_{it} = \sum_{j=1}^{4} 1[\xi_j \le \bar{\beta}_1 + \Delta\beta_2 S_{i2} + \beta_q Q_i + W_i^{'}\beta_w + U_{it}] \quad \text{for } t = 1, 2; \quad (3.2)$$

$$= \sum_{j=1}^{3} 1[\zeta_j \le \tilde{\beta}_3 + (\beta_4 - \beta_3) S_{i4} + \beta_q Q_i + \beta_d Q_i S_{i4} + W_i^{'}\beta_w + U_{it}], \ t = 3, 4. (3.3)$$

The parameters to estimate under $SD(U_{it}) = 1 \ \forall t$ (i.e., the same error term variance for all periods) are, because $\xi_1 = \zeta_1 = 0$,

$$\xi_2, \xi_3, \xi_4, \bar{\beta}_1, \quad \zeta_2, \zeta_3, \tilde{\beta}_3, \quad \gamma \equiv (\Delta\beta_2, \beta_4 - \beta_3, \ \beta_q, \beta_d, \beta_w')'.$$

In (3.3), the slope of S_4 is $\beta_4 - \beta_3$, not $\Delta\beta_4 \equiv \beta_4 - \beta_1$, due to the same reason why the intercept is $\tilde{\beta}_3 \equiv \beta_3 - \zeta_1^*$, not $\beta_1 - \zeta_1^*$: $t = 3$ becomes the base period for $t = 3, 4$ when the number of categories changes at $t = 3$.

Let

$$\delta_{ij} \equiv 1[Y_i = j] \qquad \text{and} \qquad Z_i \equiv (S_{i2}, S_{i4}, \ Q_i, Q_i S_{i4}, W_i')'.$$

The log-likelihood function of person i is

$$(S_{i1} + S_{i2}) \sum_{j=0}^{4} \delta_{ij} \ln P_{12}(Y_i = j | Z_i) \ + \ (S_{i3} + S_{i4}) \sum_{j=0}^{3} \delta_{ij} \ln P_{34}(Y_i = j | Z_i) \qquad (3.4)$$

where P_{12} denotes the likelihood for $t = 1, 2$ such that

$$\begin{aligned}
P_{12}(Y_i &= 0|Z_i) = \Phi(-\bar{\beta}_1 - Z_i'\gamma), \\
P_{12}(Y_i &= 1|Z_i) = \Phi(\xi_2 - \bar{\beta}_1 - Z_i'\gamma) - \Phi(-\bar{\beta}_1 - Z_i'\gamma), \\
P_{12}(Y_i &= 2|Z_i) = \Phi(\xi_3 - \bar{\beta}_1 - Z_i'\gamma) - \Phi(\xi_2 - \bar{\beta}_1 - Z_i'\gamma), \\
P_{12}(Y_i &= 3|Z_i) = \Phi(\xi_4 - \bar{\beta}_1 - Z_i'\gamma) - \Phi(\xi_3 - \bar{\beta}_1 - Z_i'\gamma), \\
P_{12}(Y_i &= 4|Z_i) = 1 - \Phi(\xi_4 - \bar{\beta}_1 - Z_i'\gamma),
\end{aligned}$$

whereas P_{34} denotes the likelihood for $t = 3, 4$ such that

$$\begin{aligned}
P_{34}(Y_i &= 0|Z_i) = \Phi(-\tilde{\beta}_3 - Z_i'\gamma), \\
P_{34}(Y_i &= 1|Z_i) = \Phi(\zeta_2 - \tilde{\beta}_3 - Z_i'\gamma) - \Phi(-\tilde{\beta}_3 - Z_i'\gamma), \\
P_{34}(Y_i &= 2|Z_i) = \Phi(\zeta_3 - \tilde{\beta}_3 - Z_i'\gamma) - \Phi(\zeta_2 - \tilde{\beta}_3 - Z_i'\gamma), \\
P_{34}(Y_i &= 3|Z_i) = 1 - \Phi(\zeta_3 - \tilde{\beta}_3 - Z_i'\gamma).
\end{aligned}$$

3.2 GDD for Ordinal Responses

Let Y_{it}^0 be the potential untreated response, and Y_{it}^1 the potential treated response; $Y_{it} = (1 - D_{it})Y_{it}^0 + D_{it}Y_{it}^1$. DD assumes that, in essence, the untreated response of the treatment group differs from that of the control group by a constant $\beta_q Q_i$. GDD relaxes this assumption by allowing the untreated responses to differ by $\beta_{0q}Q_i + \beta_{1q}tQ_i$. In GDD, the untreated group difference can become wider ($\beta_{1q} > 0$) or narrower ($\beta_{1q} < 0$) over time. As noted in Lee (2016b), this way of relaxing the DD assumption can be seen in other papers as well. For instance, although Jayachandran et al. (2010) did not explicitly use the word GDD, they actually used GDD with $\beta_{0q}Q_i + \beta_{1q}tQ_i$ in one of their models. Angrist and Pischke (2009, p. 238) also suggested to use group-specific pre-treatment trends, which is essentially the same idea.

To allow different pre-treatment trends across the two groups, GDD requires multiple pre-treatment periods; to allow $\beta_{1q}tQ_i$, at least two pre-treatment periods are needed. But if there is no *common* pre-treatment trend at all, then it would be hard to justify attributing the post-treatment difference to the treatment. Hence, it is more reasonable to demand three pre-treatment periods for $\beta_{1q}tQ_i$ so that there is one "degree of freedom" left after estimating the two parameters β_{0q} and β_{1q} in $\beta_{0q}Q_i + \beta_{1q}tQ_i$. Since we have only three pre-treatment periods, we will thus entertain $\beta_{0q}Q_i + \beta_{1q}tQ_i$ mainly, not $\beta_{0q}Q_i + \beta_{1q}tQ_i + \beta_{2q}t^2Q_i$ with no degree of freedom left.

With $\beta_{0q}Q_i + \beta_{1q}tQ_i$ replacing $\beta_q Q_i$, (3.1) becomes

$$Y_{it}^* = \beta_1 + \sum_{\tau=2}^{4} \Delta\beta_\tau S_{i\tau} + \beta_{0q}Q_i + \beta_{1q}tQ_i + \beta_d Q_i S_{i4} + W_i'\beta_w + U_i. \qquad (3.5)$$

Then (3.2) and (3.3) become

$$Y_{it} = \sum_{j=1}^{4} 1[\xi_j \leq \bar{\beta}_1 + \Delta\beta_2 S_{i2} \ + \beta_{0q}Q_i + \beta_{1q}tQ_i \ + W_i'\beta_w + U_i] \qquad \text{for } t = 1, 2;$$

$$= \sum_{j=1}^{3} 1[\zeta_j \leq \tilde{\beta}_3 + (\beta_4 - \beta_3)S_{i4} \ + \beta_{0q}Q_i + \beta_{1q}tQ_i \ + \beta_d Q_i S_{i4} + W_i'\beta_w + U_i], \ t = 3, 4.$$

In the parameters γ to estimate, (β_{01}, β_{1q}) replaces β_q, and tQ is added to the regressors Z. Other than these changes, the log-likelihood function (3.4) still holds.

In panel data with continuously distributed $Y_{it} = Y_{it}^*$, Lee (2016a) showed that DD and GDD can be implemented by differencing the model: the DD effect equals the slope of Q_i in the first differenced model for $\Delta Y_{it} \equiv Y_{it} - Y_{i,t-1}$, and the GDD effect equals the slope of Q_i in the second-differenced model for $\Delta^2 Y_{it} \equiv \Delta Y_{it} - \Delta Y_{i,t-1}$, as illustrated in the next paragraph. Going further, the slope of Q_i in the third differenced model is the effect allowing $\beta_{0q}Q_i + \beta_{1q}tQ_i + \beta_{2q}t^2Q_i$ in the untreated trajectory.

To see the point, difference the following model with treatment applied at $t = 4$:

$$Y_4 = \beta_4 + \beta_{0q}Q + \beta_{1q}4Q + \beta_d Q + U_4 \ \text{ and } \ Y_t = \beta_t + \beta_{0q}Q + \beta_{1q}tQ + U_t \text{ for } t \leq 3$$

$$\Longrightarrow \ \Delta Y_4 = \beta_4 - \beta_3 + \beta_{1q}Q + \beta_d Q + \Delta U_4 \quad (\text{and } \Delta Y_3 = \beta_3 - \beta_2 + \beta_{1q}Q + \Delta U_3),$$

$$\Longrightarrow \ \Delta Y_4 - \Delta Y_3 = \beta_4 - 2\beta_3 + \beta_2 + \beta_d Q + \Delta U_4 - \Delta U_3.$$

For DD, since $\beta_{1q} = 0$, β_d is the slope of Q in the ΔY_4 equation. For GDD, β_d is the slope of Q in the $\Delta Y_4 - \Delta Y_3$ equation. If $\beta_{2q}t^2Q$ were in the model as well, then β_d would not be the slope of Q in the $\Delta Y_4 - \Delta Y_3$ equation; instead, β_d would become the slope of Q in the $(\Delta Y_4 - \Delta Y_3) - (\Delta Y_3 - \Delta Y_2)$ equation.

What are the advantages and disadvantages of the differencing approach just above versus the "tQ-inserting approach" in (3.5)? They are analogous to those of fixed-effect versus random-effect approaches in panel data. Whereas differencing has the advantage of reducing endogeneity, it loses valuable time-constant regressors. In contrast, the tQ-inserting approach would be more vulnerable to endogeneity problems, but it does not lose time-constant regressors. Although somewhat involved, if desired, the differencing approach can be applied to panel data with ordered responses by collapsing ordered responses to binary in various ways; see Lee (2015) and references therein.

3.3 Empirical Analysis with Korean Disability Data

Our Korean data set is drawn from the National Survey on the Disabled that has detailed information on individual and household incomes, expenditures, types of disabilities, satisfaction with life including self-assessed health, etc. The data set was collected by the Korea Institute for Health and Social Affairs. In the pooled sample of four periods, there are 15581 individuals, among whom 4382 have $Q = 1$. Table 1 describes Y for the $Q = 0$ group, and Table 2 for the $Q = 1$ group. Naturally, Table 2 shows higher negative response proportions than Table 1.

Table 1: Observation Numbers (%): $Q = 0$ with $N = 11199$

	2005	2008	2011	2014
very healthy	28 (1.8)	75 (2.5)	138 (4.7)	95 (2.6)
healthy	331 (20.8)	384 (13.0)	793 (26.8)	910 (24.6)
average	580 (36.5)	723 (24.5)		
unhealthy	463 (29.1)	1166 (39.5)	1351 (45.6)	1679 (45.4)
very unhealthy	189 (11.9)	601 (20.4)	680 (23.0)	1013 (27.4)
total	1591	2949	2962	3697

Table 2: Observation Numbers (%): $Q = 1$ with $N = 4382$

	2005	2008	2011	2014
very healthy	9 (1.4)	43 (2.5)	39 (4.1)	36 (3.4)
healthy	82 (12.9)	202 (11.6)	214 (22.3)	215 (20.6)
average	129 (20.3)	308 (17.7)		
unhealthy	188 (29.5)	554 (31.8)	324 (33.8)	335 (32.1)
very unhealthy	229 (35.9)	633 (36.4)	383 (39.9)	459 (43.9)
total	637	1740	960	1045

Table 3 describes the regressors: age, education years, four most frequent types of disability, gender, marital status, the dummy for whether visiting doctors regularly, the number of children, the logarithm of monthly family income (KRW stands for Korean Won; \$1 $\simeq$KRW 1000), the logarithm of monthly extra expense due to disability, self-assessed SES (Socio-Economic Status) in three levels, and level of dependency graded by the respondent in five levels.

Table 3: Regressors for All Observations ($N = 15581$)

Variables		Mean (SD)	Min, Max
age		59.33 (15.53)	18, 106
education years	7∼9	0.17 (0.37)	0, 1
	10∼12	0.25 (0.44)	0, 1
	13∼	0.13 (0.34)	0, 1
disability types	physical disabilities	0.35 (0.48)	0, 1
	brain injuries	0.13 (0.34)	0, 1
	blind, low vision	0.12 (0.33)	0, 1
	deaf, hearing difficulty	0.12 (0.32)	0, 1
male		0.56 (0.50)	0, 1
married		0.50 (0.50)	0, 1
whether visiting doctors regularly		0.81 (0.39)	0, 1
# children		1.89 (1.90)	0, 10
ln(monthly family income in KRW 10000)		4.95 (0.82)	0.7, 9.2
ln(monthly extra expense in KRW 1000)		4.09 (1.32)	0.7, 9.3
SES (Socio-Eco Status)	middle	0.32 (0.47)	0, 1
	high	0.01 (0.08)	0, 1
dependency level	score 2	0.18 (0.38)	0, 1
(self-graded in scores 1∼5)	score 3	0.19 (0.39)	0, 1
	score 4	0.09 (0.29)	0, 1
	score 5	0.06 (0.25)	0, 1

The estimation results for DD and GDD are in Tables 4 and 5. Table 4 presents the treatment effect β_d, the group specific effect β_q or (β_{0q}, β_{1q}) and the cutoffs, and Table 5 shows β_w. *In Table 4, DD shows a significantly positive effect* $\beta_d = 0.09$, *but GDD does not. The significantly positive* $\beta_{1q} = 0.11$ *in GDD indicates that the untreated responses do not move parallel across the two groups: the significant DD effect is due* *to omitting* tQ. In essence, DD mistakes $\beta_{1q} = 0.11$ for $\beta_d = 0.09$ whose magnitudes are almost the same.

Except for β_d and β_q (or (β_{0q}, β_{1q})), DD and GDD estimates are almost the same in Tables 4 and 5. Those estimates are self-explanatory. The positive effects of some disability types may be surprising, but they are relative to the omitted 10 other disability types (epileptic, facial, kidney, cardiac, respiratory, autistic, etc.).

For the sake of comparison, we also added t^2Q to the GDD model allowing for a quadratic pre-treatment difference in trend, and we obtained (the other slope estimates are similar to those for DD and GDD, and thus omitted)

$$\beta_d : -0.23\ (-1.38), \quad \beta_{0q} : -0.22\ (-1.24), \quad \beta_{1q} : -0.042\ (-0.23), \quad \beta_{2q} : 0.036\ (0.81).$$

This shows still no significant treatment effect, and the estimates for β_{1q} and β_{2q} are insignificant due to the high collinearity between tQ and t^2Q.

Table 4: MLE for Self-Assessed Health

	DD	GDD
Coefficient	Estimate (tv)	Estimate (tv)
β_d	0.09 (2.02)	-0.11 (-1.50)
β_q or β_{0q}	-0.14 (-5.44)	-0.36 (-5.16)
β_{1q}		0.11 (3.41)
$\bar{\beta}_1$	2.02 (24.6)	2.06 (24.7)
$\tilde{\beta}_3$	1.77 (21.3)	1.75 (20.9)
$\beta_2 - \beta_1$	-0.22 (-6.52)	-0.24 (-6.99)
$\beta_4 - \beta_3$	-0.10 (-3.52)	-0.07 (-2.59)
ξ_2	1.05 (57.4)	1.06 (57.4)
ξ_3	1.92 (83.0)	1.93 (83.2)
ξ_4	3.09 (80.3)	3.09 (80.6)
ζ_2	1.32 (72.0)	1.32 (71.9)
ζ_3	2.74 (87.6)	2.74 (87.5)

Table 5: MLE for Self-Assessed Health (Covariates)

Variables		DD Est. (tv)	GDD Est. (tv)
age		-0.02 (-20.5)	-0.02 (-20.5)
education years	7∼9	-0.00 (-0.10)	-0.00 (-0.08)
	10∼12	0.06 (2.48)	0.06 (2.50)
	13∼	0.11 (3.48)	0.11 (3.52)
disability types	physical disabilities	0.02 (0.58)	0.02 (0.62)
	brain injuries	-0.10 (-2.89)	-0.09 (-2.79)
	blind, low vision	0.30 (9.53)	0.30 (9.57)
	deaf, hearing difficulty	0.30 (8.71)	0.30 (8.74)
male		0.13 (6.78)	0.13 (6.74)
married		-0.02 (-0.76)	-0.02 (-0.76)
whether visiting doctors regularly		-0.39 (-16.4)	-0.39 (-16.4)
# children		0.004 (0.49)	0.004 (0.56)
ln(monthly income of family in KRW 10000)		0.09 (7.16)	0.09 (7.10)
ln(monthly extra expense in KRW 1000)		-0.07 (-9.79)	-0.07 (-9.81)
SES (Socio-Eco Status)	middle	0.25 (12.4)	0.25 (12.4)
	high	0.65 (6.28)	0.65 (6.23)
dependency level	score 2	-0.11 (-4.4)	-0.11 (-4.48)
(self-graded in scores 1∼5)	score 3	-0.34 (-13.6)	-0.35 (-13.7)
	score 4	-0.57 (-17.1)	-0.58 (-17.2)
	score 5	-1.10 (-25.7)	-1.10 (-25.7)

We searched for the reason why $\beta_{1q} > 0$ happened, and we found that a new disability pension started in 2010, just before $t = 3$. The pension is for the severely disabled who are unable to work. The pension comprises a basic benefit of about \$200 per month to ensure a minimum level of expenditure, and extra benefits of about \$20 $\sim$ \$280 to cover additional expenditures. To receive the pension, the person must be at least 18 years old with the severity classification 1 to 3; if unmarried, the income should be less than about \$1,000 per month, and if married, the combined income of the couple should be less than about $1,600. Our conjecture is that this new disability pension in 2010 is the reason for $\beta_{1q} > 0$.

To better understand the DD and GDD difference, examine Figure 1 under no true effect. Suppose the four points A, B, C and D are for the treatment group, and A, B, E and F are for the control group. Because the levels of E and F are the same, DD simply assumes that the counter-factual untreated response of the treatment group at $t = 4$ is the point G that has the same level as C: consequently, DD concludes a false positive effect that equals the vertical difference $\overline{DG}$. In contrast, GDD uses A, B and C to detect the upward nonparallel trajectory of the treatment group, which means taking D as the untreated response at $t = 4$ to correctly conclude no effect.

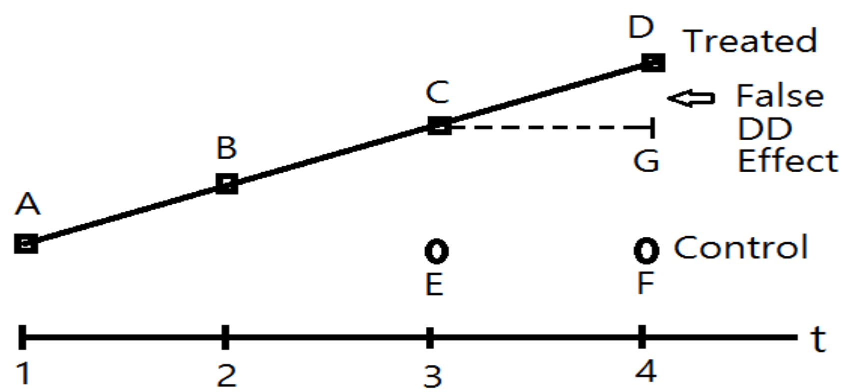

Figure 1: Positive DD under False Parallellism vs Zero GDD (True No Effect)

Had we used only two periods $t = 3, 4$ because the ordinal response category changed "annoyingly" at $t = 3$, we could not have implemented GDD; instead, DD would have given a false positive effect. By backtracking a couple of periods and then actively addressing the problem of the ordinal response category change, we were able to apply GDD and found out the misleading aspect of DD. It is advisable to try GDD by obtaining more untreated periods, rather than simply applying DD to arrive at a conclusion in haste. In the two empirical examples of Lee (2016a,b), GDD with tQ was adequate, and adding t^2Q extra was over-specifying as in the above example.

4 Two Responses from Single Latent Equation

4.1 Model and Coherence

Recall the SHARE five-category self-assessed health formats:

WHO (Y_1): very bad, bad, *fair, good, very good*;

US (Y_2): poor, *fair, good, very good*, excellent;

the three verbal descriptions (fair, good, and very good) in italics appear in both formats. These categories are generated from the same Y^* by dividing its range into five intervals. Our model and notation for these differ somewhat from (2.1):

$$\begin{aligned} Y_{1i} &= \sum_{j=1}^{4} 1[\gamma_{1j} \le Y_i^*] = \sum_{j=1}^{4} 1[\gamma_{1j} - \beta_1 - X_i'\beta_x \le U_i], \quad \gamma_{11} < \gamma_{12} < \gamma_{13} < \gamma_{14}, \\ Y_{2i} &= \sum_{j=1}^{4} 1[\gamma_{2j} \le Y_i^*] = \sum_{j=1}^{4} 1[\gamma_{2j} - \beta_1 - X_i'\beta_x \le U_i], \quad \gamma_{21} < \gamma_{22} < \gamma_{23} < \gamma_{24} \end{aligned}$$

where the cutoffs are γ's, the intercept is β_1, the slope is β_x, and X excludes unity.

If we try to use both Y_1 and Y_2 jointly for MLE, then the two sets of cutoff γ_{1j}'s and γ_{2j}'s should be ordered such that $(Y_1 = q, Y_2 = r)$ is mapped one-to-one to an interval on Y^* for each value of (q, r). This increases the number of categories, and thus reveals more information on Y^*—an example is in the next paragraph. But we do not know the cutoffs before estimation, and there is no guarantee that such a mapping exists for all observed values of (q, r). This is a "coherence" issue.

A strict form of coherence is requiring the same cutoffs for the three common verbal descriptions as in the following:

$$\begin{array}{lcccccc} & & \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \\ & \longrightarrow\longrightarrow\longrightarrow & \longmapsto\longrightarrow & \longmapsto & \longmapsto\longrightarrow & \longmapsto\longrightarrow\longrightarrow & \longmapsto\longrightarrow\longrightarrow \\ Y_1: & \text{very bad} & \text{bad} & \text{fair} & \text{good} & \text{very good} & \\ Y_2: & \multicolumn{2}{c}{\text{poor}} & \text{fair} & \text{good} & \text{very good} & \text{excellent} \\ & & & \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} \end{array} \tag{4.1}$$

There are six combined categories in (4.1), that is, six intervals on Y^*:

$$
\begin{aligned}
(Y_1, Y_2) &= (\textit{very bad},\ \textit{poor}) \iff Y^* \in (-\infty, \gamma_{11}); \\
(Y_1, Y_2) &= (\textit{bad},\ \textit{poor}) \iff Y^* \in (\gamma_{11}, \gamma_{12}) \quad \text{with } \gamma_{12} = \gamma_{21}; \\
(Y_1, Y_2) &= (fair,\ fair) \iff Y^* \in (\gamma_{12}, \gamma_{13}) \quad \text{with } \gamma_{13} = \gamma_{22} \text{ additionally}; \\
(Y_1, Y_2) &= (\textit{good},\ \textit{good}) \iff Y^* \in (\gamma_{13}, \gamma_{14}) \quad \text{with } \gamma_{14} = \gamma_{23} \text{ additionally}; \\
(Y_1, Y_2) &= (\textit{very good},\ \textit{very good}) \iff Y^* \in (\gamma_{14}, \gamma_{24}); \\
(Y_1, Y_2) &= (\textit{very good},\ \textit{excellent}) \iff Y^* \in (\gamma_{24}, \infty).
\end{aligned}
$$

The coherence in (4.1) differs from "literal concordance" in Jürges et al. (2008) or "word concordance" in Lumsdaine and Exterkate (2013): an individual's responses to both formats are verbally the same. The coherence in (4.1) also differs from "relative concordance" Jürges et al. (2008) or "numerical concordance" in Lumsdaine and Exterkate (2013): an individual's responses to both formats are the same in terms of their position in the self-assessed health scale.

In (4.1), combinations such as $(Y_1, Y_2) = (fair, \textit{good})$ and $(Y_1, Y_2) = (\textit{very good}, \textit{good})$ are not allowed. Hence, if these combinations exist in the data (they do in SHARE), the MLE cannot be done using those observations. MLE may be done excluding those incoherent observations, which, however, raises a sample selection issue. We examine MDE next that does not require any coherence.

4.2 MDE without Coherence

Coherence is something that we do not have to necessarily deal with: as long as we learn about β_x through Y_1 and Y_2, how the respondent interprets the verbal descriptions does not matter. Suppose we separately apply ordered probit to Y_1 and Y_2 under $U \sim N(0,1)$, after the location normalization of subtracting the first cutoff from the intercept and all the other cutoffs in the latent equation:

$$\begin{aligned}
Y_1 &= \sum_{j=1}^{4} 1[\gamma_{1j} - \gamma_{11} - (\beta_1 - \gamma_{11}) - X'\beta_x \le U] = \sum_{j=1}^{4} 1[\tau_{1j} - \alpha_1 - X'\beta_x \le U], \\
Y_2 &= \sum_{j=1}^{4} 1[\gamma_{2j} - \gamma_{21} - (\beta_1 - \gamma_{21}) - X'\beta_x \le U] = \sum_{j=1}^{4} 1[\tau_{2j} - \alpha_2 - X'\beta_x \le U], \\
\tau_{1j} &\equiv \gamma_{1j} - \gamma_{11},\ \alpha_1 \equiv \beta_1 - \gamma_{11}, \quad \tau_{2j} \equiv \gamma_{2j} - \gamma_{21},\ \alpha_2 \equiv \beta_1 - \gamma_{21} \quad (\tau_{11} = \tau_{21} = 0).
\end{aligned}$$

We can combine the β_x estimates in the two ordered probits using MDE as follows, which is supposed to be more efficient than the two separate ordered probits.

Let the identified parameters for the order probits be

$$\begin{aligned}
Y_1 &: \psi_1 \equiv (\tau_{12}, \tau_{13}, \tau_{14}, \alpha_1, \beta_x^{1\prime})', \\
Y_2 &: \psi_2 \equiv (\tau_{22}, \tau_{23}, \tau_{24}, \alpha_2, \beta_x^{2\prime})';
\end{aligned}$$

the MDE restriction is $\beta_x^1 = \beta_x^2$ $(\equiv \beta_x)$. In the first step of the MDE, $\hat{\psi}_1$ and $\hat{\psi}_2$ are obtained. In the second step, $\beta_x^1 = \beta_x^2$ is imposed to estimate

$$\psi \equiv (\tau_{12}, \tau_{13}, \tau_{14}, \alpha_1,\ \tau_{22}, \tau_{23}, \tau_{24}, \alpha_2,\ \beta_x')'.$$

For the second step, express $\beta_x^1 = \beta_x^2$ as

$$\begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix} = R_0\psi,\ R_0 \equiv \begin{bmatrix} I_4 & 0_{4\times4} & 0_{4\times k} \\ 0_{k\times4} & 0_{k\times4} & I_k \\ 0_{4\times4} & I_4 & 0_{4\times k} \\ 0_{k\times4} & 0_{k\times4} & I_k \end{bmatrix} \iff \begin{bmatrix} \tau_{12} \\ \tau_{13} \\ \tau_{14} \\ \alpha_1 \\ \beta_x^1 \\ \tau_{22} \\ \tau_{23} \\ \tau_{24} \\ \alpha_2 \\ \beta_x^2 \end{bmatrix} = \begin{bmatrix} I_4 & 0_{4\times4} & 0_{4\times k} \\ 0_{k\times4} & 0_{k\times4} & I_k \\ 0_{4\times4} & I_4 & 0_{4\times k} \\ 0_{k\times4} & 0_{k\times4} & I_k \end{bmatrix} \begin{bmatrix} \tau_{12} \\ \tau_{13} \\ \tau_{14} \\ \alpha_1 \\ \tau_{22} \\ \tau_{23} \\ \tau_{24} \\ \alpha_2 \\ \beta_x \end{bmatrix}$$

where I_4 is the 4×4 identity matrix, $0_{4\times4}$ is the 4×4 null matrix, and X is of dimension $k\times1$. There are $2(4+k) = 8+2k$ parameters in total in the two ordered probits, and $8+k$ parameters in the second stage MDE; the dimension of R_0 is $(8+2k)\times(8+k)$.

Using the above with ψ_j's replaced by its estimator $\hat{\psi}_j$'s, the MDE $\hat{\psi}^0_{mde}$ for ψ is

$$\hat{\psi}^0_{mde} \equiv (R_0'\hat{\Lambda}^{-1}R_0)^{-1}R_0'\hat{\Lambda}^{-1}\hat{\psi}_{12} \qquad \text{where}$$

$$\hat{\psi}_{12} \equiv (\hat{\psi}_1', \hat{\psi}_2')', \quad \hat{\Lambda} \equiv \frac{1}{N}\sum_i \hat{\eta}_i\hat{\eta}_i', \quad \hat{\eta}_i \equiv (\hat{\eta}_{i1}', \hat{\eta}_{i2}')' \to^p \eta_i \equiv (\eta_{i1}', \eta_{i2}')',$$

$$\eta_{ij} \equiv (\frac{1}{N}\sum_i s_{ij}s_{ij}')^{-1}s_{ij} \quad \text{and } s_{ij} \text{ is the ordered probit score function, } j = 1, 2.$$

It holds that, with '$\rightsquigarrow$' denoting convergence in distribution,

$$\sqrt{N}(\hat{\psi}^0_{mde} - \psi) \rightsquigarrow N\{0,\ (R_0'\Lambda^{-1}R_0)^{-1}\} \qquad \text{where } \ \Lambda \equiv E(\eta\eta');$$

$$T_N^0 \equiv N(\hat{\psi}_{12} - R_0\hat{\psi}^0_{mde})'\hat{\Lambda}^{-1}(\hat{\psi}_{12} - R_0\hat{\psi}^0_{mde}) \rightsquigarrow \chi_k^2.$$

T_N^0 is an over-identification test statistic for '$H_0 : \beta_x^1 = \beta_x^2$' that is analogous to the well-known GMM over-identification test.

The efficiency of the above 'unrestricted MDE' can be enhanced by imposing a weak form of coherence. Instead of the strict form of coherence in (4.1), a weak form of coherence requires the *same length of the Y^* range for 'fair' and 'good'* across the WHO and US formats; for 'very good', this cannot be done because its range is open-ended in the WHO format but not in the US format. Specifically, the weak coherence restrictions are, recalling (4.1),

$$(i) : \tau_{13} - \tau_{12}\ (= \gamma_{13} - \gamma_{12}) = \tau_{22}\ (= \gamma_{22} - \gamma_{21}) \qquad \text{for 'fair';}$$

$$(ii) : \tau_{14} - \tau_{13} = \tau_{23} - \tau_{22} \qquad \text{for 'good'} \tag{4.2}$$

$$\iff \tau_{22} + \tau_{14} - \tau_{13} = \tau_{23} \qquad (\text{moving } \tau_{22} \text{ to the opposite side})$$

$$\Longrightarrow \tau_{13} - \tau_{12} + \tau_{14} - \tau_{13} = \tau_{14} - \tau_{12} = \tau_{23} \qquad (\text{using } (i)).$$

In the appendix, we explain the 'restricted MDE' incorporating (4.2). The restricted MDE saves two degrees of freedom, because six cutoffs except τ_{22} and τ_{23} in (4.2) are estimated, rather than eight in the unrestricted MDE.

4.3 Empirical Analysis with SHARE 2004

The data used were drawn from SHARE 2004, release version 2.6.0 (see Börsch-Supan et al. 2005, Börsch-Supan and Jürges 2005, Börsch-Supan 2013, and Börsch-Supan et al. 2013). This particular wave of SHARE used random assignment to obtain responses to two formats of self-assessed health questions: one half of the respondents were randomized to receive one format in the beginning of the survey and the other format later, and the other half in the reverse order. The present analysis is based on the data for Austria, Germany, the Netherlands, Belgium, Spain and Greece, because the overlapping answer categories for the two formats were verbally identical in these countries (see Table 1 in Jürges et al. 2008). Table 6 presents the observation numbers (percentages) for the response categories.

Table 6: Observation Numbers for Response Categories (%): $N = 15880$

	US version					
WHO ↓	excellent	very good	good	fair	poor	*Total*
very good	1059 (6.7)	1452 (9.1)	313 (2.0)	9 (0.057)	2 (0.013)	*2835 (18)*
good	275 (1.7)	1825 (11)	4728 (30)	422 (2.7)	6 (0.038)	*7256 (46)*
fair	11 (0.069)	107 (0.67)	1240 (7.8)	2883 (18)	147 (0.93)	*4388 (28)*
bad	0 (0.000)	0 (0.00)	53 (0.33)	473 (3.0)	618 (3.9)	*1144 (7.2)*
very bad	1 (0.000)	0 (0.00)	1 (0.006)	26 (0.16)	229 (1.4)	257 *(1.6)*
Total	*1346 (8.5)*	*3384 (21)*	*6335 (40)*	*3813 (24)*	*1002 (6.3)*	*15880*

Table 7: Regressors ($N = 15880$)

Variables		Mean (SD)	Variables		Mean (SD)
female		0.55 (0.50)	limitations	1 or 2	0.13 (0.34)
age	60-69	0.28 (0.45)		3 or more	0.054 (0.23)
	70-79	0.19 (0.39)	depression	4 or higher	0.23 (0.42)
	80+	0.079 (0.27)	country	Austria	0.10 (0.30)
education	medium	0.31 (0.46)		Germany	0.18 (0.39)
	high	0.19 (0.39)		Netherlands	0.18 (0.38)
diseases	1 or 2	0.53 (0.50)		Spain	0.14 (0.34)
	3 or more	0.19 (0.40)		Greece	0.17 (0.38)
symptoms	1 or 2	0.48 (0.50)			
	3 or more	0.19 (0.39)			

Table 7 describes the regressors: gender, age (with 50-59 as the omitted base age category), the level of education using the UNESCO international classification of education, the number of chronic diseases ("diseases") ever diagnosed, the number of self-reported symptoms, and the number of limitations with ADL (Activities of Daily Living) or IADL (Instrumental Activities of Daily Living), depression score ("depression") being four or higher as measured by the Euro-Depression (a scale of depression symptoms validated for the European population), and the country dummies. The sampling was not random, and Belgium (the omitted base country category) is over-represented in the data.

The regressors were chosen following Jürges et al. (2008), but some differences should be noted between Jürges et al. (2008) and this paper. First, Jürges et al. (2008) used the SHARE 2004 release version 1, whereas we use a later version 2.6.0. Second, Jürges et al. (2008) did not use Belgium, but we do as already noted. Third, the sample size of Jürges et al. (2008) is 11643, and our sample size is much larger (15880). Fourth, high values of the response variables in Jürges et al. (2008) mean worse health, whereas they mean better health in this paper. Fifth, Jürges et al. (2008) did the location normalization for ordered probit by setting the intercept at zero, whereas we set the first threshold at zero.

Table 8: Ordered Probits and MDE (cutoffs and intercepts)

	WHO	US	MDE
Parameters	Est. (tv)	Est. (tv)	Est. (tv)
τ_{12}	1.12 (36.6)		1.12 (36.9)
τ_{13}	2.58 (76.1)		2.57 (78.0)
τ_{14}	4.30 (117)		4.28 (121)
α_1	4.35 (90.4)		4.29 (96.8)
τ_{22}		1.47 (68.5)	1.47 (70.3)
τ_{23}		2.91 (117)	2.92 (122)
τ_{24}		3.98 (141)	3.98 (148)
α_2		3.41 (86.3)	3.44 (93.7)

Tables 8 presents the ordered probit and MDE results for the cutoffs and intercepts, and Table 9 for the slopes. Judging by the t-values, MDE is more efficient than the two separate order probits, but the efficiency gain with this particular data set is rather modest. The MDE over-identification test at the bottom of Table 9 rejects. Because this is an omnibus test, what went wrong could be things other than the MDE restriction $\beta_x^1 = \beta_x^2$; e.g., the regression function might have been misspecified.

Table 9: Ordered Probits and MDE (slopes)

		WHO	US	MDE
Variables		Est. (tv)	Est. (tv)	Est. (tv)
female		0.16 (8.66)	0.14 (7.88)	0.15 (9.02)
age	60-69	-0.054 (-2.50)	-0.044(-2.10)	-0.047 (-2.41)
	70-79	-0.17 (-6.55)	-0.15 (-5.99)	-0.16 (-6.76)
	80+	-0.13 (-3.79)	-0.056 (-1.63)	-0.10 (-3.05)
education	medium	0.23 (10.0)	0.23 (10.4)	0.23 (11.2)
	high	0.39 (14.9)	0.40 (16.3)	0.40 (17.3)
diseases	1 or 2	-0.75 (-32.1)	-0.69 (-31.2)	-0.71 (-34.8)
	3 or more	-1.19 (-37.1)	-1.15 (-37.1)	-1.16 (-40.8)
symptoms	1 or 2	-0.47 (-21.3)	-0.47 (-22.4)	-0.47 (-24.1)
	3 or more	-0.89 (-28.3)	-0.90 (-29.1)	-0.89 (-31.4)
limitations	1 or 2	-0.47 (-17.3)	-0.50 (-18.6)	-0.48 (-19.6)
	3 or more	-1.07 (-26.3)	-1.11 (-27.4)	-1.10 (-29.3)
depression	4 or higher	-0.45 (-20.1)	-0.42 (-19.2)	-0.43 (-21.4)
country	Austria	-0.40 (-12.1)	-0.28 (-8.58)	-0.33 (-11.0)
	Germany	-0.64 (-21.6)	-0.61 (-21.1)	-0.62 (-23.4)
	Netherlands	-0.21 (-7.46)	-0.19 (-7.06)	-0.20 (-7.95)
	Spain	-0.33 (-10.9)	-0.32 (-10.6)	-0.33 (-11.9)
	Greece	-0.072 (-2.57)	-0.12 (-4.21)	-0.090 (-3.54)
Over-ID test statistic:		66.2 (p-value 0.000)		

We also tried the restricted MDE described in the appendix, but the over identification test statistic value is huge (338) to reject the weak form of coherence. This means that the individual answers are incoherent across the WHO and US formats. It would be a futile exercise to try to learn more by imposing coherence on the two sets of answers.

5 Conclusions

In this paper, we addressed the problem of how to jointly use ordinal responses with different numbers of categories without equalizing the category numbers. We also applied 'generalized difference in differences' (GDD) to ordinal responses that generalizes the popular difference in differences (DD) by allowing nonparallel untreated response trajectories across the treatment and control groups.

We distinguished three cases of different ordinal responses, depending on how they are generated: (i) by two independent latent equations, (ii) by two related latent equations, and (iii) by a single common latent equation. Since the second case has been addressed in the literature using panel or cross-section multiple equation set-ups, we focused on (i) with repeated cross-sections and (iii) with two cross-section equations.

For (i), we showed the identified parameters when the number of the ordinal categories changes once over time, and then applied GDD using Korean data. In our empirical analysis, DD showed a significant effect whereas GDD did not. We attributed this discrepancy to the false parallel trajectory assumption of DD. It is thus recommended to obtain multiple pre-treatment periods and implement GDD, instead of settling with DD which requires only one pre-treatment period (and one post).

For (iii), we showed how to handle two ordinal responses with different verbal descriptions, despite that both were generated by a common latent equation. MDE that does not require 'coherence/concordance' in the two ordinal responses was proposed, and another MDE that requires a weak form of coherence/concordance was also proposed. The methods were then applied to a wave in SHARE, where the efficiency gain of MDE over two separate ordered probits turned out to be modest.

APPENDIX

Identified Parameters in Repeated Cross-Sections with Varying Categories

Substituting the Y_{it}^* equation in (3.1) into the Y_{it} equations and subtracting ξ_1^* and ζ_1^* from both sides gives, with $\tilde{\beta}_1 \equiv \beta_1 - \zeta_1^*$,

$$\begin{aligned} Y_{it} &= \sum_{j=1}^{4} 1[\xi_j \ \leq \bar{\beta}_1 + \sum_{\tau=2}^{4} \Delta\beta_\tau S_{i\tau} + \beta_q Q_i + \beta_d Q_i S_{i4} + W_i'\beta_w + U_{it}] \qquad \text{for } t = 1, 2; \\ &= \sum_{j=1}^{3} 1[\zeta_j \ \leq \tilde{\beta}_1 + \sum_{\tau=2}^{4} \Delta\beta_\tau S_{i\tau} + \beta_q Q_i + \beta_d Q_i S_{i4} + W_i'\beta_w + U_{it}] \qquad \text{for } t = 3, 4. \end{aligned}$$

The subtraction by ξ_1^* and ζ_1^* makes the intercept change as t changes from 2 to 3.

Observe

$$\begin{aligned} &\sum_{\tau=2}^{4} \Delta\beta_\tau S_{i\tau} = \Delta\beta_2 S_{i2} \qquad \text{for } t = 1, 2 \text{ in the first line;} \\ &\sum_{\tau=2}^{4} \Delta\beta_\tau S_{i\tau} = \Delta\beta_3 S_{i3} + \Delta\beta_4 S_{i4} \qquad \text{for } t = 3, 4 \text{ in the second line.} \end{aligned}$$

Whereas the first term with only S_{i2} is all right, the second with both S_{i3} and S_{i4} needs a modification, because $S_{i3} + S_{i4} = 1$ when $t = 3, 4$ that is perfectly collinear with the regressor 1 for $\tilde{\beta}_1$. Hence, when $t = 3, 4$, rewrite the intercept part as

$$\begin{aligned} &\tilde{\beta}_1 + \sum_{\tau=2}^{4} \Delta\beta_\tau S_{i\tau} = \tilde{\beta}_1 + \Delta\beta_3(1 - S_{i4}) + \Delta\beta_4 S_{i4} = (\tilde{\beta}_1 + \Delta\beta_3) + (\Delta\beta_4 - \Delta\beta_3)S_{i4} \\ &= \beta_1 - \zeta_1^* + \beta_3 - \beta_1 + (\beta_4 - \beta_3)S_{i4} = \tilde{\beta}_3 + (\beta_4 - \beta_3)S_{i4}. \end{aligned}$$

Now only $(1, S_4)$ appears as regressors, not $(1, S_3, S_4)$.

Restricted MDE under Weak Coherence

Replace ψ by

$$\psi_r \equiv (\tau_{12}, \tau_{13}, \tau_{14}, \alpha_1, \ \tau_{24}, \alpha_2, \ \beta_x')',$$

and R_0 by the following R_1 so that $(\psi_1', \psi_2') = R_1\psi_r$ that equals

$$\begin{bmatrix} \tau_{12} \\ \tau_{13} \\ \tau_{14} \\ \alpha_1 \\ \beta_x^1 \\ \tau_{22} \\ \tau_{23} \\ \tau_{24} \\ \alpha_2 \\ \beta_x^2 \end{bmatrix} = R_1 \begin{bmatrix} \tau_{12} \\ \tau_{13} \\ \tau_{14} \\ \alpha_1 \\ \tau_{24} \\ \alpha_2 \\ \beta_x \end{bmatrix} \quad \text{where} \quad \underset{(8+2k)\times(6+k)}{R_1} \equiv \begin{bmatrix} I_4 & 0_{4\times 2} & 0_{4\times k} \\ 0_{k\times 4} & 0_{k\times 2} & I_k \\ -1,1,0,0 & 0,0 & 0_{1\times k} \\ -1,0,1,0 & 0,0 & 0_{1\times k} \\ 0_{2\times 4} & I_2 & 0_{2\times k} \\ 0_{k\times 4} & 0_{k\times 2} & I_k \end{bmatrix}.$$

Then

$$\hat{\psi}^1_{mde} \equiv (R_1'\hat{\Lambda}^{-1}R_1)^{-1}R_1'\hat{\Lambda}^{-1}\hat{\psi}_{12};\ \ T_N^1 \equiv N(\hat{\psi}_{12} - R_1\hat{\psi}^1_{mde})'\hat{\Lambda}^{-1}(\hat{\psi}_{12} - R_1\hat{\psi}^1_{mde}) \rightsquigarrow \chi^2_{k+2}.$$

The restrictions $\tau_{13} - \tau_{12} = \tau_{22}$, $\tau_{14} - \tau_{12} = \tau_{23}$ and $\beta_x^1 = \beta_x^2$ can be tested with T_N^1.

REFERENCES

Angrist, J.D. and J.S. Pischke, 2009, Mostly harmless econometrics, Princeton University Press.

Börsch-Supan, A., 2013, Survey of health, ageing and retirement in Europe (SHARE) Wave 1, Release version: 2.6.0., SHARE-ERIC, Data set, DOI 10.6103/SHARE.w1.260.

Börsch-Supan, A., A. Brugiavini, H. Jürges, J. Mackenbach, J. Siegrist and G. Weber, 2005, Health, ageing and retirement in Europe—first results from the Survey of Health, Ageing and Retirement in Europe, Mannheim Research Institute for the Economics of Aging (MEA).

Börsch-Supan, A. and H. Jürges, 2005, The survey of health, ageing and retirement in Europe—methodology, Mannheim Research Institute for the Economics of Aging (MEA).

Börsch-Supan, A., M. Brandt, C. Hunkler, T. Kneip, J. Korbmacher, F. Malter, B. Schaan, S. Stuck and S. Zuber, 2013, Data Resource Profile: The survey of health, ageing and retirement in Europe (SHARE), International Journal of Epidemiology 42, 992-1001.

Crossley, T.F. and S. Kennedy, 2002, The reliability of self-assessed health status, Journal of Health Economics 21, 643-658.

Greene, W., M.N. Harris, B. Hollingsworth and T.A. Weterings, 2014, Heterogeneity in ordered choice models: a review with applications to self-assessed health, Journal of Economic Surveys 28, 109-133.

Jayachandran, S., A. Lleras-Muney and K.V. Smith, 2010, Modern medicine and the twentieth century decline in mortality: evidence on the impact of sulfa drugs, American Economic Journal: Applied Economics 2, 118-146.

Jürges, H., 2007, True health vs response styles: exploring cross-country differences in self-reported health, Health Economics 16, 163-178.

Jürges, H., M. Avendano and J.P. Mackenbach, 2008, Are different measures of self-rated health comparable? An assessment in five European countries, European Journal of Epidemiology 23, 773-781.

Lee, M.J., 2010, Micro-econometrics: methods of moments and limited dependent variables, Springer.

Lee, M.J., 2015, Panel conditional and multinomial logit estimators, in The Oxford Handbook of Panel Data, 202-232, edited by B. Baltagi, Oxford University Press.

Lee, M.J., 2016a, Generalized difference in differences with panel data and least squares estimator, Sociological Methods and Research 45, 134-157.

Lee, M.J., 2016b, Matching, regression discontinuity, difference in differences, and beyond, Oxford University Press.

Lindeboom, M. and E. van Doorslaer, 2004, Cut-point shift and index shift in self-reported health, Journal of Health Economics 23,1083-1099.

Lumsdaine, R.L. and A. Exterkate, 2013, How survey design affects self-assessed health responses in the Survey of Health, Ageing, and Retirement in Europe (SHARE), European Economic Review 63, 299-307.

Moors, G., N.D. Kieruj and J.K. Vermunt, 2014, The effect of labeling and numbering of response scales on the likelihood of response bias, Sociological Methodology 44, 369-399.

Weijters, B., E. Cabooter and N. Schillewaert, 2010, The effect of rating scale format on response styles: the number of response categories and response category labels, International Journal of Research in Marketing 27, 236-247.