Quality & Quantity **33:** 203–213, 1999.

Modeling Item Nonresponse in Questionnaires

PHILIP HANS FRANSES[1,⋆], IRMA GELUK[2] and PAUL VAN HOMELEN[2]
[1]*Econometric Institute, Erasmus University Rotterdam, PO Box 1738, NL-3000 DR Rotterdam, The Netherlands;* [2]*Institute for Research and Investment Services, ROBECO–RaboBank*

Abstract. The statistical analysis of empirical questionnaire data can be hampered by the fact that not all questions are answered by all individuals. In this paper we propose a simple practical method to deal with such item nonresponse in case of ordinal questionnaire data, where we assume that item nonresponse is caused by an incomplete set of answers between which the individuals are supposed to choose. Our statistical method is based on extending the ordinal regression model with an additional category for nonresponse, and on investigating whether this extended model describes and forecasts the data well. We illustrate our approach for two questions from a questionnaire held amongst a sample of clients of a financial investment company.

Key words: ordered regression, item nonresponse.

1. Introduction and Motivation

Attitude questionnaires often concern a set of Q questions, which can be answered by A answers. Frequently, Q can be as large as 25, and A often takes the value of 3, 4, 5 or 7. For each question, N individuals are asked to mark one of the A possible answers. Sometimes these answers correspond with degrees of agreement, that is, for example, when $A = 5$, answer 1 corresponds with "strongly agree" and answer 5 with "strongly disagree' In other cases the answers can correspond with explicitly stated attitudes, that is, for example, when $A = 3$, answer 1 corresponds with "I invest in those products which are advised to me by a financial consultant" and answer 3 with "I actively follow international financial developments and I make my own choices". Our application in Section 4 deals with questionnaire data of the latter type, but our statistical method can also be applied to data of the first type. The answers are often measured on an ordinal scale, that is, one can rank the answers from 1 to A with A having a higher rank. For example, in the $A = 3$ example above, answer 3 corresponds with an active investor, while answer 1 matches with an inactive investor.

When Q becomes large, it is frequently encountered in practice that not all N individuals give answers to all Q questions. Hence, one encounters missing data. Additionally, it is often observed that there is item nonresponse, see Little

⋆ Author for correspondence. This paper was mainly written while the first author was enjoying the kind hospitality of ROBECO, Rotterdam.

and Schenker (1995) for the terminology. This means for example that N_1 ($\in N$) individuals answer questions 1, 2, ..., $Q-1$, while N_2 ($\in N$, and $N_1 \neq N_2$) individuals answer questions 2, 3, ..., Q. In case of such item response, one thus has $N_1, \cdots, N_Q$ individuals who give answers to the questions 1 to Q. One may now decide to consider the cross-sectional sample $N_R = N_1 \cap N_2 \cap \cdots \cap N_Q$, but this sample can become quite small. Hence, deleting all individuals who do not respond to at least one question may lead to a loss of information, which is possibly relevant to other questions. This is particularly relevant in case the missing data are not missing at random, see, for example, Little and Rubin (1989) and Little and Schenker (1995).

In this paper we propose a statistical method which can handle partial nonresponse, by explicitly modeling the missing data themselves. We assume the availability of N complete or partially complete questionnaires concerning Q questions with A possible answers, and we assume knowledge of K characteristics of these individuals. The characteristics serve as explanatory variables for the individual response. To save space, we set $Q = 1$, that is, our method can be applied to one question at a time. Extensions to jointly evaluating Q questions (using for example Principle Component Analysis) is postponed for further research. Our approach is based on the application of the Ordered Regression Model [ORM] to analyze response behavior, see McKelvey and Zavoina (1975) for an introduction to this model and Long (1997) for a recent survey. Other important references are McCullagh (1980) and Winship and Mare (1984). The variable to be explained in this model is ordinal, that is 1, 2, ..., A, where answer A is ranked higher (on a certain scale) than answers $A-1$, $A-2$, and so on. Basically, our method amounts to examining whether the dependent ordinal variable is better measured as an ordinal variable with $A+1$ values, where the extra category corresponds with the nonresponse. Since the additional category can be located in between answer a and $a+1$, with $a = 1, 2, \ldots, A-1$, or below or above answers 1 and A, respectively, we need an empirical specification method to help to make decisions in practice.

The outline of our paper is as follows. In Section 2, we discuss some of the main aspects of an Ordered Regression Model. In Section 3, we present our method to deal with item nonresponse, and we put forward a specification strategy that can be used in practice. In Section 4, we apply our method to 2 questions from an actual survey amongst about 25000 clients of a financial investment company. In Section 5, we conclude our paper with some remarks.

2. Ordered Regression Model

In this section, we review the key aspects of an ORM. For illustrative purposes, and without loss of generalization, we set $A = 3$ throughout this paper, also since it corresponds with our applications in Section 4. We also use a sample question which comes close to those in Section 4.

Consider the structural model

$$\begin{aligned} y_i^* &= \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_K x_{K,i} + \epsilon_i, \\ &= X_i\beta + \varepsilon_i, \end{aligned} \tag{1}$$

where $i = 1, 2, \cdots, N$. The ε_i is an error term, and the x_1 to x_K are variables which measure the K individual characteristics. The y_i^* variable is a so-called latent variable as it is not observed directly.

For illustration, suppose that y_i^* measures "activity", where larger values of y_i^* correspond with higher activity, and suppose that an active investor is someone who regularly changes his or her investment portfolio. As the y_i^* variable is not observed directly, one can obtain information on y_i^* through a questionnaire involving ordinal answers. For example, one may ask individuals to mark the answer which most closely matches with their attitude. In this case, sample answers are

$y_i = 1$: "I never change my portfolio",
$y_i = 2$: "I adjust my portfolio only when important economic events take place",
$y_i = 3$: "I regularly change my portfolio".

The respondent is asked to convey agreement with answer 1, 2 or 3. The answers y_i can now be linked to the unobserved y_i^* according to the measurement equation

$$\begin{aligned} y_i &= 1 \quad \text{if } -\infty \le y_i^* < \tau_1, \\ y_i &= 2 \quad \text{if } \tau_1 \le y_i^* < \tau_2, \\ y_i &= 3 \quad \text{if } \tau_2 \le y_i^* < \infty. \end{aligned} \tag{2}$$

Model (1)–(2) is called an ordinal regression model [ORM], see Long (1997, Chapter 5) and McKelvey and Zavoina (1975), amongst others. The τ_1 and τ_2 parameters are called thresholds.

Given (1) and (2), the probability of the observed value y_1, conditional on the explanatory variables, is

$$\begin{aligned} \Pr(y_i = 1 \mid X_i) &= \Pr(-\infty \le y_i^* < \tau_1 \mid X_i) \\ &= \Pr(-\infty - X_i\beta \le \varepsilon_i < \tau_1 - X_i\beta \mid X_i) \\ &= \Pr(\varepsilon_i < \tau_1 - X_i\beta \mid X_i) - \Pr(\varepsilon_i \le -\infty - X_i\beta \mid X_i) \\ &= F(\tau_1 - X_i\beta) - 0 \\ &= F(\tau_1 - X_i\beta), \end{aligned} \tag{3}$$

where F is a cumulative distribution function [cdf]. Similarly,

$$\Pr(y_i = 2 \mid X_i) = F(\tau_2 - X_i\beta) - F(\tau_1 - X_i\beta) \tag{4}$$

and

$$\begin{aligned} \Pr(y_i = 3 \mid X_i) &= 1 - \Pr(y_i = 1 \mid X_i) - \Pr(y_i = 2 \mid X_i) \\ &= 1 - F(\tau_2 - X_i\beta). \end{aligned} \tag{5}$$

When F is the normal cdf, one has an ordered probit model, and when F refers to the logistic distribution, the model is called the ordered logit model.

The parameters in (1)–(2) are only identified when either β_0 is set equal to zero, or when τ_1 or τ_2 equals zero. In our application below we will be interested in the values of τ_1 and τ_2, and therefore we will choose to set $\beta_0 = 0$. The variance of $\varepsilon_i \mid X_i$ is $\pi^2/3$ in the ordered probit model, and 1 in the ordered logit model.

The parameters (β, τ_1 and τ_2) can be estimated by maximum likelihood [ML]. The relevant expressions for the first and second order conditions are given in Maddala (1983: 48–49). Pratt (1981) proves that the likelihood is concave, and hence that ML estimation routine will converge.

3. Modeling Item Nonresponse

The ORM in (1)–(2) assumes that the respondents themselves are aware of their y_i^* value and that they also know the values τ_1 and τ_2, which lead them to choose either y_i = 1, 2 or 3. It is however possible that an individual finds his or hers opinion or attitude not to be amongst the given set of answers. In turn, this leads to missing data. For example, it may be that y_i^* is better classified into an ordinal variable with four instead of three answer categories. If this fourth option is not included, the respondents can either opt for one of the given attitudes, or decide not to answer. In the first case the researcher faces the problem that it is difficult (if not impossible) to disentangle whether respondents choose $y_i = 1, 2$ or 3, because they really want to, or that they choose so because there are no better alternatives.

In case an individual decides not to answer, which we will denote as $y_i = 0$, it may be that he or she *does* give an answer, albeit outside the scope of the three given answers. If many individuals do so, the questionnaire obviously excludes an important answer category, which means that the missing data are not missing at random. Statistically speaking, the ORM parameter estimates of τ_1 and τ_2 as well as of β will be biased. It therefore makes sense to focus attention on the additional answer category

$$y_i = 0 \qquad \text{“No response”}.$$

In this paper we assume knowledge of the individual characteristics X_i for all individuals, including those who do not respond to one or more questions. If these are not observed, one has to resort to a potentially complicated model as the censored ORM. We leave an analysis of this model for further research.

The practical question is now how we can incorporate the $y_i = 0$ category into the ORM. There are four possibilities, i.e., the answers can be arranged as {0, 1, 2, 3}, {1, 0, 2, 3}, {1, 2, 0, 3} or {1, 2, 3, 0}. The question is which of the four sequences should be selected. To abstain from variable selection issues, we assume that the same K variables are used throughout, i.e., all models have the same K explanatory variables on the right hand side.

As the first step, we recommend to consider a binary regression model [BRM] for

$$\begin{aligned} y_i^{**} = 0: &\quad \text{when } y_i = 0, \\ y_i^{**} = 1: &\quad \text{when } y_i = 1, 2 \text{ or } 3, \end{aligned}$$

with X_i as the set of explanatory variables. When no variables amongst the X_i variables are relevant (according to a priori specified criteria), the BRM is not helpful to discriminate between response and nonresponse. These criteria can be based on t-ratios of the parameter, but also on pseudo-R^2 measures, see Windmeijer (1995) for a recent survey. In that case, one may just as well delete the nonrespondents for further analysis.

However, if nonrespondents do differ from respondents, we advocate as the second step the use of three BRMs for

$$\begin{aligned} y_i^{**} = 0: &\quad \text{No response} \\ y_i^{**} = 1: &\quad \text{Response is attitude } j, \end{aligned}$$

where j can be 1, 2, and 3. For each value of j, the outcome can be that the no response category does or does not differ from answer j. If there is *no* difference, one can add the nonrespondents to the jth answer category. If there is a difference, one can choose between the rank order $\{0, j\}$ or $\{j, 0\}$, for $j = 1, 2,$ or 3.

In practice, it may not be easy to make a choice between the rankings $\{0, j\}$ and $\{j, 0\}$. One option is to make the choice dependent on the expected sign of the effect of one or more explanatory variables. For example, if the unobserved variable y_i^* is supposed to measure "experience", one may expect that the variable "age" has a positive impact. When the estimated parameter for "age" is positive in the BRMs for 0 versus 2 and 3, but negative in the BRM for answer 0 versus 1, one may hypothesize the plausibility of the sequence $\{1, 0, 2, 3\}$. Another option is to consider the average values of the explanatory variables in the four categories and see whether a certain pattern can be observed.

Of course, as a third option, one may also decide to estimate all four possible models. In that case, one considers the measurement equation as

$$\begin{aligned} y_i = A &\quad \text{if } -\infty \leq y_i^* < \lambda_1, \\ y_i = B &\quad \text{if } \lambda_1 \leq y_i^* < \lambda_2, \\ y_i = C &\quad \text{if } \lambda_2 \leq y_i^* < \lambda_3, \\ y_i = D &\quad \text{if } \lambda_3 \leq y_i^* < \infty, \end{aligned} \tag{6}$$

which extends (2) with an additional answer category, and where $\{A, B, C, D\}$ can be $\{0, 1, 2, 3\}$, $\{1, 0, 2, 3\}$, $\{1, 2, 0, 3\}$ or $\{1, 2, 3, 0\}$. The log-likelihood values of the four models can be compared, and since these models all contain the same K explanatory variables, one can select the model with the highest log-likelihood value. Finally, one can examine whether the confidence intervals around $\hat{\lambda}_1$, $\hat{\lambda}_2$ and $\hat{\lambda}_3$ show overlap. If so, one can test whether restrictions such as $\lambda_1 = \lambda_2$ hold.

Before we turn to our applications in the next section, we mention that a practitioner should exercise care when estimating the parameters in (in fact, any) ORM when the thresholds (τ or λ) may be close to each other. For example, strictly speaking, it should hold for (1) that $\hat{\tau}_2 > \hat{\tau}_1$. However, when the underlying τ_1 and τ_2 values are very close, it can occur in one of the iteration steps that $\hat{\tau}_1$ exceeds $\hat{\tau}_2$ (because the value of the gradient is relatively large). Negative probabilities can then occur, and most estimation programs rightfully collapse. This can be taken as an indication that the explanatory variables are not helpful to distinguish between two answer categories, and the ORM should be modified by combining these categories.

4. Applications

We will illustrate some of the suggested approaches for two examples. We use a data set on a questionnaire that is held amongst 24989 clients of the ROBECO investment company. For reasons of confidentiality, we assign different names to most variables, and we scale the observations of some variables using a (here: known but unspecified) monotone transformation. Also, the questionnaire contains many more questions, and we focus only on a very small subset of these.

The first question is assumed to deal with the latent variable "activity" or "dynamic behavior". Individuals are asked whether they agree with one of the following statements:

$y_i = 1$: "I am only interested in the long run: I seldom change my portfolio"
$y_i = 2$: "Important economic events may force me to change my portfolio"
$y_i = 3$: "I regularly change my portfolio".

Nonresponse to this question concerns 856 individuals (which is about 3.4%). Useful explanatory variables for "activity" are hypothesized to be

$x_{1,i}$: Age,
$x_{2,i}$: Number of changes in portfolio (in the past year),
$x_{3,i}$: Number of mutations (in the past year),
$x_{4,i} - x_{7,i}$: Investment in products of type 1, 2, 3 or 4 (dummy variables).

The second question deals with "geographical interest". Individuals can agree with one of the following statements:

$y_i = 1$: "I have no interest in other countries",
$y_i = 2$: "Other countries do interest me because of higher returns to be obtained",
$y_i = 3$: "Investing in other countries allows me to diversify my risk".

Nonresponse to this question concerns 1751 individuals (which is about 6.0%). Useful explanatory variables for "geographical interest" are hypothesized to be

$x_{1,i}$: Age,
$x_{8,i}$: Number of years of relationship with company,
$x_{4,i} - x_{5,i}$: Investment in products of type 1 or 2 (dummy variables),
$x_{9,i}$: Investment in product of type 5 (dummy variable).

where investment in product of type 2 provides information on the actual geographic activity of the individuals. Hence, a positive sign of the corresponding parameter is expected.

To be able to investigate the out-of-sample forecasting performance of the models, we randomly select 16660 individuals (which is about two-third of the sample), and we leave 8329 individuals for forecast evaluation. In a final step, we will compare the forecasting performance with that of models where the nonrespondents are simply excluded. In that case, we have less individuals in our estimation sample, and also less in our evaluation sample, since we delete nonrespondents from 16660 and 8329 individuals, respectively. Given the substantial size of all samples, we assume that we can safely compare the parameter estimates and the out-of-sample results.

For both questions we find that a binary regression model for nonresponse versus any of the three responses a highly relevant (based on t-ratios which indicate significance at the 0.5% level). Hence, it seems that we cannot dismiss the nonresponse category. This confirms that the item nonresponse corresponds with missing data which are not missing at random, see Little and Schenker (1995).

4.1. ACTIVITY

When we consider the three pairwise BRMs for the question on "activity", we find no pair $\{0, j\}$ for which there are only insignificant variables. Hence, we cannot set nonresponse equal to answer 1, 2 or 3. In fact, we find that several variables (age, number of changes in portfolio, products 1 and 2) have a significant effect, and that this effect is consistently negative or positive across the three BRMs. This implies that the nonresponse category should be placed before or after the sequence $\{1, 2, 3\}$. As the age variable appears to have a negative impact, and the average age in the group of nonrespondents is highest amongst all answers, we hypothesize that the ORM for "activity" (when we include the nonrespondents) contains the four answers ranked as $\{0, 1, 2, 3\}$. The estimation results, obtained by maximum likelihood, for (1) with (6) are

$$\begin{aligned} y_i^* = & \underset{(0.064)}{-0.498x_{1,i}} + \underset{(0.021)}{0.640x_{2,i}} + \underset{(0.014)}{0.089x_{3,i}} - \underset{(0.027)}{0.254x_{4,i}} + \\ & + \underset{(0.028)}{0.425x_{5,i}} - \underset{(0.029)}{0.450x_{6,i}} - \underset{(0.045)}{0.304x_{7,i}}, \end{aligned} \tag{7}$$

where estimated standard errors are given in parentheses. Furthermore, we have

$$\underset{(0.053)}{\hat{\lambda}_1 = -3.543}, \ \underset{(0.040)}{\hat{\lambda}_2 = 0.533}, \text{ and } \underset{(0.048)}{\hat{\lambda}_3 = 2.656}.$$

Clearly, the estimated thresholds λ_1, λ_2 and λ_3 are significantly different from each other.

When we delete nonrespondents from the estimation sample of 16660, we are left with 16085 individuals. To examine the effects of deleting these nonrespondents, we estimate the parameters in the ORM (1) with (2). We obtain

$$\begin{aligned} y_i^* = & \underset{(0.066)}{-0.357x_{1,i}} + \underset{(0.022)}{0.667x_{2,i}} + \underset{(0.012)}{0.102x_{3,i}} - \underset{(0.031)}{0.387x_{4,i}} + \\ & + \underset{(0.038)}{0.421x_{5,i}} - \underset{(0.033)}{0.580x_{6,i}} - \underset{(0.049)}{0.403x_{7,i}}, \end{aligned} \tag{8}$$

where estimated standard errors are given in parentheses. Furthermore, we find

$$\underset{(0.042)}{\hat{\tau}_1 = 0.570}, \text{ and } \underset{(0.049)}{\hat{\tau}_2 = 2.724}.$$

Comparing (7) with (8), we observe that there are several parameters which obtain slightly different values across models, and that the "age" variable (x_1) shows the largest difference. The thresholds $\hat{\tau}_1$ and $\hat{\tau}_2$ do not seem to differ much from $\hat{\lambda}_2$ and $\hat{\lambda}_3$, thereby clearly indicating that $\{0, 1, 2, 3\}$ is an appropriate ranking.

In Table I we report the out-of-sample classification of models (7) and (8). From the upper left matrix, we can see that the explanatory variables are not helpful to forecast who is a nonrespondent or who gives answer 3. Defining the hit rate as the sum of the diagonal elements of this matrix, we notice that the out-of-sample hit rate of model (7) is 0.63. For model (8) the number of individuals in the hold-out sample is 8048. The upper right matrix in Table I shows that there is not much difference between (7) and (8) with respect to forecasting. The hit rate is approximately the same, and so are the percentages of misclassification.

Table I. Out-of-sample forecasting performance (classification) of various ORMs (cells contain frequencies)

	Observed answers							
	Model with nonresponse				Model without nonresponse			
Model prediction	0	1	2	3	0	1	2	3
Activity								
0	0	0	0	0	NA	NA	NA	NA
1	0.03	0.58	0.24	0.04	NA	0.58	0.24	0.04
2	0	0.04	0.05	0.02	NA	0.06	0.06	0.02
3	0	0	0	0	NA	0	0	0
Geographical interest								
0	0.02	0.09	0.06	0.05	NA	NA	NA	NA
1	0	0.01	0	0	NA	0	0	0
2	0.05	0.19	0.30	0.23	NA	0.31	0.38	0.30
3	0	0	0	0	NA	0	0	0

NA: Not available.
Clockwise starting with the upper left matrix, the corresponding models are (7), (8), (10) and (9). The ORM for activity, which includes nonresponse, has the ranking {0, 1, 2, 3} for the answers (where 0 corresponds with nonresponse), while for geographical interest this ranking is {1, 0, 2, 3}.

5. Geographical Interest

For the second example question, on geographical interest, we find in the pairwise BRMs that not many parameters are significant, especially not in the BRM for nonresponse versus answer 1. An important variable in the latter BRM is the dummy variable for product of type 2. In the BRM for 0 versus 1, it obtains a negative sign, while in the BRM for nonresponse versus answer 2 it obtains the expected positive sign. Hence, for this sample question, we hypothesize that {1, 0, 2, 3} is a plausible categorization. To verify this conjecture we also estimate the parameters in the other three ORMs as in (6), and compare the values of the log-likelihoods.

For the ranking {1, 0, 2, 3}, we obtain the following parameter estimates:

$$\underset{}{y_i^*} = \underset{(0.058)}{-0.851x_{1,i}} \; \underset{(0.049)}{+0.258x_{8,i}} \; \underset{(0.030)}{+0.449x_{4,i}} \; \underset{(0.031)}{+0.365x_{5,i}} \; \underset{(0.035)}{+0.225x_{9,i}}, \tag{9}$$

where estimated standard errors are given in parentheses. Furthermore, we find

$$\underset{(0.037)}{\hat{\lambda}_1 = -0.875}, \quad \underset{(0.037)}{\hat{\lambda}_2 = -0.539}, \quad \text{and} \quad \underset{(0.038)}{\hat{\lambda}_3 = 1.013}.$$

The estimated thresholds λ_1, λ_2 and λ_3 are clearly different from each other. The log-likelihood of this model is -1.249. The log-likelihood values of the ORMs

for the rankings $\{0, 1, 2, 3\}$, $\{1, 2, 0, 3\}$ and $\{1, 2, 3, 0\}$ are -1.250, -1.257, and -1.260, respectively. Hence, it seems that (9) is to be preferred. In words, this means that nonrespondents to this question prefer an answer which is in between answer 1 and 2.

When we delete the nonrespondents in the estimation sample, we are left with 15499 individuals. An ORM for the answers 1, 2 and 3, with the explanatory variables as in (9) yields

$$\begin{matrix} y_i^* = & -0.726x_{1,i} & +0.254x_{8,i} & +0.454x_{4,i} & +0.330x_{5,i} & +0.242x_{9,i}, \\ & (0.063) & (0.052) & (0.030) & (0.033) & (0.036) \end{matrix} \tag{10}$$

and the thresholds are estimated as

$$\begin{matrix} \hat{\tau}_1 = -1.006, & \text{and} & \hat{\tau}_2 = 0.661. \\ (0.038) & & (0.039) \end{matrix}$$

We observe that again the age variable obtains a different parameter value. Additionally, we notice that the estimated thresholds vary substantially across (9) and (10). This indicates that the in-sample classification of individuals is quite different across the two cases where nonrespondents are included or not.

In the bottom panel of Table I, we compare the out-of-sample forecasting (classification) performance of models (9) and (10). We find that the hit rate for (10) (based on 7739 individuals) is better, but that this seems due to the fact that (10) almost only predicts answer category 2. Model (9) assigns too many individuals to the nonresponse category.

6. Conclusion

In this paper we proposed a simple method to deal with item nonresponse in questionnaires. For this purpose, we use an extended ordered regression model, which includes an answer category that corresponds with nonresponse. Our applications to two questions (which were part of a large-scale survey) showed that parameter estimates can differ across models, and that out-of-sample forecasting performance can also differ (although in our case not that much).

In this paper we abstained from analyzing two issues that certainly deserve further attention. The first is to show through extensive simulations the effects of neglecting nonresponse on parameter estimates and out-of-sample classification in the case of the ordered regression model. The second is concerned with the variable selection issue. How to select the most appropriate explanatory variables in each model in each round of the specification strategy is left for further research.

References

Little, R.A.J. & Rubin, D.B. (1989). The analysis of social science data with missing values. *Sociological Methods and Research* 18: 292–326.

Little, R.A.J. & Schenker, N. (1995). Missing data. In: G. Arminger, C. C. Clogg & M. E. Sobel (eds), *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. New York: Plenum Press.

Long, J. Scott (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage Publications.

Maddala, George S. (1983). *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

McCullagh, Peter (1980). Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society B* 42: 109–142.

McKelvey, Richard D. & Zavoina, William (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology* 4: 103–120.

Pratt, John W. (1981). Concavity of the log likelihood. *Journal of the American Statistical Association* 76: 103–106.

Windmeijer, Frank A.G. (1995). Goodness-of-fit measures in binary responses models. *Econometric Reviews* 14: 101–116.

Winship, Christopher and Mare, Robert D. (1984). Regression models with ordinal variables. *American Sociological Review* 49: 512–525.