

A Note on the Usefulness of the Behavioural Rasch Selection Model for Causal Inference in the Social Sciences

Matthew P. Rabbitt¹

Economic Research Service, U.S. Department of Agriculture, 355 E Street, SW,
Washington, DC, 20024-3221, USA

E-mail: matthew.rabbitt@ers.usda.gov

Abstract. Social scientists are often interested in examining causal relationships where the outcome of interest is represented by an intangible concept, such as an individual's well-being or ability. Estimating causal relationships in this scenario is particularly challenging because the social scientist must rely on measurement models to measure individual's properties or attributes and then address issues related to survey data, such as omitted variables. In this paper, the usefulness of the recently proposed behavioural Rasch selection model is explored using a series of Monte Carlo experiments. The behavioural Rasch selection model is particularly useful for these types of applications because it is capable of estimating the causal effect of a binary treatment effect on an outcome that is represented by an intangible concept using cross-sectional data. Other methodology typically relies on summary measures from measurement models that require additional assumptions, some of which make these approaches less efficient. Recommendations for application of the behavioural Rasch selection model are made based on results from the Monte Carlo experiments.

1. Introduction

The social sciences are often interested in measuring properties or attributes of individuals with no clearly defined physical referent (intangible concepts), such as attitudes, values, beliefs, well-being, health, or ability. For the measurement of these concepts to be valid and reliable, it must be supported by both social science and measurement theory. Social science theory allows the researcher to clearly define the concept that is to be measured and develop an instrument while measurement theory provides the foundation for establishing and validating the instrument's corresponding scale. Many social scientists rely on item response theory (IRT) models, such as the Rasch model [3] to develop and validate their scales.

Researchers in the social and physical sciences undertake the measurement of concepts with and without physical referents to confirm, reject, and refine research hypotheses. Findings from the testing of hypotheses advance theory and inform policymakers. Properly testing a hypothesis requires the careful design of an experiment, whether it is in the laboratory or using a social survey. Physical scientists often have the advantage of using randomized experiments with carefully constructed treatment and control groups. While randomized experiments are arguably the best way to test a hypothesis, they are often costly, time consuming, and for social scientists, one must consider the ethical

¹ The views expressed in this paper are those of the author and do not necessarily reflect those of the Economic Research Service or the U.S. Department of Agriculture.



issues associated with human subjects. Alternatively, researchers may rely on natural experiments or statistical methodology to identify causal relationships when randomized experiments are not feasible.

Estimating causal relationships is particularly challenging when using survey data, as key characteristics essential to social science models may be omitted. When this occurs bias is introduced into the analysis, potentially contributing to counterintuitive or misleading findings. This occurs because the treatment and control groups are not randomized, ensuring the sample is not representative of the population intended for analysis. Self-selection bias ensues because individuals select into treatment and control groups based on their observable and unobservable characteristics.

This paper examines the usefulness of a recently proposed model, the behavioural Rasch selection model [2], by conducting a Monte Carlo study of the model's properties. The Monte Carlo study undertaken here is the first to examine the behavioural Rasch selection model and will be used to provide guidance for researchers interested in implementing the model in their own analyses. The behavioural Rasch selection model is particularly useful for social science researchers because it is capable of estimating the causal effect of a binary treatment effect on an outcome that is represented by an intangible concept using cross-sectional data. The behavioural Rasch selection model maintains all of the Rasch model assumptions while incorporating a multivariate behavioural component and addressing selection on contemporaneous unobservable characteristics. Identification of the model is achieved using exclusion restrictions in the form of instrumental variables.

2. Methodology

IRT models, such as the Rasch model, are particularly useful for measuring an individual's properties or attributes when no clear physical referent exists by formulating them as latent traits. For the purposes of this paper, the Rasch model is framed to measure an individual's ability; however, it could just as easily be altered to measure an individual's well-being, health, food security, or attitudes and beliefs. The Rasch measurement model assumes individual i 's underlying unobservable (latent) index of ability is θ_i , with the property that higher values of the index correspond to greater ability. Assuming there are N individuals administered J binary questions (e.g., yes or no.) that capture different levels of ability, then the probability that the individual affirms the j^{th} question is

$$P(Y_{ij} = 1 | \theta_i, \delta_j) = \frac{\exp(\theta_i - \delta_j)}{1 + \exp(\theta_i - \delta_j)}, i = 1, \dots, N; j = 1, \dots, J \quad (1)$$

where $\exp(\cdot)$ is the exponential function and δ_j is a threshold or "calibration" parameter. The model further assumes that the individual's responses to each question are independent, conditional on θ_i , and that the factor loadings (discrimination parameters) are constrained to be equal across all items and normalized to one.

An attractive property of the Rasch model for statistical inference is that if all individuals' responses to the questions follow this model and if the individuals answer all J items, then their latent ability can be ranked and compared using simple counts of the number of affirmed items. Formally, the count of affirmed items, commonly referred to as the raw score, is a sufficient statistic for θ_i .

Summary measures have proven very useful for social scientists interested in examining the relationships between an intangible concept and other variables. Rasch scores may be used to estimate continuous regression models, such as Ordinarily Least Squares (OLS). However, for these models to be identified, all observations with extreme values (all affirmative or negative responses) to an instrument's questions (the majority of respondents for many instruments and surveys) may have to be dropped from the analysis sample prior to estimation depending on how the Rasch model is estimated (e.g., joint or conditional maximum likelihood.). Alternatively, the raw score may be used to estimate count data regression models for samples of individuals with no missing responses to an instrument's questions. Missing responses may also be imputed, but the accuracy of these imputations must be assessed.

The Rasch model can also be extended to incorporate a multivariate behavioural component for cases in which social scientists are interested in examining the relationship between the latent trait and other variables. Specifically, one may assume that individual i 's latent ability, θ_i , depends on an observed treatment (T_i), control characteristics (X_i), and an unobserved variable (e_i) such that

$$\theta_i = \beta_T T_i + \beta_X X_i + e_i. \quad (2)$$

It is common practice to assume that $e_i \sim N(0, \sigma^2)$ for identification purposes. Combining specifications (1) and (2) yields the behavioural Rasch model² [5].

Using the behavioural Rasch model has several advantages over summary measures for this type of inference. Methodologically, the behavioural Rasch model is similar to a regression model where the dependent variable is the Rasch score. The main difference is that the behavioural Rasch model estimates specifications (1) and (2) jointly, while regression models involving the Rasch score estimate these parameters in two steps. Regression models involving Rasch scores may also draw different samples, as individuals with extreme scores may be excluded, depending on the methodology used to produce the Rasch scores. The behavioural Rasch model, on the other hand, can incorporate individuals with extreme scores. Another advantage of the behavioural Rasch model is its ability to impute missing values by virtue of the Rasch measurement model's properties.

Estimating the causal effect of a treatment on an intangible outcome is further complicated by the endogeneity of the decision to select into the treatment group (to be treated). Behavioural Rasch models are able to control for selection on an individual's observable characteristics, but do not account for selection on an individual's unobservable characteristics. More formally, the individual's decision to select into the treatment group is

$$T_i = \begin{cases} 1 & \text{if } \alpha_X X_i + \alpha_Z Z_i + u_i > 0 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where T_i and X_i are defined above, Z_i is a set of instrumental variables, and u_i is an unobserved variable. For identification purposes, one may assume $u_i \sim N(0, 1)$ ³; however, it need not be. The resulting model is consistent with a probit model for the decision to select into the treatment group.

The endogeneity of the treatment can be formalized, as in [2] and [4], by assuming the error-component in specification (2) can be decomposed into u_i and e_i^* , such that $e_i = \lambda u_i + e_i^*$. Thus, the individual's latent ability index can be respecified as

$$\theta_i^* = \beta_T T_i + \beta_X X_i + \lambda u_i + e_i^* \quad (4)$$

where λ is an unknown factor loading (selection) parameter to be estimated, and e_i^* represents the new unobserved variable after controlling for observed and unobserved variables (heterogeneity). It is assumed that $e_i^* \sim N(0, \eta^2)$. Correlation between the treatment variable (T_i) and the latent ability is generated through the factor loading parameter, λ . If λ is estimated to be nonzero, u_i influences the individual's selection into the treatment group and the likelihood of affirming a question, rendering the baseline behavioural Rasch model inconsistent and biased.

An estimator for the behavioural Rasch selection model, which combines specifications (1)-(3), was proposed by [2] and tested on survey data. In this paper, [2] derived the likelihood function for the behavioural Rasch selection model, which is as follows:

² The behavioral Rasch model is equivalent to the person-level explanatory Rasch model described in [5].

³ Alternatively, one may assume that u_i is time-invariant and estimate a model that combines specifications (1)-(4) using panel data methods. For an example of this approach, see [1].

$$L(\alpha, \beta, \delta, \eta^2, \lambda) = \prod_{i=1}^N \left\{ T_i \int_{-\alpha_X X_i - \alpha_Z Z_i}^{\infty} \int_{-\infty}^{\infty} \prod_{j=1}^J \left[\frac{\exp(q_{ij}(\theta_i^* - \delta_j))}{1 + \exp(q_{ij}(\theta_i^* - \delta_j))} \right] \frac{1}{\eta} \phi\left(\frac{e_i^*}{\eta}\right) de_i^* \varphi(u_i) du_i \right. \\ \left. + (1 - T_i) \int_{-\infty}^{-\alpha_X X_i - \alpha_Z Z_i} \int_{-\infty}^{\infty} \prod_{j=1}^J \left[\frac{\exp(q_{ij}(\theta_i^* - \delta_j))}{1 + \exp(q_{ij}(\theta_i^* - \delta_j))} \right] \frac{1}{\eta} \phi\left(\frac{e_i^*}{\eta}\right) de_i^* \varphi(u_i) du_i \right\} \quad (5)$$

where the ‘i’ subscript denotes the ith sample household (i = 1, ..., N). The proposed model allows for all of the parameters to be jointly estimated using full information maximum likelihood (FIML). It is useful to note that the factor loading parameter, λ, is a sufficient statistic for exogeneity of the treatment decision.

In the current paper, I estimate the likelihood function specified in (5) using the Stata code provided by [2]. The Stata program approximates the two-dimensional integral using Gauss-Steen and Gauss-Hermite quadrature techniques for the first and second integrals, respectively. All models programs were executed using State 13 2-Core Multiprocessor statistical software.

3. Monte Carlo Study Design

The data generating process was designed to examine the performance of the behavioural Rasch selection model with samples of varying size (in terms of number of individuals). The instrument length was held constant across the samples to assess how many individuals would need to be surveyed in real-world data for reliable inference. Four separate experiments were performed using 500, 1,000, 5,000, and 10,000 observations and an instrument that consisted of 5 questions. For each sample size, 1,000 repetitions were completed and 8 quadrature points were used for each level of integration to approximate the integrals.

Data for each experiment were generated according the following assumptions.

1. $X_i \sim U(0,1)$
2. $Z_i \sim U(0,1)$
3. $T_i^* = \alpha_X X_i + \alpha_Z Z_i + u_i > 0 : u_i \sim N(0,1)$
4. $Y_{ij} = \frac{\exp(\beta_T T_i + \beta_X X_i + \lambda u_i + e_i^* - \delta_j)}{1 + \exp(\beta_T T_i + \beta_X X_i + \lambda u_i + e_i^* - \delta_j)} : e_i^* \sim N(0, \eta^2)$
5. $[\alpha_X, \alpha_Z, \alpha_{\text{Constant}}] = [1, 1, -1]$
6. $[\beta_{\text{Treatment}}, \beta_X, \beta_{\text{Constant}}, \delta_1, \delta_2, \delta_3, \delta_4, \delta_4, \eta^2, \lambda] = [1, 1, -1, -0.6, -0.3, 0.0, 0.3, 0.6, \exp(1), 1]$

4. Results

Estimates for the Monte Carlo experiments were obtained by estimating the behavioural Rasch selection model using simulated data generated by the procedure outlined above. Summary statistics for the experiments are reported in Table 1, and include each parameter’s average, standard deviation, and mean square error. The top, middle, and bottom panels contain summary information for the selection, outcome, and error-component equations, respectively. The columns report the true value and summary information, separately, for each of the experiments by sample size.

Estimates from the selection equation presented in the top panel of Table 1, which was estimated jointly with the other equations, performed very well, regardless of sample size. Average parameter values were close to their true values, even in smaller sample sizes, suggesting modest bias. For example, the percent bias in the smallest sample (N = 500) ranged from .03 to 1 percent, while bias in the largest sample (N = 10,000) ranged from 0.1 to 0.2 percent. As the sample size increases, the standard deviation of the selection equation parameters declines in magnitude.

Table 1. Monte Carlo Summary Statistics for the Behavioural Rasch Selection Model: Average, Standard Deviation, and Mean Square Error of the Coefficients.

	True Value	Sample Size			
		500	1,000	5,000	10,000
<u>Selection (Treatment) Equation</u>					
α_X	1.000	1.010 (0.206) [0.043]	0.999 (0.137) [0.019]	1.003 (0.067) [0.005]	1.001 (0.045) [0.002]
α_Z	1.000	0.997 (0.204) [0.041]	1.009 (0.145) [0.021]	1.000 (0.064) [0.004]	1.002 (0.044) [0.002]
α_{Constant}	-1.000	-1.006 (0.162) [0.026]	-1.007 (0.110) [0.012]	-1.002 (0.050) [0.003]	-1.002 (0.035) [0.001]
<u>Outcome Equation</u>					
$\beta_{\text{Treatment}}$	1.000	0.979 (0.933) [0.871]	1.031 (0.615) [0.379]	0.988 (0.268) [0.072]	0.999 (0.191) [0.037]
β_X	1.000	1.031 (0.541) [0.293]	0.996 (0.370) [0.137]	1.010 (0.154) [0.024]	1.003 (0.113) [0.013]
β_{Constant}	-1.000	-1.006 (0.364) [0.132]	-1.015 (0.241) [0.059]	-0.997 (0.109) [0.012]	-1.002 (0.077) [0.006]
δ_1	-0.600	-0.608 (0.175) [0.031]	-0.606 (0.124) [0.015]	-0.598 (0.056) [0.003]	-0.601 (0.038) [0.001]
δ_2	-0.300	-0.304 (0.174) [0.030]	-0.303 (0.121) [0.015]	-0.299 (0.055) [0.003]	-0.299 (0.037) [0.001]
δ_4	0.300	0.305 (0.170) [0.029]	0.296 (0.117) [0.014]	0.302 (0.056) [0.003]	0.299 (0.039) [0.002]
δ_5	0.600	0.605 (0.169) [0.029]	0.602 (0.120) [0.014]	0.602 (0.055) [0.003]	0.600 (0.039) [0.001]
<u>Error Components</u>					
$\ln(\eta^2)$	1.000	0.923 (0.269) [0.078]	0.977 (0.154) [0.024]	0.993 (0.064) [0.004]	0.995 (0.044) [0.002]
λ	1.000	1.027 (0.606) [0.368]	0.990 (0.402) [0.162]	1.008 (0.172) [0.029]	1.001 (0.125) [0.015]

Note: Models estimated using simulated data with varying number of individuals and a constant instrument length of 5 items. Standard deviations are in parenthesis and mean square errors are in brackets.

The primary coefficient of interest, an estimate of the treatment’s effect, is shown in the middle panel of Table 1. The behavioural Rasch selection model is able to obtain reliable estimates of this effect, even in smaller samples, albeit biased downwards (with the exception of the N = 1,000 experiment

where it is biased upwards). The model appears to perform best with samples that are greater than or equal to 5,000 observations. For the experiments with the two largest sample sizes, the percent bias ranges from -1.2 percent to -0.1 percent. Significant gains in terms of mean square error are also observed moving from a sample size of 1,000 to 5,000 or more observations.

The bottom panel of Table 1 contains estimates of the model's error components. Of particular interest are the parameter estimates for the factor loading, λ , which describes the degree of selection in one's sample. Estimates of the selection parameter are underestimated in the experiment with the smallest sample size and overestimated in the other experiments. The percentage bias for this parameter was relatively small across all experiments, ranging from 2.7 percent to 0.1 percent in the experiments with 500 and 10,000 observations, respectively. Once again, significant gains in terms of mean square error are observed for this parameter moving from a sample size of 1,000 to 5,000 or more observations. Estimates of the variance for the latent trait are reasonable in samples with 1,000 or more observations.

5. Conclusion

Social scientists are often interested in examining causal relationships for the purposes of informing theory or policy. Examination of these types of relationships become even more challenging when the outcome of interest is represented by an intangible concept. When intangible concepts are involved in research, the social scientist must appeal to measurement methodology, such as item response theory, to develop and validate a scale. After a scale is constructed, the social scientist may use various approaches to model the relationships of interest. In the current paper, I examine the usefulness of applying the recently proposed behavioural Rasch selection model to situations in which the researcher wishes to estimate the causal effect of a binary treatment effect on an outcome that is represented by an intangible concept using cross-sectional data.

Results from the Monte Carlo experiments suggest that the behavioural Rasch selection model is consistent and has modest bias in samples of size 500 or more. While the Monte Carlo experiments did not assess the effect of instrument length, the choice of an instrument with only 5 questions provides confidence that performance will only improve as the length of the instrument increases. While the behavioural Rasch selection model performed reasonably in all of the experiments, significant gains, in terms of mean square error, were observed for samples with 5,000 or more observations. Thus, it is recommended that the model be applied to samples with at least 5,000 observations. For those with smaller samples, careful attention should be paid to the statistical power they will have in their sample. In these cases it may be useful to conduct a power analysis.

Acknowledgements

The author would like to thank David C. Ribar at the Melbourne Institute of Applied Economic and Social Research for helpful comments.

References

- [1] Moffitt, R. A., & Ribar, D. C. (2016). Rasch Analyses of Very Low Food Security Among Households and Children in the Three City Study. *Southern Economic Journal*, 82(4), 1123-1146.
- [2] Rabbitt, M. P. (2013). Measuring the Effect of Supplemental Nutrition Assistance Program Participation on Food Insecurity Using a Behavioral Rasch Selection Model. Working Paper No. 13-20. University of North Carolina at Greensboro, Department of Economics.
- [3] Rasch, G. (1960). Studies in Mathematical Psychology: I. Probabilistic Models for Some Intelligence and Attainment Tests.
- [4] Terza, J. V. (2009). Parametric Nonlinear Regression with Endogenous Switching. *Econometric Reviews*, 28(6), 555-580.
- [5] Wilson, M., & De Boeck, P. (2004). Descriptive and Explanatory Item Response Models. *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*, 43-74.