Measurement 120 (2018) 193–205

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Causal inference with latent variables from the Rasch model as outcomes

Matthew P. Rabbitt

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

ARTICLE INFO

Keywords:
Causal inference
Food insecurity
Item response theory
Latent regression
Rasch model
Selection
SNAP

ABSTRACT

This article discusses and compares several methods for estimating the parameters of a latent regression model when one of the explanatory variables is an endogenous binary (treatment) variable. Traditional methods based on two-stage least squares and the Tobit selection model where the dependent variable is an estimate of the latent variable from the Rasch model are compared to the behavioral Rasch selection model. The properties of these methods are examined using simulated data and empirical examples are included to demonstrate the usefulness of the behavioral Rasch selection model for research in the social sciences. The simulations suggest the latent regression model parameters are more accurately and precisely estimated by the behavioral Rasch selection model than by two-stage least squares or the Tobit selection model. The empirical examples demonstrate the importance of addressing endogenous explanatory variables in latent regressions for Item Response Theory (IRT) models when estimating causal differences in the latent variable or examining differential item functioning.

1. Introduction

Inference in modern empirical research is often based on parameters from regression models of outcomes represented by a latent variable, such as a person's ability, health, food insecurity, or well-being, on a set of explanatory variables. Often these relationships are modeled indirectly using linear or censored regression models where the dependent variable is an estimate of the latent variable. Estimates of the latent variable can be obtained by combining item responses to an instrument using models from Item Response Theory (IRT; [36]), such as the Rasch model [26]. While this approach is straightforward, modeling the desired relationships between the latent variable and a set of explanatory variables directly using an IRT model offers greater precision and accuracy [9,38] because the measurement and latent regression (behavioral) models are jointly estimated, rather than estimated in two steps.

Early research on modeling latent regressions in IRT models focused on the validity of the distributional assumptions of the latent variable [2] and comparing the distribution of ability across groups [29]. Multivariate latent regressions models were later developed for the dichotomous Rasch model [18,19,20,38], polytomous Rasch model [39], and loglinear Rasch model [10]. For a comprehensive discussion of these models, see De Boeck and Wilson [13]. Even though these models rely on observational data to estimate the parameters of the latent regression, they do not address endogeneity in the latent regression model's explanatory variables. Endogeneity can occur in observational data because of omitted variables and measurement error. Failure to account for these sources of bias will render estimates of the latent regression model parameters biased and inconsistent. The behavioral Rasch selection model (BRSM; [22]) addresses endogenous binary variables in the latent regression model for the dichotomous Rasch model using an instrumental variables approach.

This article describes and compares several methods for estimating the parameters of a latent regression model with a binary endogenous variable. The methods considered include the BRSM, two-stage least squares (2SLS) and Tobit [34] selection model (TSM). The properties of these methods are compared using simulated item response data. Empirical examples are also considered to demonstrate the usefulness of the BRSM for estimating the causal effect of a binary endogenous variable on a latent variable, and correcting for the endogeneity of the group indicator variable in analyses of differential item functioning (DIF). The simulated data and empirical analyses make two notable contributions. First, this is the first time simulated data has been used to compare the BRSM, 2SLS, and TSM under the assumption that the data is generated with an endogenous binary variable. Second, the empirical analysis is the first time the endogeneity of the DIF group indicator has been addressed.

The remainder of this article is organized as follows. In the next

Abbreviations: BRM, behavioral Rasch model; BRSM, behavioral Rasch selection model; CPS-FSS, Current Population Survey Food Security Supplement; DIF, differential item functioning; EAP, expected a posteriori; HFSSM, household food security survey module; IRT, Item Response Theory; MML, marginal maximum likelihood; POM, potential outcomes model; RMSE, root mean square error; RCM, Rubin causal model; SNAP, Supplemental Nutrition Assistance Program; TSM, Tobit selection model

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

https://doi.org/10.1016/j.measurement.2018.01.044
Received 5 May 2017; Received in revised form 22 January 2018; Accepted 23 January 2018
Available online 02 February 2018
0263-2241/ Published by Elsevier Ltd.

section, I discuss how the BRSM is developed from the Rasch model and other methods for estimating the latent regression model parameters. The following section describes the simulation analyses and discusses the results. Next, I describe the data used in the empirical examples and how the BRSM is useful in two particular settings: estimating the causal effect of a binary endogenous variable on a latent variable, and correcting for endogeneity in analyses of DIF. Results from the empirical examples are discussed with emphasis on their implications for future applications of the BRSM. Lastly, I discuss the implications of my findings from the simulations and empirical examples.

2. An illustrative Rasch model

For illustrative purposes, I consider the Rasch model for dichotomous responses in this article; however, the methods presented can be modified to include more complex models, such as the two-parameter IRT logistic model [5] or models that allow for polytomous responses [1,3]. The model is framed to measure a person's ability; however, it could just as easily be modified to measure a person's health, food insecurity, or any other latent trait. The model assumes person i's underlying latent index of ability, denoted by θ_i, represents the person's location on the continuum of ability. Higher values of θ_i are associated with greater ability. Assuming there are J binary items administered to the person that capture different levels of ability, then the probability that the person affirms the j^{th} item is

$$P(Y_{ij}=1|\theta_i)=\frac{\exp\left(\theta_i-\sum_{k=1}^{J}\delta_k X_{kj}\right)}{1+\exp\left(\theta_i-\sum_{k=1}^{J}\delta_k X_{kj}\right)}, \tag{1}$$

where exp(•) is the exponential function, δ_k is an item-difficulty parameter, and X_{kj} equals one if k = j, and zero otherwise. Item-difficulty parameters represent each items location on the continuum of ability, and are assumed to take on different values for each item. Higher item-difficulty parameter values are consistent with items that capture greater ability. The model further assumes that the person's responses are independent, conditional on latent ability, and the item-discrimination parameters are constrained to be equal across all items and normalized to one.

The item-difficulty parameters can be estimated using a maximum likelihood method called marginal maximum likelihood (MML; [6]) because the parameters of the distribution of latent ability are integrated out. For a detailed discussion of the methods for estimating the Rasch and other IRT model parameters, see [36]. The person's ability parameter can be estimated, conditional on the values of the item-difficulty parameters, using maximum likelihood or Bayesian methods. Bayesian methods for obtaining estimates of the person's ability parameters are particularly useful when these estimates are to be used in linear or censored regression models, where the dependent variable is the estimate of person ability, since estimates of ability can be obtained for persons with all zero and perfect scores on an instrument. Persons with extreme responses to an instrument are particularly important in these regressions since their responses may reveal important differences in behavior that are associated with ability. Maximum likelihood methods have been developed that "assign" values of ability to persons with extreme responses; however, they depend on which statistical software is used [12].

The most commonly used methods for estimating IRT model parameters are MML for estimating the item-difficulty parameters and the Bayesian expected a posteriori (EAP) method for estimating the person-ability parameters [36]. A benefit of the Bayesian EAP method for obtaining estimates of the person-ability parameters is that linear and censored regression models using them as dependent variables do not have to be adjusted for estimation error from the measurement model prior to testing for differences in ability based on the explanatory variables.

All analyses contained in this article were performed using Stata 15 Multiprocessor. The number of quadrature points was set at 15 for the numerical methods required to estimate the item-difficulty and person-ability parameters. When the numerical methods required evaluating integrals of more than one dimension, 15 quadrature points were used for each dimension. Using 15 quadrature points has been shown to produce reasonably accurate parameter estimates in this type of analysis [7,31].

2.1. Incorporating person-level covariates into the Rasch model

The Rasch model, described above, has been used extensively for scale development in the measurement sciences; however, it can also be used to directly examine the relationship between latent ability and a series of explanatory variables. A multivariate behavioral component, consisting of person-level explanatory variables, can be incorporated into the Rasch model by respecifying the person's latent ability index (θ_i) as

$$\theta_i=\beta_T T_i+\beta_X' X_i+e_i,\ \text{with}\ e_i\sim \text{i. i. d.}\ \ N(0,\sigma^2), \tag{2}$$

where T_i is an observed treatment indicator and X_i is a vector of observed person-level explanatory variables. The term treatment is used very broadly in economics and other fields. Essentially, it covers any variable whose effect on some outcome is the object of study. Substituting Eq. (2) into Eq. (1) yields

$$P(Y_{ij}=1|T_i,X_i,e_i)=\frac{\exp\left(\beta_T T_i+\beta_X' X_i+e_i-\sum_{k=1}^{J}\delta_k X_{kj}\right)}{1+\exp\left(\beta_T T_i+\beta_X' X_i+e_i-\sum_{k=1}^{J}\delta_k X_{kj}\right)}. \tag{3}$$

The model described in Eq. (3) is referred to as the behavioral Rasch model (BRM) in this article. The term "behavioral" is used to emphasize that the BRM includes a latent regression model in addition to the measurement model. Alternatively, the BRM has also been described as a generalized Rasch model [38] and a person-explanatory Rasch model [13]. The BRM parameters can be estimated using the MML method.

The latent regression specified in Eq. (2) can also be estimated using the predicted person-ability parameters as the dependent variable in a regression model. Any measurement error resulting from the estimation of the person-ability parameters must be addressed for this approach to be feasible if the ability parameters are estimated using a maximum likelihood method. A common method for addressing this measurement error is to assume it is orthogonal to the person-level explanatory variables (i.e., classical measurement error). Since person-ability parameters can theoretically take on any value on the real line it is common for these models to be estimated using a linear regression model [9,38]. Yet, data limitations or poor instrument design may constrain the estimates of the person-ability parameters to a smaller interval on the real line.

If an instrument is poorly designed or survey data contains a selective sample of persons with high or low levels of ability, then there may be restricted coverage of items or persons on the continuum of latent ability, respectively. If this occurs, person-ability parameter estimates can be grouped at the bounds of the feasible range for estimation. Censoring will occur resulting in biased and inconsistent estimates of the regression model parameters. This can be addressed by estimating a censored regression model, such as the Tobit model. Under the Tobit model with censoring at the lower level, values of latent ability at or below the censoring threshold are unobservable and assigned a value of c, while latent ability above the threshold is observed and set equal to the person's ability parameter estimate.