The Canadian Journal of Statistics
Vol. 47, No. 4, 2019, Pages 653–667
La revue canadienne de statistique

Instrumental variable estimation in ordinal probit models with mismeasured predictors

Jing GUAN[1], Hongjian CHENG[1], Kenneth A. BOLLEN[2,3], D. Roland THOMAS[4], and Liqun WANG[5*]

[1]*School of Mathematics, Tianjin University, Tianjin 300075, China*
[2]*Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3270, U.S.A.*
[3]*Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3270, U.S.A.*
[4]*Sprott School of Business, Carleton University, Ottawa, Ontario K1S 5B6, Canada*
[5]*Department of Statistics, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada*

Key words and phrases: Instrumental variable; latent predictors; likelihood method; measurement error; ordinal dependent variable; probit model.

MSC 2010: Primary 62J12; secondary 62F12.

Abstract: Researchers in the medical, health, and social sciences routinely encounter ordinal variables such as self-reports of health or happiness. When modelling ordinal outcome variables, it is common to have covariates, for example, attitudes, family income, retrospective variables, measured with error. As is well known, ignoring even random error in covariates can bias coefficients and hence prejudice the estimates of effects. We propose an instrumental variable approach to the estimation of a probit model with an ordinal response and mismeasured predictor variables. We obtain likelihood-based and method of moments estimators that are consistent and asymptotically normally distributed under general conditions. These estimators are easy to compute, perform well and are robust against the normality assumption for the measurement errors in our simulation studies. The proposed method is applied to both simulated and real data. *The Canadian Journal of Statistics* 47: 653–667; 2019 © 2019 Statistical Society of Canada

Résumé: Les chercheurs en médecine, en santé et en sciences sociales sont régulièrement confrontés à des variables ordinales comme des mesures auto-déclarées de santé ou de bonheur. Pour modéliser une variable réponse ordinale, il est fréquent qu'une erreur de mesure affecte les covariables telles que l'attitude, le revenu familial ou des variables rétrospectives. Ignorer de telles erreurs de mesure peut causer un biais préjudiciable à l'estimation des paramètres. Les auteurs proposent une approche par variable instrumentale pour l'estimation d'un modèle probit pour une réponse ordinale et des prédicteurs mal mesurés. Ils obtiennent des estimateurs par le maximum de vraisemblance et la méthode des moments qui sont convergents et asymptotiquement normaux sous des conditions générales. À l'aide de simulations, les auteurs constatent que les estimateurs proposés sont faciles à calculer, offrent de bonnes performances et sont robustes à une violation de l'hypothèse de normalité des erreurs de mesure. Ils appliquent leur méthode à des données simulées et réelles. *La revue canadienne de statistique* 47: 653–667; 2019 © 2019 Société statistique du Canada

---

Additional Supporting Information may be found in the online version of this article at the publisher's website.
* *Author to whom correspondence may be addressed.*
E-mail: liqun.wang@umanitoba.ca

1. INTRODUCTION

Ordinal dependent variables are common in the medical, health, social, and behavioural sciences (Agresti, 2002, 2010). For instance, Tosteson, Stefanski & Schafer (1989) modeled ordinal health outcomes as a function of pollution measures that are subject to error. Mohanty (2014) investigated an ordinal happiness variable as a function of income and attitudes. Cranfield & Magnusson (2003) used an ordered probit model to analyze the willingness of Canadian consumers to pay higher premiums for pesticide-free products. Varin & Czado (2010) studied a mixed probit model for ordinal longitudinal data and applied it to the analysis of determinants of migraine severity in a group of Canadian patients. Baetschmann, Staub & Winkelmann (2015) used an ordered logit model to study life satisfaction in Germany. Lu, Poon & Cheung (2015) proposed a test for the multiple comparison of several treatments with a control in clinical trials and applied the method to compare the efficacy of intravenous treatment with various doses of fentanyl and lidocaine for reducing the pain on injection of propofol in surgical patients.

More real data examples and applications of the various categorical response models may be found in Agresti (2002, 2010). McCullagh (1980) developed a maximum likelihood method for parameter estimation in ordinal response models. More recently, an overview and survey of methods for ordered categorical data analysis was provided by Liu & Agresti (2005). See also Peyhardi, Trottier & Guedon (2015) for a discussion of methodology developments in categorical response models.

In particular, Zhu et al. (2014) used a logistic model to predict the pathologic responses of a group of breast cancer patients. where the predictors included certain tumour pathologic variables as well as haemoglobin variables measured by ultrasound-guided near-infrared optical tomography. In that study the original response variable took five ordinal values which were dichotomized so that a logistic model could be used. Apparently the original responses contained much more information and therefore using them would lead to more accurate predictions. Moreover, some predictors such as haemoglobin variables can hardly be measured precisely.

In practice, many practical data analyses involve predictor variables, such as pollution, attitudes and family income, that cannot be measured precisely and instead only proxy observations are available to estimate model parameters. It is well known that statistical methods which ignore such measurement error result in biased and inconsistent estimates, and therefore misleading conclusions (Carroll et al., 2006). While the problem of measurement error in regression models with continuous or binary response variables has been intensively studied, the same cannot be said for models with ordinal response variables as well as measurement error in the covariates. For example, see Huang & Wang (2001), Xu, Ma & Wang (2015) and Yi (2017), and the references therein. Some authors, such as Li & Hsiao (2004), Abarin & Wang (2012) and Li & Wang (2012), have studied this problem in the context of generalized linear models, while Chen, Yi & Wu (2014) investigated this misclassification problem in a longitudinal data model with both a categorical response and covariates.

In this article we consider a probit model with ordinal response and predictor variables that are measured with error. In particular, we use an instrumental variable (IV) approach to derive consistent and asymptotically normally distributed estimators for the unknown parameters. This is achieved by using a linear projection of the unobserved true predictors onto the instrumental variables to create an auxiliary probit model with observed predictors, so that we can use the standard likelihood method. Within this framework we can also derive method of moments estimators (MMEs) which are easier to compute numerically. Our simulation studies demonstrate that the proposed estimators have satisfactory finite sample performances under standard model assumptions as well as under some mis-specified models.

Methods that involve the use of IVs have been widely used in measurement error models. One advantage of this approach is that these methods require weaker assumptions than most other (consistent) estimators in measurement error models. In fact, the main requirement is that the

IVs be correlated with the observed covariate but independent of the measurement and equation error. For example, a popular method in the literature is to use replicate data to estimate the measurement error variances. In this case one member of the replicated data qualifies as an IV. More generally, if longitudinal or repeatedly measured data are available, then the measurements at different time points can serve as IVs.

The IV method has been used in binary regression models by Stefanski & Buzas (1995) and Buzas & Stefanski (1996a), and in generalized linear models by Buzas & Stefanski (1996b) and Abarin & Wang (2012). In particular, Huang & Wang (2001) constructed corrected-score estimators for binary logistic models, while Xu, Ma & Wang (2015) proposed a semiparametric efficient method. Moreover, the IV method is also used in censored linear regression models by Wang & Hsiao (2007) and in general nonlinear models by Wang & Hsiao (2011). Our simulation studies show that our proposed estimators exhibit performance that is similar to that of the estimator described Xu, Ma & Wang (2015) in a semiparametric model and also to that of the maximum likelihood estimator of Buzas & Stefanski (1996a) in a probit model where all the variables are assumed to follow normal distributions. However, in their work no estimate of the covariance matrix of the parameter estimator was provided. In addition to its computational simplicity, a further advantage of our proposed procedure is that it produces consistent estimators for the measurement error variances, which are not available if the alternative methods mentioned above are used.

In statistics, two commonly used methods in general nonlinear measurement error models are regression calibration (Carroll & Stefanski, 1990) and simulation extrapolation (SIMEX) (Cook & Stefanski, 1994). However, these methods give only *approximately* consistent estimators and, therefore, are only applicable when the measurement errors are small. Furthermore, extra simulation procedures such as use of the bootstrap are required to compute the variances of the estimators.

The remainder of this article is organized as follows. In Section 2 we introduce the model and its assumptions. In Section 3 we develop an IV estimator and establish its consistency and asymptotic normality. We also propose an alternative estimator based on the method of moments. In Section 4 we employ Monte Carlo simulations to study the finite sample properties of our proposed estimators under various conditions. In Section 5, we apply our methods to analyze a health survey data set. Finally, conclusions and discussion are presented in Section 6. All mathematical proofs and derivations are provided in the accompanying Appendix.

2. THE MODEL

Let the observed response variable $Y = \sum_{j=1}^{J-1} \mathbf{1}\left(t_j \leq Y^*\right)$, where $\mathbf{1}(\cdot)$ is the indicator function and $-\infty < t_1 < t_2 < \cdots < t_{J-1} < \infty$ are unknown thresholds and $J \geq 2$. Thus Y is an ordinal variable taking on values $j = 0, 1, 2, \ldots, J-1$. Obviously Y is a binary random variable when $J = 2$. The underlying latent response Y^* is related to predictor variables $\boldsymbol{X}^*$ and $\boldsymbol{X}$ via

$$Y^* = \alpha_1 + \boldsymbol{\alpha}_2'\boldsymbol{X}^* + \boldsymbol{\alpha}_3'\boldsymbol{X} + \epsilon, \tag{1}$$

where $\boldsymbol{X}^*$ and $\boldsymbol{X}$ are mismeasured and correctly measured predictors respectively, $\boldsymbol{\alpha} = (\alpha_1, \boldsymbol{\alpha}_2', \boldsymbol{\alpha}_3')'$ is the vector of unknown parameters, $\epsilon \sim N(0, \sigma_\epsilon^2)$ and is uncorrelated with $\boldsymbol{X}^*$ and $\boldsymbol{X}$. Note that subtracting a constant from the latent variable Y^* and $t_j, j = 1, 2, \ldots, J-1$ or dividing these quantities by a positive constant will give the same values of the observed response Y. Therefore, for the purpose of model identifiability, the location and scale of Y^* can be set to $t_1 = 0$ and $\sigma_\epsilon^2 = 1$. Another way of normalizing the location of Y^* is to set the intercept α_1 equal to 0 and leave t_1 unrestricted. However, the normalization is not a restriction of generality and is similar to the standard logistic model where the standard logistic distribution is used, for example, Long (1997). See also Agresti (2002, pp. 277–279) for the latent variable modeling of ordinal responses.

Furthermore, assume that we observe

$$\boldsymbol{W} = \boldsymbol{X}^* + \boldsymbol{U}, \tag{2}$$

where the measurement error $\boldsymbol{U} \sim N(0, \boldsymbol{\Sigma}_{uu})$ and is independent of $\boldsymbol{X}^*$, $\boldsymbol{X}$ and ϵ. Throughout the article we will denote the mean by $\boldsymbol{\mu}$ and the variance–covariance matrix by $\boldsymbol{\Sigma}$ with subscripts to indicate the random variables of interest. For instance, the mean of a random vector $\boldsymbol{X}$ is $\boldsymbol{\mu}_x$, its variance-covariance matrix is $\boldsymbol{\Sigma}_{xx}$, and the covariance matrix between two random vectors $\boldsymbol{X}$ and $\boldsymbol{W}$ is $\boldsymbol{\Sigma}_{xw}$.

Measurement error can induce a lack of model identifiability so that the unknown parameters in the model cannot be uniquely determined by the sampling distribution of the data and therefore cannot be estimated consistently. To ensure identifiability, usually extra data or information is required in addition to the main sample, for example, validation data, replicate measurements, or instrumental data. For examples see Carroll et al. (2006) and Wang & Hsiao (2011).

In this article we assume that an IV $\boldsymbol{Z}$ is available that is correlated with $\boldsymbol{X}^*$ but independent of $\boldsymbol{U}$ and ϵ. Let $\boldsymbol{\beta}_2 = \boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{zx^*}$, $\boldsymbol{\beta}_1 = \boldsymbol{\mu}_{x^*} - \boldsymbol{\beta}_2'\boldsymbol{\mu}_z$ and write the linear projection of $\boldsymbol{X}^*$ on $\boldsymbol{Z}$ as

$$\boldsymbol{X}^* = \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2'\boldsymbol{Z} + \boldsymbol{\delta}, \tag{3}$$

where the projection error $\boldsymbol{\delta}$ satisfies $E(\boldsymbol{\delta}) = 0$ and $\mathrm{Cov}(\boldsymbol{\delta}, \boldsymbol{Z}) = 0$ by construction. We also assume that $\dim(\boldsymbol{Z}) \geq \dim(\boldsymbol{X}^*)$ and the covariance matrices $\boldsymbol{\Sigma}_{zx^*}, \boldsymbol{\Sigma}_{zz}$ have full column rank. The first assumption is necessary so that we have at least as many IVs as $\boldsymbol{X}^*$ covariates that are measured with error. This is a well-known requirement for IV estimation in order to be able to find unique estimates of each coefficient. The full column rank assumption is necessary to ensure that no IVs are redundant and that they are sufficiently correlated with $\boldsymbol{X}^*$. The IVs can be continuous or discrete variables. For further discussion of these IV assumptions see, for instance, Bollen (2012) or Greene (2012). Our primary interest is to obtain consistent estimators of the regression parameter $\boldsymbol{\alpha}$ in the probit model specified in Equation (1).

3. INSTRUMENTAL VARIABLE ESTIMATION

In this section we derive consistent estimators of unknown parameters in the model specified in Equations (1)–(3) based on the i.i.d. random sample $(Y_i, \boldsymbol{X}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i), i = 1, 2, \ldots, n$. We first show that the parameter $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2')$ in the instrumental equation, that is, Equation (3), is straightforward to estimate by combining Equations (2) and (3). Indeed, since $\boldsymbol{Z}$ is uncorrelated with $\boldsymbol{\delta}$ and $\boldsymbol{U}$, by Equations (2) and (3) we have $\boldsymbol{\beta}_2 = \boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{zw}$ and $\boldsymbol{\beta}_1 = \boldsymbol{\mu}_w - \boldsymbol{\beta}_2'\boldsymbol{\mu}_z$. Therefore $\boldsymbol{\beta}$ can be consistently estimated by the corresponding sample moments of $\boldsymbol{Z}_i$ and $\boldsymbol{W}_i$. In the following we derive two consistent estimators for $\boldsymbol{\alpha}$ using a likelihood approach and the method of moments.

3.1. A Likelihood-based Estimator

First, we substitute Equation (3) into Equation (1) to obtain

$$Y^* = \tilde{\gamma}_1 + \tilde{\boldsymbol{\gamma}}_2'\boldsymbol{Z} + \tilde{\boldsymbol{\gamma}}_3'\boldsymbol{X} + V, \tag{4}$$

where $\tilde{\gamma}_1 = \alpha_1 + \boldsymbol{\beta}_1'\boldsymbol{\alpha}_2$, $\tilde{\boldsymbol{\gamma}}_2 = \boldsymbol{\beta}_2\boldsymbol{\alpha}_2$, $\tilde{\boldsymbol{\gamma}}_3 = \boldsymbol{\alpha}_3$, and $V = \boldsymbol{\alpha}_2'\boldsymbol{\delta} + \epsilon$ has mean 0 and variance $\sigma_v^2 = \boldsymbol{\alpha}_2'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\alpha}_2 + 1$. After dividing both sides of Equation (4) by σ_v, it follows that the result becomes a standard ordinal probit model and therefore $\boldsymbol{\gamma} = (\gamma_1, \boldsymbol{\gamma}_2', \boldsymbol{\gamma}_3')'$ with $\boldsymbol{\gamma}_j = \sigma_v^{-1}\tilde{\boldsymbol{\gamma}}_j, j = 1, 2, 3,$

can be consistently estimated by the method of maximum likelihood outlined in McCullagh (1980).

To obtain the estimator for $\boldsymbol{\alpha}$, we rewrite its relation with $\boldsymbol{\gamma}$ as

$$\boldsymbol{\alpha} = \sigma_v \boldsymbol{L}\boldsymbol{\gamma}, \tag{5}$$

where

$$\boldsymbol{L} = \begin{pmatrix} 1 & -\boldsymbol{\beta}_1'\boldsymbol{M} & 0 \\ 0 & \boldsymbol{M} & 0 \\ 0 & 0 & \boldsymbol{I} \end{pmatrix}, \tag{6}$$

Here $\boldsymbol{I}$ is an identity matrix, $\boldsymbol{M} = (\boldsymbol{\Sigma}_{wz}\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{zw})^{-1}\boldsymbol{\Sigma}_{wz}$ and $\boldsymbol{\Sigma}_{zw} = \boldsymbol{\Sigma}_{zx^*}$ which has full rank by assumption. We show in the Appendix that the standard deviation of V is

$$\sigma_v = \left(\boldsymbol{\gamma}_2'\boldsymbol{\Sigma}_{zz}\boldsymbol{\gamma}_2 + 2\boldsymbol{\gamma}_2'\boldsymbol{\Sigma}_{zx}\boldsymbol{\gamma}_3 - \boldsymbol{\gamma}_2'\boldsymbol{M}'\boldsymbol{\Sigma}_{wx}\boldsymbol{\gamma}_3 + 1 - \eta\rho\boldsymbol{\Sigma}_{wy}'\boldsymbol{M}\boldsymbol{\gamma}_2\right)^{-1/2}, \tag{7}$$

where

$$\eta = \left(\boldsymbol{\gamma}_2'\boldsymbol{\Sigma}_{zz}\boldsymbol{\gamma}_2 + 2\boldsymbol{\gamma}_2'\boldsymbol{\Sigma}_{zx}\boldsymbol{\gamma}_3 + \boldsymbol{\gamma}_3'\boldsymbol{\Sigma}_{xx}\boldsymbol{\gamma}_3 + 1\right)^{1/2}, \tag{8}$$

$\rho = \left\{\sum_{j=1}^{J-1} \phi\left[\Phi^{-1}(p_j)\right]\right\}^{-1}$, $p_j = P(Y \le j-1), j = 1, 2, \dots, J-1$; $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution functions, respectively, and $\Phi^{-1}(\cdot)$ denotes the inverse function of $\Phi(\cdot)$.

From Equations (5)–(8) we can see that, besides $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$ also depends on other parameters such as $p_1, p_2, \dots, p_{J-1}$ and $\boldsymbol{\Sigma}_{wy}$, which can be estimated by their corresponding sample moments. However, the derivation of the asymptotic covariance matrix for the estimator for $\boldsymbol{\alpha}$ will be complicated if these moment estimators are correlated with the estimator of $\boldsymbol{\gamma}$. In order to overcome this problem we propose to obtain these estimators using separate subsamples $Y_i, \boldsymbol{W}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i, i = 1, 2, \dots, n_1$ and $Y_i, \boldsymbol{W}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i, i = n_1 + 1, n_1 + 2, \dots, n$, where $1 < n_1 < n$ is such that $n_1 \to \infty$ as $n \to \infty$ and n/n_1 tends to a positive constant.

Thus all the model parameters can be estimated as follows: First, estimate the first two moments of $\boldsymbol{X}, \boldsymbol{W}, \boldsymbol{Z}$ using the sample moments of $\boldsymbol{X}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i, i = 1, 2, \dots, n$. Second, evaluate estimators $\hat{p}_j = \sum_{i=1}^{n_1} \mathbf{1}(Y_i \le j-1)/n_1$, $\hat{\boldsymbol{\mu}}_{wy} = \sum_{i=1}^{n_1} \boldsymbol{W}_i Y_i/n_1$, $\hat{\boldsymbol{\mu}}_w = \sum_{i=1}^{n_1} \boldsymbol{W}_i/n_1$, $\hat{\mu}_y = \sum_{i=1}^{n_1} Y_i/n_1$ and $\hat{\boldsymbol{\Sigma}}_{wy} = \hat{\boldsymbol{\mu}}_{wy} - \hat{\mu}_w\hat{\mu}_y$ using the data $Y_i, \boldsymbol{W}_i, i = 1, 2, \dots, n_1$. Third, calculate the maximum likelihood estimator $\hat{\boldsymbol{\gamma}}$ using the data $Y_i, \boldsymbol{X}_i, \boldsymbol{Z}_i, i = n_1 + 1, n_1 + 2, \dots, n$. Finally, calculate the estimates $\hat{\sigma}_v$ via Equations (7) and (8) and $\hat{\boldsymbol{\alpha}}$ using Equation (5).

Now we investigate the asymptotic properties of the likelihood-based IV estimator $\hat{\boldsymbol{\alpha}}$ defined in Equation (5). Following the standard regression literature we consider the conditional asymptotic distribution of $\hat{\boldsymbol{\alpha}}$ given the observed predictors $\boldsymbol{X}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i,\ i = 1, 2, \dots, n$, so that their moments are treated as known for the sake of notational convenience. Let $\boldsymbol{\psi} = (p_1, p_2, \dots, p_{J-1}, \boldsymbol{\mu}_{wy}', \mu_y)'$ and $\hat{\boldsymbol{\psi}} = \sum_{i=1}^{n_1} T_i/n_1$, where $T_i = \left(\mathbf{1}(Y_i \le 0), \dots, \mathbf{1}(Y_i \le J-2), \boldsymbol{W}_i'Y_i, Y_i\right)'$. Then by the central limit theorem $\sqrt{n_1}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_\psi)$ as $n_1 \to \infty$, where $\boldsymbol{\Sigma}_\psi = \text{plim}_{n_1\to\infty} \sum_{i=1}^{n_1}(T_i - \bar{T})(T_i - \bar{T})'/n_1$. Furthermore, under general conditions the maximum likelihood estimator of $\boldsymbol{\gamma}$ satisfies $\sqrt{n-n_1}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_\gamma)$ as $n - n_1 \to \infty$, where $\boldsymbol{\Sigma}_\gamma$ can be estimated in the usual likelihood set-up and is a part of the standard output of many statistical software packages, including R.

The asymptotic distribution of $\hat{\boldsymbol{\alpha}}$ can be obtained via the delta-method based on the first-order Taylor expansion

$$\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = \frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\psi}'}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) + \frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\gamma}'}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}). \tag{9}$$

However, derivation of the asymptotic covariance matrix of $\hat{\boldsymbol{\alpha}}$ would be complicated if the estimators $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\psi}}$ were correlated. This is the reason we use separate samples to estimate them. Hence, we have the following asymptotic behaviour for our proposed estimator, the proof of which is provided in the accompanying Appendix.

Theorem 1. *Suppose* $1 < n_1 < n$ *satisfies* $\lim_{n\to\infty} n/n_1 = c_1 \in (1, \infty)$ *and let* $\boldsymbol{\Sigma}_\psi$ *and* $\boldsymbol{\Sigma}_\gamma$ *be the asymptotic covariance matrix of* $\hat{\boldsymbol{\psi}}$ *and* $\hat{\boldsymbol{\gamma}}$*, respectively. Then under the specified assumptions, and conditional on* $\boldsymbol{X}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i, i = 1, 2, \ldots, n$,

$$\sqrt{n}\,(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{d} N\big(0, c_1 \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{\psi}}\boldsymbol{A}' + c_2 \boldsymbol{B}\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\gamma}}}\boldsymbol{B}'\big) \text{ as } n \to \infty,$$

where $c_2 = c_1/(c_1 - 1)$,

$$\boldsymbol{A} = \frac{\rho\sigma_{y^*}\boldsymbol{\alpha}\boldsymbol{\alpha}_2'}{2}\left(\rho\boldsymbol{\Sigma}_{wy}\Phi^{-1}(p_1), \ldots, \rho\boldsymbol{\Sigma}_{wy}\Phi^{-1}(p_{J-1}), \boldsymbol{I}, -\boldsymbol{\mu}_w\right), \tag{10}$$

$$\boldsymbol{B} = \sigma_v \boldsymbol{L} - \frac{\sigma_v\boldsymbol{\alpha}}{2}(0, \boldsymbol{d}_1, \boldsymbol{d}_2), \tag{11}$$

$$\begin{aligned}\boldsymbol{d}_1 = {} & 2\boldsymbol{\alpha}_2'\boldsymbol{\Sigma}_{wz} + 2\boldsymbol{\alpha}_3'\boldsymbol{\Sigma}_{xz} - \boldsymbol{\alpha}_3'\boldsymbol{\Sigma}_{xw}\boldsymbol{M} - \sigma_{y^*}\rho\boldsymbol{\Sigma}_{wy}'\boldsymbol{M} \\ & -\sigma_{y^*}^{-1}\rho\boldsymbol{\Sigma}_{wy}'\boldsymbol{\alpha}_2(\boldsymbol{\alpha}_2'\boldsymbol{\Sigma}_{wz} + \boldsymbol{\alpha}_3'\boldsymbol{\Sigma}_{xz}),\end{aligned} \tag{12}$$

$$\begin{aligned}\boldsymbol{d}_2 = {} & 2\boldsymbol{\alpha}_2'\boldsymbol{\Sigma}_{wz}\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{zx} - \boldsymbol{\alpha}_2'\boldsymbol{\Sigma}_{wx} \\ & -\sigma_{y^*}^{-1}\rho\boldsymbol{\Sigma}_{wy}'\boldsymbol{\alpha}_2(\boldsymbol{\alpha}_2'\boldsymbol{\Sigma}_{wz}\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{zx} + \boldsymbol{\alpha}_3'\boldsymbol{\Sigma}_{xx}),\end{aligned} \tag{13}$$

and

$$\sigma_{y^*}^2 = \boldsymbol{\alpha}_2'\boldsymbol{\Sigma}_{wz}\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{zw}\boldsymbol{\alpha}_2 + 2\boldsymbol{\alpha}_2'\boldsymbol{\Sigma}_{wz}\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{zx}\boldsymbol{\alpha}_3 + \boldsymbol{\alpha}_3'\boldsymbol{\Sigma}_{xx}\boldsymbol{\alpha}_3 + \sigma_v^2. \tag{14}$$

Note that we used the split samples to calculate $\hat{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\gamma}}$ in order to obtain a simpler formula for the asymptotic covariance matrix of $\hat{\boldsymbol{\alpha}}$. However, splitting the sample reduces the effective sample size and results in a loss of efficiency in the estimator $\hat{\boldsymbol{\alpha}}$. In applications where the sample size is relatively small, the entire sample can be used at all stages in order to obtain more efficient estimates. In that case extra computation such as using the bootstrap has to be employed in order to obtain the variances of the estimators.

3.2. A Method of Moments Estimator

The estimator discussed in the previous subsection requires calculation of the maximum likelihood estimators in the reduced model identified in Equation (4). In this subsection we propose a computationally simpler, moments-based estimator under the additional assumption that the error-free predictor $\boldsymbol{X}$ and instrumental variable $\boldsymbol{Z}$ are normally distributed. For this purpose we denote $\tilde{\boldsymbol{X}} = (\boldsymbol{X}^{*\prime}, \boldsymbol{X}')'$, $\tilde{\boldsymbol{W}} = (\boldsymbol{W}', \boldsymbol{X}')'$, $\tilde{\boldsymbol{Z}} = (\boldsymbol{Z}', \boldsymbol{X}')'$, $\tilde{\boldsymbol{U}} = (\boldsymbol{U}', 0)'$ and $\tilde{\boldsymbol{\alpha}}_2 = (\boldsymbol{\alpha}_2', \boldsymbol{\alpha}_3')'$. We show in the Appendix that under the model specified in Equations (1)–(3) we have the following moment identities:

$$\tilde{\boldsymbol{\alpha}}_2 = \rho\sigma_{y^*}\tilde{\boldsymbol{M}}\boldsymbol{\Sigma}_{\tilde{z}\tilde{z}}^{-1}\boldsymbol{\Sigma}_{\tilde{z}y}, \tag{15}$$

where $\tilde{\boldsymbol{M}} = (\boldsymbol{\Sigma}_{\tilde{w}\tilde{z}}\boldsymbol{\Sigma}_{\tilde{z}\tilde{z}}^{-1}\boldsymbol{\Sigma}_{\tilde{z}\tilde{w}})^{-1}\boldsymbol{\Sigma}_{\tilde{w}\tilde{z}}$,

$$\alpha_1 = t_1 - \sigma_{y^*}\Phi^{-1}(p_1) - \boldsymbol{\mu}_{\tilde{w}}'\tilde{\boldsymbol{\alpha}}_2, \tag{16}$$

where $t_1 = 0, p_1 = P(Y \leq t_1)$ and

$$\sigma_{y^*} = \left(1 - \rho^2 \boldsymbol{\Sigma}'_{\tilde{w}y} \tilde{\boldsymbol{M}} \boldsymbol{\Sigma}^{-1}_{\tilde{z}\tilde{z}} \boldsymbol{\Sigma}_{\tilde{z}y}\right)^{-1/2}. \tag{17}$$

Therefore a MME for $\boldsymbol{\alpha}$ can be obtained by substituting the corresponding sample moments of $Y_i, \tilde{\boldsymbol{W}}_i, \tilde{\boldsymbol{Z}}_i$ into the above equations. The asymptotic covariance matrix of this MME can also be derived in a manner analogous to the corresponding covariance matrix for the likelihood-based estimator that we discussed in Section 3.1.

3.3. An Estimator of Measurement Error Variances

Although our primary focus is consistent estimation of the regression coefficient $\boldsymbol{\alpha}$, sometimes researchers are also interested in estimating other parameters such as the measurement error variances. In this subsection we derive such variance estimators for the special case where the components of measurement error $\boldsymbol{U}$, say $U_1, U_2, \ldots, U_p$, are uncorrelated. Specifically, let the measurement error covariance matrix $\boldsymbol{\Sigma}_{uu} = \text{diag}(\sigma^2_{u_1}, \sigma^2_{u_2}, \ldots, \sigma^2_{u_p})$ and $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)'$ the elementary vector with ith entry 1 and all others 0. Then, as we show in the Appendix,

$$\sigma^2_{u_i} = \frac{\boldsymbol{e}'_i(\boldsymbol{\Sigma}_{ww}\boldsymbol{\alpha}_2 + \boldsymbol{\Sigma}_{wx}\boldsymbol{\alpha}_3 - \rho\sigma_{y^*}\boldsymbol{\Sigma}_{wy})}{\boldsymbol{e}'_i\boldsymbol{\alpha}_2}, \tag{18}$$

$i = 1, 2, \ldots, p$, where σ_{y^*} is given by

$$\sigma_{y^*} = \frac{\rho\boldsymbol{\Sigma}'_{wy}\boldsymbol{\alpha}_2}{2} + \left\{\left(\frac{\rho\boldsymbol{\Sigma}'_{wy}\boldsymbol{\alpha}_2}{2}\right)^2 + \boldsymbol{\alpha}'_2\boldsymbol{\Sigma}_{wx}\boldsymbol{\alpha}_3 + \boldsymbol{\alpha}'_3\boldsymbol{\Sigma}_{xx}\boldsymbol{\alpha}_3 + 1\right\}^{1/2}. \tag{19}$$

Therefore, a consistent estimator of $\sigma^2_{u_i}$ can be obtained by substituting the consistent estimator of $\boldsymbol{\alpha}$ and sample moments of $\boldsymbol{W}_i, \boldsymbol{X}_i, Y_i$ into Equations (18) and (19). Note that Equations (18) and (19) do not require the normality assumption for $\boldsymbol{X}, \boldsymbol{Z}$ and therefore they can be used with either the likelihood-based estimator of $\boldsymbol{\alpha}$ or the MME (under the normality assumption).

4. A SIMULATION STUDY

In this section, we report results from a simulation study to investigate the finite sample properties of our proposed estimators. We calculated the likelihood-based instrumental variable estimator (IVE) and the MME. For comparison purposes, we also evaluated the naive maximum likelihood estimator that ignored the measurement error (NAE). For each estimator, we evaluated its bias (BIAS), its root mean squared error (RMSE) and the estimated 95% coverage probability (CP), where the standard errors of the IVE were calculated using the asymptotic variance formula specified in Theorem 1 . We considered three different sample sizes $n = 200$, 500, 1,000 with $n_1 = n/2$ so that $c_1 = c_2 = 2$. In each simulation run, we carried out 1,000 Monte Carlo replications.

We used a model with two predictors, both of them mismeasured, so that the observed predictors were $W_i = X^*_i + U_i$, $i = 1, 2$, where $X^*_1 \sim N(10, 5)$, $X^*_2 \sim N(9, 4)$, $U_1 \sim N(0, 0.5)$, $U_2 \sim N(0, 1)$ and all these variables were independent. In addition, there were two instrumental variables that we generated independently, defined as $Z_i = (X^*_i - \beta_{i1} - \delta_i)/\beta_{i2}, i = 1, 2$, where $(\beta_{11}, \beta_{12}) = (1, 0.8)$, $(\beta_{21}, \beta_{22}) = (-3, 1.2)$ and $\delta_1, \delta_2 \sim N(0, 1)$. Finally, the latent response Y^* was generated using the true parameter values $\boldsymbol{\alpha}' = (-4.5, 0.3, 0.4)$ and the observed response Y took $J = 5$ possible categorical values. The numerical results are summarized in Table 1.

TABLE 1: Simulation results for three methods of estimation with normally distributed measurement error and instrumental variables. The true values of the parameters estimated were $\alpha_1 = -4.5$, $\alpha_2 = 0.3$, $\alpha_3 = 0.4$, $\sigma^2_{u_1} = 0.5$ and $\sigma^2_{u_2} = 1$

Sample size, n	Method	Item	Parameter estimated α_1	α_2	α_3	$\sigma^2_{u_1}$	$\sigma^2_{u_2}$
200	IVE	BIAS	−0.1486	0.0107	0.0091	0.2111	−0.0030
		RMSE	1.0318	0.0678	0.0841	0.8080	0.6943
		CP	94.9%	94.6%	94.3%	—	—
	MME	BIAS	−0.1153	0.0089	0.0060	0.0166	−0.0246
		RMSE	0.7801	0.0512	0.0624	0.3198	0.3228
		CP	94.8%	94.4%	94.7%	—	—
	NAE	BIAS	1.3188	−0.0430	−0.1018	—	—
		RMSE	1.4366	0.0578	0.1096	—	—
		CP	34.8%	79.1%	26.8%	—	—
500	IVE	BIAS	−0.0498	0.0043	0.0030	0.0908	−0.0521
		RMSE	0.5785	0.0399	0.0484	0.5486	0.4910
		CP	94.7%	95.3%	94.8%	—	—
	MME	BIAS	−0.0613	0.0043	0.0041	0.0037	−0.0044
		RMSE	0.4557	0.0309	0.0366	0.2171	0.1969
		CP	94.9%	95%	95.5%	—	—
	NAE	BIAS	1.3538	−0.0455	−0.1026	—	—
		RMSE	1.3970	0.0513	0.1056	—	—
		CP	2.7%	50.9%	2.2%	—	—
1,000	IVE	BIAS	−0.0247	0.0016	0.0033	0.0320	−0.0206
		RMSE	0.4055	0.0275	0.0336	0.4353	0.3438
		CP	95.5%	93.8%	94.2%	—	—
	MME	BIAS	−0.0178	0.0012	0.0024	−0.0096	−0.0077
		RMSE	0.3327	0.0215	0.0264	0.1602	0.1352
		CP	95.1%	95.7%	94.2%	—	—
	NAE	BIAS	1.3684	−0.0470	−0.1024	—	—
		RMSE	1.3884	0.0495	0.1038	—	—
		CP	0%	15.9%	0%	—	—

From Table 1 we can see that, overall, both IVE and MME exhibited satisfactorily small BIAS and RMSE for all sample sizes, and these measures decreased rapidly as the sample size increased. The only exception was that the bias values of $\hat{\alpha}_1$ when the sample size was 200 seemed a little larger than we expected. This probably occurred because splitting the sample reduces the effective sample size. The naive estimator NAE exhibited significantly larger values of BIAS and RMSE than the other two estimators, and both measures remained roughly constant

TABLE 2: Simulation results for three methods of estimation; sensitivity to the assumption of normally distributed measurement error and/or instrumental variables. The true values of the parameters estimated were $\alpha_1 = -4.5$, $\alpha_2 = 0.3$, $\alpha_3 = 0.4$, $\sigma^2_{u_1} = 1.7$, and $\sigma^2_{u_2} = 1.7$

Sample size, n	Method	Item	Parameter estimated α_1	α_2	α_3	$\sigma^2_{u_1}$	$\sigma^2_{u_2}$
200	IVE	BIAS	−0.1064	0.0076	0.0100	−0.0820	−0.1330
		RMSE	0.9849	0.0644	0.0824	1.1870	0.8865
		CP	94.7%	94.2%	96.2%	—	—
	MME	BIAS	−0.1019	0.0070	0.0088	−0.0724	−0.0672
		RMSE	0.7742	0.0497	0.0650	0.5608	0.5538
		CP	95.3%	94.7%	94.9%	—	—
	NAE	BIAS	2.2994	−0.0992	−0.1474	—	—
		RMSE	2.3579	0.1047	0.1528	—	—
		CP	1.1%	14.6%	4%	—	—
500	IVE	BIAS	−0.0530	0.0022	0.0064	−0.1040	−0.0511
		RMSE	0.5670	0.0397	0.0491	0.8917	0.6252
		CP	94.8%	94.9%	94.4%	—	—
	MME	BIAS	−0.0562	0.0031	0.0051	−0.0469	−0.0395
		RMSE	0.4613	0.0316	0.0380	0.3840	0.3000
		CP	94.7%	94.6%	95.3%	—	—
	NAE	BIAS	2.3406	−0.1010	−0.1503	—	—
		RMSE	2.3638	0.1032	0.1523	—	—
		CP	0%	0.5%	0%	—	—
1,000	IVE	BIAS	−0.0026	0.0015	0.0002	−0.0636	−0.0730
		RMSE	0.4041	0.0278	0.0340	0.6008	0.4285
		CP	94.3%	95.1%	95.7%	—	—
	MME	BIAS	−0.0046	0.0011	0.0011	−0.0457	−0.0430
		RMSE	0.3326	0.0219	0.0271	0.2397	0.2019
		CP	95.8%	95.7%	96.3%	—	—
	NAE	BIAS	2.3615	−0.1019	−0.1516	—	—
		RMSE	2.3725	0.1029	0.1525	—	—
		CP	0%	0%	0%	—	—

as the sample size increased. For the regression coefficients belonging to $\boldsymbol{\alpha}$, the IVE had larger RMSE than the MME for any given sample size; again, this result is likely due to use of the split samples. In other simulation studies, which we have not reported here, where the IVE was estimated using the full sample, the IVE had smaller BIAS and RMSE than the MME in all cases. Notice, also, that with respect to the estimated measurement error variances the MME generally exhibited smaller BIAS and RMSE than the IVE. Finally, the estimated 95%

TABLE 3: Analysis results using IVE for the health survey data collected from 200 residents of Tianjin.

Parameter	Estimate	95% CI	Parameter	Estimate	95% CI
α_1	64.847	(62.823, 66.871)	α_5	−0.257	(−0.289, −0.225)
α_2	0.234	(0.077, 0.392)	α_6	−0.067	(−0.074, −0.060)
α_3	−1.158	(−1.219, −1.098)	α_7	0.350	(0.315, 0.385)
α_4	0.627	(0.583, 0.671)	α_8	−0.722	(−0.761, −0.683)

coverage probabilities of both IVE and MME were close to the nominal value, while those of the naive estimator were markedly smaller and quickly decreased to zero as sample size increased. Presumably this behaviour is because the naive estimator converged to a different value than the true parameter value.

To examine the sensitivity of our proposed estimators with respect to the normality assumption for measurement error U_i, and for the instruments Z_i in the MME, we also considered various scenarios where either U_i or Z_i were generated from a nonnormal distribution. As an example, in Table 2 we report our numerical results for the scenario $U_1, U_2 \sim t(5)$ independently and $\delta_1, \delta_2 \sim U(-1.5, 1.5)$ independently. The various numerical results exhibited overall patterns that were similar to those found in Table 1. When only Z_i was nonnormal, the MME generated larger BIAS but approximately the same RMSE, while the IVE was unaffected, an observed result that is in line with existing theory. When only U_i was nonnormal, both IVE and MME produced slightly larger BIAS and RMSE. However, when both U_i and Z_i were nonnormal, both IVE and MME exhibited noticeably larger BIAS but only slightly increased RMSE. Therefore, in terms of RMSE, both estimators showed a relatively high degree of robustness against departures from the normal distribution in U_i or in Z_i. The R source code that we used to compute the results found in Tables 1 and 2 are included in the accompanying Supplementary Materials.

5. APPLICATION TO HEALTH SURVEY DATA

We also used our proposed IVE to analyze a dataset from the 2013 national public health survey of China. The main purpose of the survey was to understand the health status and public health care service in both the urban and rural populations. Our dataset consisted of the results from 200 questionnaires sampled from Tianjin. The response variable is the individual respondent's self-assessment of their health status on the five-point scale 0, 1, 2, 3, 4 representing responses ranging from poor (0) to excellent (5). Following the literature we included seven predictor variables in our model: X_1, logarithmic family income; X_2, age in years; X_3, marital status (1 if married and 0 otherwise); X_4, residence type (1 for rural residence and 0 otherwise); X_5, education completed (1 for no formal schooling, 2 for elementary school, 3 for middle school, 4 for high school, 5 for technical school, 6 for secondary vocational school, 7 for college and 8 for university); X_6, current employment status (1 if employed and 0 otherwise); and X_7, chronic disease status (1 if present and 0 otherwise). We assumed that the family income variable, X_1, was measured with error and used the logarithm of family expenditures on culture and entertainment (Z_1), education (Z_2) and health care (Z_3), respectively, as instrumental variables.

The estimated effects of these various factors and the corresponding 95% confidence intervals for the regression coefficients are reported in Table 3. As we expected, health status was positively associated with family income, marital and employment status, and negatively associated with age and chronic disease status. In addition, urban residents tended to feel healthier than rural residents. Our data set and the R source code that we used to derive the estimates summarized in Table 3 may be found in the associated Supplementary Materials.

6. CONCLUSIONS

Although the problem of measurement error in binary response models has been extensively studied in the literature, studies of the same problem in more general categorical response models appear to be rare. We have proposed an approach involving the use of IVs that leads to consistent estimators for an ordered probit model with covariate measurement error. One estimator is based on the use of maximum likelihood, while an alternative approach, which is computationally simpler, is based on the method of moments. Both estimators are consistent and asymptotically normally distributed under general conditions. Our simulation studies show that these methods are effective in correcting for the bias that is caused by measurement error. Moreover, the proposed estimators performed well for a sample size of 200 or more, and they appear to be robust against departures from the assumption of a normal distribution for the measurement error. An application of these methods to the analysis of self-reported health survey data further demonstrated the usefulness of our proposed methods. In order to obtain an explicit formula for the asymptotic covariance matrix of the likelihood-based estimator, we used a split-sample strategy which reduced the efficiency of the estimator. More efficient estimators can be computed using the full sample, but extra efforts such as the use of the bootstrap would be required to derive the variance of our IV estimator.

APPENDIX

The following is a summary of assumptions that are sufficient to prove Theorem 1.

A1. The random error $\epsilon \sim N(0,1)$ and is independent of $\boldsymbol{X}^*$ and $\boldsymbol{X}$.
A2. The measurement error $\boldsymbol{U} \sim N(0, \boldsymbol{\Sigma}_{uu})$ and is independent of $\boldsymbol{X}^*$, $\boldsymbol{X}$ and ϵ.
A3. The instrumental variable $\boldsymbol{Z}$ is correlated with $\boldsymbol{X}^*$ but independent of $\boldsymbol{U}$ and ϵ.
A4. The matrices $\boldsymbol{\Sigma}_{zx^*}$ and $\boldsymbol{\Sigma}_{zz}$ have full column rank.
A5. The observed data $Y_i, \boldsymbol{X}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i, i = 1, 2, \ldots, n$ are independent and identically distributed.

We begin by first establishing Equation (7). Since $\boldsymbol{U}$ is independent of $\boldsymbol{X}$ and ϵ, by Equations (1) and (2) it follows that

$$\boldsymbol{\Sigma}_{wy^*} = \boldsymbol{\Sigma}_{x^*x^*}\boldsymbol{\alpha}_2 + \boldsymbol{\Sigma}_{x^*x}\boldsymbol{\alpha}_3 \tag{A.1}$$

and

$$\begin{aligned}\sigma^2_{y^*} &= \boldsymbol{\alpha}_2'\boldsymbol{\Sigma}_{x^*x^*}\boldsymbol{\alpha}_2 + 2\boldsymbol{\alpha}_2'\boldsymbol{\Sigma}_{x^*x}\boldsymbol{\alpha}_3 + \boldsymbol{\alpha}_3'\boldsymbol{\Sigma}_{xx}\boldsymbol{\alpha}_3 + \sigma^2_\epsilon \\ &= \boldsymbol{\alpha}_2'\boldsymbol{\Sigma}_{wy^*} + \boldsymbol{\alpha}_2'\boldsymbol{\Sigma}_{x^*x}\boldsymbol{\alpha}_3 + \boldsymbol{\alpha}_3'\boldsymbol{\Sigma}_{xx}\boldsymbol{\alpha}_3 + 1.\end{aligned} \tag{A.2}$$

Next, since Y^* and $\boldsymbol{W}$ are jointly normal, by direct integration we obtain

$$\boldsymbol{\Sigma}_{wy^*} = \sigma_{y^*}\rho\boldsymbol{\Sigma}_{wy}, \tag{A.3}$$

where $\rho = \left\{\sum_{j=1}^{J-1} \phi[\Phi^{-1}(p_j)]\right\}^{-1}$ and

$$p_j = P(Y^* < t_j) = P(Y \le j-1),$$

$j = 1, 2, \ldots, J-1$. See also Equation (12) in Olsson, Drasgow & Dorans (1982). Substituting Equation (A.3) into Equation (A.2) and applying the transformation specified in Equation (5) of the main article yields the result

$$\sigma^2_{y^*} = \sigma_{y^*}\sigma_v\rho\boldsymbol{\Sigma}_{wy}'\boldsymbol{M}\boldsymbol{\gamma}_2 + \sigma^2_v\boldsymbol{\gamma}_2'\boldsymbol{M}'\boldsymbol{\Sigma}_{wx}\boldsymbol{\gamma}_3 + \sigma^2_v\boldsymbol{\gamma}_3'\boldsymbol{\Sigma}_{xx}\boldsymbol{\gamma}_3 + 1. \tag{A.4}$$

On the other hand, from Equation (4) in the main article we have

$$\sigma_{y*}^2 = \sigma_v^2\left(\boldsymbol{\gamma}_2'\boldsymbol{\Sigma}_{zz}\boldsymbol{\gamma}_2 + 2\boldsymbol{\gamma}_2'\boldsymbol{\Sigma}_{zx}\boldsymbol{\gamma}_3 + \boldsymbol{\gamma}_3'\boldsymbol{\Sigma}_{xx}\boldsymbol{\gamma}_3 + 1\right). \tag{A.5}$$

Substituting Equation (A.5) into Equation (A.4) and solving the resulting expression for σ_v^2 proves the required result identified in Equation (7).

Proof of Theorem 1. We use the delta-method to derive the asymptotic covariance matrix for the estimator of $\boldsymbol{\alpha}$, since it is easy to see from Equations (5)–(8) that $\boldsymbol{\alpha}$ is continuously differentiable with respect to $\boldsymbol{\psi}$ and $\boldsymbol{\gamma}$.

First, since $\boldsymbol{\alpha} = \sigma_v \boldsymbol{L}\boldsymbol{\gamma}$ depends on $\boldsymbol{\psi}$ only through σ_v, we have

$$\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\psi}'} = \boldsymbol{L}\boldsymbol{\gamma}\frac{\partial \sigma_v}{\partial \boldsymbol{\psi}'}. \tag{A.6}$$

Furthermore, since

$$\frac{\partial \rho}{\partial p_j} = \frac{\Phi^{-1}(p_j)}{\left(\sum_{j=1}^{J-1}\phi[\Phi^{-1}(p_j)]\right)^2} = \rho^2\Phi^{-1}(p_j),$$

we have

$$\frac{\partial \sigma_v}{\partial p_j} = \frac{\sigma_v^3}{2}\eta\boldsymbol{\gamma}_2'\boldsymbol{M}'\boldsymbol{\Sigma}_{wy}\frac{\partial \rho}{\partial p_j} = \frac{\sigma_v^3}{2}\eta\rho^2\Phi^{-1}(p_j)\boldsymbol{\gamma}_2'\boldsymbol{M}'\boldsymbol{\Sigma}_{wy}. \tag{A.7}$$

Similarly, since $\boldsymbol{\Sigma}_{wy} = \boldsymbol{\mu}_{wy} - \boldsymbol{\mu}_w\mu_y$, the partial derivatives of σ_v with respect to $\boldsymbol{\mu}_{wy}'$ and μ_y are

$$\frac{\partial \sigma_v}{\partial \boldsymbol{\mu}_{wy}'} = \frac{\sigma_v^3}{2}\eta\rho\boldsymbol{\gamma}_2'\boldsymbol{M}' \tag{A.8}$$

and

$$\frac{\partial \sigma_v}{\partial \boldsymbol{\mu}_y} = -\frac{\sigma_v^3}{2}\eta\rho\boldsymbol{\gamma}_2'\boldsymbol{M}'\boldsymbol{\mu}_w, \tag{A.9}$$

respectively. It follows from Equations (A.6)–(A.9) that the partial derivative of $\boldsymbol{\alpha}$ with respect to $\boldsymbol{\psi}'$, evaluated at the true parameter value, is

$$\begin{aligned}
\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\psi}'} &= \frac{\sigma_v^3\eta\rho}{2}\boldsymbol{L}\boldsymbol{\gamma}\left(\rho\Phi^{-1}(p_1)\boldsymbol{\gamma}_2'\boldsymbol{M}'\boldsymbol{\Sigma}_{wy}, \ldots, \rho\Phi^{-1}(p_{J-1})\boldsymbol{\gamma}_2'\boldsymbol{M}'\boldsymbol{\Sigma}_{wy}, \boldsymbol{\gamma}_2'\boldsymbol{M}', -\boldsymbol{\gamma}_2'\boldsymbol{M}'\boldsymbol{\mu}_w\right)\\
&= \frac{\sigma_{y*}\rho\boldsymbol{\alpha}}{2}\left(\rho\Phi^{-1}(p_1)\boldsymbol{\alpha}_2'\boldsymbol{\Sigma}_{wy}, \ldots, \rho\Phi^{-1}(p_{J-1})\boldsymbol{\alpha}_2'\boldsymbol{\Sigma}_{wy}, \boldsymbol{\alpha}_2', -\boldsymbol{\alpha}_2'\boldsymbol{\mu}_w\right)\\
&= \boldsymbol{A}.
\end{aligned}$$

Now, the partial derivative of $\boldsymbol{\alpha}$ with respect to $\boldsymbol{\gamma}'$ is

$$\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\gamma}'} = \sigma_v\boldsymbol{L} + \boldsymbol{L}\boldsymbol{\gamma}\frac{\partial \sigma_v}{\partial \boldsymbol{\gamma}'}. \tag{A.10}$$

Also, $\partial\sigma_v/\partial\gamma_1 = 0$,

$$\frac{\partial \sigma_v}{\partial \boldsymbol{\gamma}_2'} = -\frac{\sigma_v^3}{2}\left\{2\boldsymbol{\gamma}_2'\boldsymbol{\Sigma}_{zz} + 2\boldsymbol{\gamma}_3'\boldsymbol{\Sigma}_{xz} - \boldsymbol{\gamma}_3'\boldsymbol{\Sigma}_{xw}\boldsymbol{M} - \eta\rho\boldsymbol{\Sigma}_{wy}'\boldsymbol{M} - \eta^{-1}\rho\boldsymbol{\Sigma}_{wy}'\boldsymbol{M}\boldsymbol{\gamma}_2(\boldsymbol{\gamma}_2'\boldsymbol{\Sigma}_{zz} + \boldsymbol{\gamma}_3'\boldsymbol{\Sigma}_{xz})\right\}$$

and

$$\frac{\partial \sigma_v}{\partial \boldsymbol{\gamma}_3'} = -\frac{\sigma_v^3}{2}\left\{2\boldsymbol{\gamma}_2'\boldsymbol{\Sigma}_{zx} - \boldsymbol{\gamma}_2'\boldsymbol{M}'\boldsymbol{\Sigma}_{wx} - \eta^{-1}\rho\boldsymbol{\Sigma}_{wy}'\boldsymbol{M}\boldsymbol{\gamma}_2(\boldsymbol{\gamma}_2'\boldsymbol{\Sigma}_{zx} + \boldsymbol{\gamma}_3'\boldsymbol{\Sigma}_{xx})\right\}.$$

Evaluated at the true values of the model parameters, the above partial derivatives are

$$\frac{\partial \sigma_v}{\partial \boldsymbol{\gamma}_2'} = -\frac{\sigma_v^2}{2}\boldsymbol{d_1} \tag{A.11}$$

and

$$\frac{\partial \sigma_v}{\partial \boldsymbol{\gamma}_3'} = -\frac{\sigma_v^2}{2}\boldsymbol{d_2}. \tag{A.12}$$

It follows from Equations (A.10)–(A.12) that the partial derivative of $\boldsymbol{\alpha}$ with respect to $\boldsymbol{\gamma}'$, evaluated at the true values of the parameters, is

$$\begin{aligned}\frac{\partial \boldsymbol{\alpha}}{\partial \tilde{\boldsymbol{\gamma}}'} &= \sigma_v \boldsymbol{L} - \frac{\sigma_v^2}{2}\boldsymbol{L}\boldsymbol{\gamma}\left(0, \boldsymbol{d_1}, \boldsymbol{d_2}\right)\\ &= \boldsymbol{B}.\end{aligned}$$

Finally, Equation (9) in the main article can be written as

$$\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = \boldsymbol{A}\sqrt{\frac{n}{n_1}}\sqrt{n_1}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) + \boldsymbol{B}\sqrt{\frac{n}{n-n_1}}\sqrt{n-n_1}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) + o_p(1)$$

and the required result follows from the central limit theorem. ■

Proof of the MME. The model specified in Equations (1) and (2) of the main article can be written as

$$Y^* = \alpha_1 + \tilde{\alpha}_2'\tilde{X} + \epsilon \tag{A.13}$$

and $\tilde{W} = \tilde{X} + \tilde{U}$.

First, by Equation (4) of the main article we have $\boldsymbol{\Sigma}_{zy^*} = \boldsymbol{\Sigma}_{zz}\boldsymbol{\gamma}_2$ and by Equation (5) it follows that $\boldsymbol{\alpha}_2 = \boldsymbol{M}\boldsymbol{\gamma}_2 = \boldsymbol{M}\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{zy^*}$. Then similar to Equation (A.3) above, under normality we have $\boldsymbol{\Sigma}_{zy^*} = \rho\sigma_{y^*}\boldsymbol{\Sigma}_{zy}$, which implies Equation (15) in the main article. Thus, substituting Equation (15) into Equation (A.2) leads to the result

$$\sigma_{y^*}^2 = \rho\sigma_{y^*}\boldsymbol{\Sigma}_{wy}'\boldsymbol{\alpha}_2 + 1 = \rho^2\sigma_{y^*}^2\boldsymbol{\Sigma}_{wy}'\boldsymbol{M}\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{zy} + 1, \tag{A.14}$$

which implies Equation (17) in the main article. Finally, because $p_1 = P(Y^* < t_1) = \Phi(\frac{t_1-\mu_{y^*}}{\sigma_{y^*}})$, we have $\mu_{y^*} = t_1 - \sigma_{y^*}\Phi^{-1}(p_1)$. On the other hand, by Equations (1) and (2), $\mu_{y^*} = \alpha_1 + \boldsymbol{\alpha}_2'\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_x = \boldsymbol{\mu}_w$, which establishes Equation (16), thereby proving the required result concerning the MME. ■

Proof of the estimator of the ME variance. First, by Equation (A.1) we have

$$\boldsymbol{\Sigma}_{wy^*} = \boldsymbol{\Sigma}_{x^*x^*}\boldsymbol{\alpha}_2 + \boldsymbol{\Sigma}_{wx}\boldsymbol{\alpha}_3 = \boldsymbol{\Sigma}_{ww}\boldsymbol{\alpha}_2 - \boldsymbol{\Sigma}_{uu}\boldsymbol{\alpha}_2 + \boldsymbol{\Sigma}_{wx}\boldsymbol{\alpha}_3$$

and therefore by Equation (A.3)

$$\begin{aligned}\boldsymbol{\Sigma}_{uu}\boldsymbol{\alpha}_2 &= \boldsymbol{\Sigma}_{ww}\boldsymbol{\alpha}_2 + \boldsymbol{\Sigma}_{wx}\boldsymbol{\alpha}_3 - \boldsymbol{\Sigma}_{wy^*} \\ &= \boldsymbol{\Sigma}_{ww}\boldsymbol{\alpha}_2 + \boldsymbol{\Sigma}_{wx}\boldsymbol{\alpha}_3 - \rho\sigma_{y^*}\boldsymbol{\Sigma}_{wy}.\end{aligned}$$

Then Equation (18) follows simply by multiplying both sides of this equation by $\boldsymbol{e}_i'$. Then, using Equations (A.2) and (A.3) it follows that

$$\sigma_{y^*}^2 = \sigma_{y^*}\rho\boldsymbol{\Sigma}_{wy}'\boldsymbol{\alpha}_2 + \boldsymbol{\alpha}_2'\boldsymbol{\Sigma}_{wx}\boldsymbol{\alpha}_3 + \boldsymbol{\alpha}_3'\boldsymbol{\Sigma}_{xx}\boldsymbol{\alpha}_3 + 1.$$

Solving this equation for σ_{y^*} gives Equation (19), which proves the required result. ■

ACKNOWLEDGEMENTS

We thank an Associate Editor and two referees for their comments. We are particularly grateful to the Editor for her constructive and helpful comments and suggestions that improved previous versions of this article. The research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and Tianjin University.

BIBLIOGRAPHY

Abarin, T. & Wang, L. (2012). Instrumental variable approach to covariate measurement error in generalized linear models. *Annals of the Institute of Statistical Mathematics*, 64, 475–493.

Agresti, A. (2002). *Categorical Data Analysis*, 2nd ed., John Wiley & Sons, New York.

Agresti, A. (2010). *Analysis of Ordinal Categorical Data*, 2nd ed., John Wiley & Sons, New York.

Baetschmann, G., Staub, K. E., & Winkelmann, R. (2015). Consistent estimation of the fixed effects ordered logit model. *Journal of the Royal Statistical Society Series A*, 178, 685–703.

Bollen, K. A. (2012). Instrumental variables in sociology and the social sciences. *Annual Review of Sociology*, 38, 37–72.

Buzas, J. S. & Stefanski, L. A. (1996a). Instrumental variable estimation in generalized linear measurement error models. *Journal of the American Statistical Association*, 91, 999–1006.

Buzas, J. S. & Stefanski, L. A. (1996b). Instrumental variable estimation in a probit measurement error model. *Journal of Statistical Planning and Inference*, 55, 47–62.

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd ed., Chapman & Hall, London.

Carroll, R. J. & Stefanski, L. A. (1990). Approximate quasi-likelihood estimation in models with surrogate predictors. *Journal of the American Statistical Association*, 85, 652–663.

Chen, Z., Yi, G. Y., & Wu, C. (2014). Marginal analysis of longitudinal ordinal data with misclassification in both response and covariates. *Biomedical Journal*, 56, 69–85.

Cook, J. R. & Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, 89, 1314–1328.

Cranfield, J. A. L. & Magnusson, E. (2003). Canadian consumer's willingness-to-pay for pesticide free food products: An ordered probit analysis. *International Food and Agribusiness Management Review*, 6, 13–30.

Greene, W. H. (2012). *Econometric Analysis*, 7th ed., Prentice Hall, Boston.

Huang, Y. J. & Wang, C. Y. (2001). Consistent functional methods for logistic regression with errors in covariates. *Journal of the American Statistical Association*, 96, 1469–1482.

Li, H. & Wang, L. (2012). Consistent estimation in generalized linear mixed models with measurement error. *Journal of Biometrics and Biostatistics*, S7, 007, 10.4172/2155-6180.S7-007.

Li, T. & Hsiao, C. (2004). Robust estimation of generalized linear models with measurement errors. *Journal of Econometrics*, 118, 51–65.

Liu, I. & Agresti, A. (2005). The analysis of ordered categorical data: An overview and a survey of recent developments. *Test*, 14, 1–73.

Long, S. J. (1997). Regression models for categorical and limited dependent variables. *Advanced Quantitative Techniques in the Social Sciences*, Vol. 7, Sage Publications, Thousand Oaks, CA.

Lu, T. -Y., Poon, W. -Y., & Cheung, S. H. (2015). Multiple comparisons with a control for a latent variable model with ordered categorical responses. *Statistical Methods in Medical Research*, 24, 949–967.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society Series B*, 42, 109–142.

Mohanty, M. S. (2014). What determines happiness? income or attitude: Evidence from the U.S. longitudinal data. *Journal of Neuroscience, Psychology, and Economics*, 7, 80–102.

Olsson, U., Drasgow, F., & Dorans, N. J. (1982). The polyserial correlation coefficient. *Psychometrika*, 47, 337–347.

Peyhardi, J., Trottier, C., & Guedon, Y. (2015). A new specification of generalized linear models for categorical responses. *Biometrika*, 102, 889–906.

Stefanski, L. A. & Buzas, J. S. (1995). Instrumental variable estimation in binary regression measurement error models. *Journal of the American Statistical Association*, 90, 541–550.

Tosteson, T. D., Stefanski, L. A., & Schafer, D. W. (1989). A measurement-error model for binary and ordinal regression. *Statistics in Medicine*, 8, 1139–1147.

Varin, C. & Czado, C. (2010). A mixed autoregressive probit model for ordinal longitudinal data. *Biostatistics*, 11, 127–138.

Wang, L. & Hsiao, C. (2007). Two-stage estimation of limited dependent variable models with errors-in-variables. *Econometrics Journal*, 10, 426–238.

Wang, L. & Hsiao, C. (2011). Method of moments estimation and identifiability of semiparametric nonlinear errors-in-variables models. *Journal of Econometrics*, 165, 30–44.

Xu, K., Ma, Y., & Wang, L. (2015). Instrument assisted regression for errors in variables models with binary response. *Scandinavian Journal of Statistics*, 42, 104–117.

Yi, G. Y. (2017). *Statistical Analysis with Measurement Error or Misclassification—Strategy, Method and Application*. Springer-Verlag, New York.

Zhu, Q., Wang, L., Tannenbaum, S., Ricci, A., DeFusco, P., & Hegde, P. (2014). Pathologic response prediction to neoadjuvant chemotherapy utilizing pretreatment near-infrared imaging parameters and tumor pathologic criteria. *Breast Cancer Research*, 16, 456. http://dx.doi.org/10.1186/s13058-014-0456-0.

Received 12 March 2016
Accepted 21 April 2019