Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Statistical inference for the unbalanced two-way error component regression model with errors-in-variables

Lili Yue [a,b], Gaorong Li [b,*], Junhua Zhang [c]

[a] *College of Applied Sciences, Beijing University of Technology, Beijing 100124, China*
[b] *Beijing Institute for Scientific and Engineering Computing, Beijing University of Technology, Beijing 100124, China*
[c] *College of Mechanical Engineering, Beijing Information Science & Technology University, Beijing 100192, China*

ARTICLE INFO

Article history:
Received 27 October 2016
Accepted 1 June 2017
Available online xxxx

AMS 2000 subject classifications:
primary 62F03
secondary 62F12

Keywords:
Errors-in-variables
Unbalanced two-way error component regression model
Bias-corrected technique
Restricted condition

ABSTRACT

In this paper, we investigate the estimation and testing problems of unbalanced two-way error component regression model with errors-in-variables. The estimation of the unknown parameter is given based on the bias-corrected technique, and the asymptotic property of the resulting estimator is shown under some regularity conditions. For the hypothesis testing problem of restricted condition, a test statistic is constructed through the difference of the corrected residual sums of squares under the null and alternative hypotheses. Then we develop an adjusted test statistic, which follows an asymptotically standard Chi-squared distribution. Some simulation studies and a real data analysis are carried out to assess the performance of the proposed methods.

1. Introduction

The two-way error component regression model has received much attention in statistics. As pointed out by Yue, Shi, and Song (2017), the two-way error component regression model is a special mixed model, which has been widely used in econometrics, computer science, market researches and regional economic surveys, etc. The model takes the following form

$$y_{it} = \beta_0 + x_{it}^{\mathrm{T}}\beta + \mu_i + \nu_t + \varepsilon_{it}, \quad i = 1, \ldots, n,\ t = 1, \ldots, T_i, \tag{1.1}$$

where y_{it} is the response variable at the time t on the ith individual, the intercept β_0 and the slope vector $\beta = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ are unknown parameters of interest, x_{it} denotes a $p \times 1$ vector associated with y_{it}, the individual effect μ_i and the time effect ν_t are random variables, ε_{it} is the random error. We assume that ε_{it} is independent identically distributed (i.i.d.) across individuals and time points, with mean zero and finite variance σ_ε^2. If T_i is the same for $i = 1, \ldots, n$, model (1.1) is called the balanced two-way error component regression model and has been studied in some literatures, for example, Fan and Wang (2008), Wu and Li (2014), Yue, Shi, and Li (2017) and Yue, Shi, and Song (2017). In practice, due to the sampling design or missing data, balanced panel data is almost impossible to obtain, and the examples can be found in Baltagi (2008). For the unbalanced Panel data model, readers may refer to Baltagi and Song (2006), Baltagi et al. (2002), Davis (2002), Oya (2004), Wu et al. (2015), and among others.

In practice, however, the covariates cannot often be observed directly, instead, one can obtain their surrogates. If we simply replace the true covariates in some classical regression models with the observed surrogate variables, we will obtain

* Corresponding author.
E-mail addresses: liliyue@emails.bjut.edu.cn (L. Yue), ligaorong@bjut.edu.cn (G. Li), zjh@bjut.edu.cn (J. Zhang).

http://dx.doi.org/10.1016/j.jkss.2017.06.001
1226-3192/

inconsistent estimation and inefficient testing procedures. The presentation of the errors-in-variables (EV) models can be found in Carroll et al. (2006), Cui and Chen (2003), Cui et al. (2004), Fuller (1987), Liu (2011), and Li et al. (2016), etc. In this paper, we consider the two-way error component regression model (1.1), and assume that the explanatory variable x_{it} is measured with additive errors. That is, instead of the true p-dimensional fixed covariate x_{it}, the surrogate variable ω_{it} is observed as

$$\omega_{it} = x_{it} + \delta_{it}, \tag{1.2}$$

where δ_{it} is the measurement error, and is independent of ε_{it}. We assume that δ_{it} is i.i.d., and $E(\delta_{it}) = 0,\ \mathrm{Cov}(\delta_{it}) = \sigma_\delta^2 I_p (1 \le i \le n,\ 1 \le t \le T_i)$. Here and after, σ_δ^2 is assumed to be known, otherwise, valid estimation can be constructed and the specific details can be found in Carroll et al. (2006), Griliches and Hausman (1986), Li et al. (2016), and Xiao et al. (2010). When the time effects do not exist in models (1.1) and (1.2), Shao et al. (2011) proposed the GMM estimation for the unknown parameter vector. If $\sigma_\delta^2 = 0$, Baltagi et al. (2002) provided some estimation procedures for the regression coefficients in model (1.1), for example, least squares estimator, within estimator and generalized least squares estimator; Wu et al. (2015) considered the unbalanced one-way error component regression model, and studied the parameter estimation and testing problem for the existence of individual effects. Since x_{it} in model (1.1) cannot be exactly observed, the existing two methods in Baltagi et al. (2002) and Wu et al. (2015) are not feasible. One purpose of this paper is to use the group method in Shao et al. (2011) and the estimation method in Wu et al. (2015) to study model (1.1) with measurement errors. This extension is rather important, since the time effects and measurement errors are likely to play important roles in practice. In addition, the existence of time effects and errors-in-variables will lead to the complicated techniques to use. Note that the above literatures did not study the restricted estimation problem for the regression coefficients in models (1.1) and (1.2).

In many important statistical applications, the regression parametric vector is constrained by additional restricted condition in linear models. We shall assume that the constraint is of the form

$$H\beta = d, \tag{1.3}$$

where H is a known $k \times p$ matrix with full row rank and d is a $k \times 1$ known vector. For the statistical inference of the EV model under the restriction (1.3), Wei (2012) and Zhang et al. (2011) studied the partially linear varying-coefficient errors-in-variables model, Tang et al. (2013) discussed the linear EV model with random censored data, Yue, Shi, and Li (2017) considered the special case of the model (1.1) with measurement errors. However, the estimation and testing problems of models (1.1) and (1.2) have not been considered under the restricted condition (1.3).

Based on the bias-corrected technique and Lagrange multiplier method, we propose some estimation procedures for the regression coefficient with or without linear constraints in this paper for models (1.1) and (1.2), and study their asymptotic distributions under some regularity conditions. Then, by the difference of the bias-corrected residual sums of squares under the null and alternative hypotheses, we develop a test approach for testing the validity of the restricted condition (1.3), and derive the asymptotic distribution of the proposed test statistic. In the Monte Carlo experiments and real data analysis, we show that the proposed methods work satisfactorily.

In this paper, we assume that n tends to infinity and T_i for $1 \le i \le n$ is fixed, and it has been assumed in many literatures, such as, Baltagi (2008), Wu and Li (2014) and Yue, Shi, and Li (2017). For the sake of statements, we first introduce some notations as follows. A^{T} denotes the transpose of matrix A, $\|A\| = [\mathrm{tr}(A^{\mathrm{T}}A)]^{\frac{1}{2}}$ represents the Frobenius norm of matrix A, $A > 0$ means a positive-definite matrix, $A^{\otimes 2} = AA^{\mathrm{T}}$. The notation "$\xrightarrow{\mathcal{L}}$" stands for convergence in distribution, "$\xrightarrow{P}$" denotes convergence in probability. The rest of this article is organized as follows. In Section 2, the bias-corrected method is proposed, the asymptotic properties of estimators are studied. The testing procedure for the restricted condition is developed in Section 3. Simulation studies are conducted in Section 4 to assess the performance of our proposed methods. The proposed estimation and testing procedures are also applied to a real data example in Section 5. Some concluding remarks are summarized in Section 6. All the proofs of the main results are provided in Appendix.

2. Methodology and asymptotic properties

2.1. Bias-corrected method

Assume that there are L disjoint subsets $\mathcal{G}_1, \ldots, \mathcal{G}_L$ of $\{1, \ldots, n\}$ such that the observed time length is identical for each $i \in \mathcal{G}_l$ with $l = 1, \ldots, L$. That is, the panel dataset is balanced for the subjects with individuals in $\mathcal{G}_l$. In group $\mathcal{G}_l$, the number of individuals is n_l, the time length is T_l. For each group $\mathcal{G}_l$, we rewrite models (1.1) and (1.2) into the vector form as

$$\begin{cases} y_{li} = \beta_0 \mathbf{1}_{T_l} + X_{li}\beta + \mu_{li}\mathbf{1}_{T_l} + \nu_l + \varepsilon_{li}, \\ W_{li} = X_{li} + \Delta_{li}, \quad i \in \mathcal{G}_l, \end{cases} \tag{2.1}$$

where $y_{li} = (y_{li1}, \ldots, y_{liT_l})^{\mathrm{T}}$, $X_{li} = (x_{li1}, \ldots, x_{liT_l})^{\mathrm{T}}$, $x_{lij} = (x_{lij1}, \ldots, x_{lijp})^{\mathrm{T}}$, $j = 1, \ldots, T_l$, $\mathbf{1}_{T_l}$ is the $T_l \times 1$ vector of ones, $\nu_l = (\nu_{l1}, \ldots, \nu_{lT_l})^{\mathrm{T}}$, $\varepsilon_{li} = (\varepsilon_{li1}, \ldots, \varepsilon_{liT_l})^{\mathrm{T}}$, $W_{li} = (\omega_{li1}, \ldots, \omega_{liT_l})^{\mathrm{T}}$, $\omega_{lij} = (\omega_{lij1}, \ldots, \omega_{lijp})^{\mathrm{T}}$, $\Delta_{li} = (\delta_{li1}, \ldots, \delta_{liT_l})^{\mathrm{T}}$, $\delta_{lij} = (\delta_{lij1}, \ldots, \delta_{lijp})^{\mathrm{T}}$.