Journal of the Korean Statistical Society 43 (2014) 79–90

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Equivalence of two tests in varying coefficient partially linear errors in variable model with missing responses

Hu Yang, Xiaochao Xia *

College of Mathematics and Statistics, Chongqing University, 401331, PR China

ARTICLE INFO

Article history:
Received 28 November 2012
Accepted 4 May 2013
Available online 30 May 2013

AMS 2000 subject classifications:
primary 62G05
secondary 62G10

Keywords:
VCPLM
Errors in variables
Missing response
Profile Lagrange multiplier
Corrected residual sum of squares
Wilk's phenomenon

ABSTRACT

This paper considers two tests on varying coefficient partially linear errors-in-variables models (VCPLM-EV) with missing responses under the linear constraint. The restricted estimator for the parametric component is derived and proven to share asymptotically normal distribution. In order to test the linear constraint, two statistics based on the profile Lagrange multiplier method and the corrected residual sum of squares method respectively, are proposed. It is of interest to obtain that the magnitudes of the two statistics are equal exactly and follow the asymptotical chi-square distribution. This reveals a new type of Wilk's phenomenon in VCPLM-EV models with missing response. Finally, some numerical examples are carried out to illustrate relevant performances.

1. Introduction

Over the last two decades, semiparametric modeling has received extensive attention and has been applied to a lot of fields because it is not only capable of avoiding the so-called "curse of dimensionality" of nonparametric modeling, but also of reducing the risk of misspecification of parametric modeling. It can balance the interpretation of linear models and flexibility of nonparametric models, and it covers many specific models, such as partially linear models, additive models, varying coefficient models and their hybrids. Among them, varying-coefficient partially linear model (VCPLM) is an important one having the form

$$Y = Z^{\tau}\boldsymbol{\beta} + X^{\tau}\boldsymbol{\alpha}(T) + \varepsilon \tag{1.1}$$

where Y is the response, and X, Z, T are random covariates. It is usually assumed that T is univariate, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^{\tau}$ is a vector of d-dimensional unknown parameters, $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot), \ldots, \alpha_q(\cdot))^{\tau}$ is a vector of q-dimensional unknown coefficient functions, ε is the random error with conditional zero mean and finite variance given (Z, X, T). As for model (1.1) there is a lot of literature, such as Ahmad, Leelahanon, and Li (2005), Fan and Huang (2005), You and Zhou (2006), Zhang, Lee, and Song (2002) and so forth.

In data analysis, the data of some variables may not be accurately collected. It may be measured with errors, which leads to a class of errors-in-variable (EV) models. Errors-in-variables will result in biased estimate for the parameter in model (1.1). Usually, we consider that the covariate has additive measurement errors, i.e. $W = Z + U$. For this kind of model, You and Chen (2006) proposed the estimation for the parametric component via applying the technique of correction for attenuation.

* Corresponding author. Tel.: +86 13983365416.
E-mail addresses: xia_xiao_chao@126.com, xxc@cqu.edu.cn (X. Xia).

1226-3192/$ – see front matter
http://dx.doi.org/10.1016/j.jkss.2013.05.001

Hu, Wang, and Zhao (2009) further considered the estimation for the parameters and constructed the confidence region by using the empirical likelihood method. Zhao and Xue (2009) applied the empirical likelihood method to analyze longitudinal data. Wei (2012) and Zhang, Li, and Xue (2011) both drew statistical inference for the parametric component under restricted conditions and obtained the same theoretical results. There is a mass of literature on other kinds of EV models, for example, see Fuller (1987) on parametric models, Liang (2000) and Liang, Härdle, and Carroll (1999) on the partially linear model.

On the other hand, missing data may be encountered in practical applications. When response is subject to missingness, there has been extensive research on this aspect. Wang, Linton, and Härdle (2004) proposed a class of semiparametric estimators in partially linear models and made some statistical inference including the semiparametric efficiency, empirical likelihood, adjusted empirical likelihood, and bootstrap empirical likelihood. Wang and Sun (2007) further developed the imputation, semiparametric regression surrogate and inverse marginal probability weighted approaches. Sun, Wang, and Dai (2009) investigated the model checking problem for a partial linear model when some responses are missing at random. More details on applications for the semiparametric model and missing data can be found in the monograph by Tsiatis (2006) and the references therein.

However, most of efforts are devoted to accounting for the data with either covariate measurement error or missing response, but relatively little work has been done to address both simultaneously. This motivates us to consider the following model in the current paper. Suppose that the response variable Y has a missing observation. We introduce an indicator variable δ such that $\delta = 1$ if Y is observable, otherwise $\delta = 0$. Consider the VCPLM-EV model

$$\begin{cases} Y = Z^\tau \boldsymbol{\beta} + X^\tau \boldsymbol{\alpha}(T) + \varepsilon \\ W = Z + U \end{cases} \tag{1.2}$$

with the following missing mechanism

$$\Pr(\delta = 1|Y, Z, X, T) = \Pr(\delta = 1|Z, X, T) = \pi(Z, X, T) \tag{1.3}$$

for some unknown $\pi(Z, X, T)$. Note that Y is missing at random if Z is exactly observed, otherwise it is not missing at random if the surrogate W of covariate Z is used, see Liang, Wang, and Carroll (2007). Recently, some efforts have been made to analyze model (1.2) under the condition (1.3). For example, Liang et al. (2007) investigated the estimation and made an empirical likelihood inference for the parametric component in semiparametric partially linear models. Sun, Zhang, and Du (2012) studied the parametric estimation under monotonicity constrains in the special case where $X \equiv 1$. For semi-varying coefficient partially linear EV models, Wei (2010) investigated the estimates for the parameters via imputation and surrogate methods. Wei and Mei (2012) further made an empirical likelihood inference on the parametric components. In the parametric regression setting, Li, Ma, and Carroll (2012) proposed a functional generalized method of moment approach for the marginal analysis of longitudinal data.

When some prior information on the regression coefficient is available, taking both the missing response variable and the covariate measurement error into account is of interest. Usage of such information may improve upon the efficiency of the estimator. There exist some works on EV models under restricted conditions on regression coefficients, see Zhang et al. (2011) for instance. In this paper, we consider the following constraint

$$\mathbf{A}\boldsymbol{\beta} = \mathbf{d} \tag{1.4}$$

where $\mathbf{A}$ is a $k \times d$ known matrix, and $\text{rank}(\mathbf{A}) = k \leqslant d$, $\mathbf{d}$ is a $k \times 1$ known vector. Here we assume that the measurement error U is independent of (Y, Z, X, T) and with $E(U) = 0$ and $\text{Cov}(U) = \Sigma_{UU}$ in order to make the model identifiable. If Σ_{UU} is unknown, we can take advantage of the technique of repeated measurement proposed by Liang et al. (2007).

Our task is to estimate the parameters and to test whether linear constraints (1.4) are fulfilled under the missing response from the missing mechanism (1.3). In the current paper, the latter is of main interest. The results are summarized as follows. Firstly, a restricted estimator is derived and it is proven to follow the asymptotically normal distribution under the constraints. From our results, we find that the noncentrality parameter η in Theorem 5 of Zhang et al. (2011) is not correct, and we correct it in Remark 2 of Section 3. Secondly, two statistics focused on two different approaches are constructed to test the validity of the restricted conditions. We demonstrate that the magnitudes of these two test statistics are exactly equal and share the same asymptotically chi-squared distribution. The second is the main finding of this paper, which can be extended to common measurement error models. It turns out that the theoretical results are reliable according to our numeric examples.

The rest of this paper is organized as follows. In Section 2, the restricted estimators for the parametric part and nonparametric part are put forward. In Section 3, two statistics are constructed to test the effectiveness. In Section 4, some examples are carried out to illustrate the performance of the proposed estimators with a finite sample. The proofs of results are relegated to Section 5.

2. Restricted estimation

Let $(Y_i, \delta_i, Z_i, X_i, T_i)_{i=1}^n$ be a random sample of incomplete data from model (1.2). Then we have

$$\delta_i Y_i = \delta_i Z_i^\tau \boldsymbol{\beta} + \delta_i X_i^\tau \boldsymbol{\alpha}(T_i) + \delta_i \varepsilon_i. \tag{2.1}$$