ARTICLE IN PRESS
Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

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

Binary response correlated random coefficient panel data models[☆]

Yichen Gao[a], Cong Li[b,c,*], Zhongwen Liang[d]

[a] International School of Economics and Management, Capital University of Economics and Business, Beijing 100070, PR China
[b] School of Economics, Shanghai University of Finance & Economics, 777 Guoding Road, Shanghai 200433, PR China
[c] Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai 200433, PR China
[d] Department of Economics, University at Albany, SUNY, Albany, NY 12222, USA

ARTICLE INFO

Article history:
Available online xxxx

JEL classification:
C14
C33

Keywords:
Random coefficient
Binary choice
Nonparametric
Panel data
Specification test

ABSTRACT

In this paper, we consider binary response correlated random coefficient (CRC) panel data models which are frequently used in the analysis of treatment effects and demand of products. We focus on the nonparametric identification and estimation of panel data models under unobserved heterogeneity which is captured by random coefficients and when these random coefficients are correlated with regressors. Our identification conditions and estimation are based on the framework of the model with a special regressor, which is a novel approach proposed by Lewbel (1998, 2000) to solve the heterogeneity and endogeneity problem in the binary response models. With the help of the additional information on the special regressor, we can transform a binary response CRC model to a linear moment relation. We also construct a semiparametric estimator for the average slopes and derive the $\sqrt{n}$-normality result. Further, we propose a nonparametric method to test the correlations between random coefficients and regressors. Simulations are given to show the finite sample performance of our estimators and test statistics.

1. Introduction

Recently, the correlated random coefficient model has drawn much attention. As stated in Heckman et al. (2010), "The correlated random coefficient model is the new centerpiece of a large literature in microeconometrics". In this paper, we consider binary response CRC panel data models in the form of

$$y_{it} = \mathbf{1}(v_{it}^\top \gamma + x_{it}^\top \beta_i + u_{it} > 0),$$
$$(i = 1, \ldots, n;\ t = 1, \ldots, T) \tag{1.1}$$

where $\mathbf{1}(\cdot)$ is the indicator function, v_{it} denotes regressors with constant coefficient γ, x_{it} denotes regressors with random coefficient β_i, and u_{it} is the error term. x_{it} can include 1 as a component. Thus, the binary choice panel data model with fixed effects is a special case of this model. We allow the general correlation between the random coefficient β_i and the regressors x_{it}. We focus on the nonparametric identification and estimation of the mean of random slope β_i in this model, which will be more specific in later sections.

Binary choice panel data models are widely used by applied researchers. One reason is its direct economic interpretability as the decision making process of individuals. Another reason is that given the advantage of panel data with multiple observations of the same individual over several time periods, it is possible to take into account unobserved heterogeneity. The common approach is to include an individual-specific heterogeneous effect variable additively, which leads to a correlated random effects model or a fixed effects model. The advantage of this approach is that we can eliminate the unobservable variable by taking the difference between different time periods and get the fixed effects estimator for linear models easily, see e.g. Arellano (2003) and Hsiao (2003). This also resolves the incidental parameter problem in linear panel data models. The method of taking difference can also be extended to nonlinear panel data models in certain context, see Bonhomme (2012). Though it is convenient to deal with unobserved heterogeneity additively, economic models imply many different non-additive forms, see Browning and Carro (2007) and Imbens (2007). Among them, one class is the random coefficient model which arises from the demand analysis with the consideration of the individual heterogeneity.

Random coefficient models have the multiplicative individual heterogeneity. They are popular in empirical analysis of

☆ We thank a co-editor, and two anonymous referees for their constructive suggestions which helped us improve this paper substantially.

* Corresponding author at: School of Economics, Shanghai University of Finance & Economics Guoding Road, Shanghai 200433, PR China.
E-mail address: li.cong@mail.shufe.edu.cn (C. Li).

http://dx.doi.org/10.1016/j.jeconom.2015.03.008
0304-4076/

treatment effects and the demand of products. In the analysis of treatment effect, under certain circumstances, the binary choice fixed-effects model can be transformed to a linear random coefficient model with the average treatment effect being the mean of a random coefficient. For instance, in one of the commenting papers for Angrist (2001), Hahn (2001) gives an example on this transformation and discusses the consistency of the fixed effects estimator. Wooldridge (2005) further allows the correlation between regressors and random coefficients and gives the conditions that assure the consistency of the fixed effects estimator. Recently, Hsiao et al. (2012) investigate the identification and estimation of linear correlated random coefficient panel data models, which is closely related with this paper and serves as a basic piece of our method.

In the literature of demand analysis, Berry et al. (1995) propose to use the random coefficients logit multinomial choice model to study the demand of automobiles which has become the major vehicle of the demand analysis. However, they leave the correlation between the random coefficients and the regressors unconsidered, and put assumptions on the functional form of the distributions of the unobservable variables. In this paper, we study random coefficient binary choice models without specifying the functional form of the distribution of unobservable variables. Also, we allow for non-zero correlation between regressors and random coefficients. For simplicity, we only consider binary choice models.

Recently, there is a growing literature on CRC models. Graham and Powell (2012) discuss the identification and estimation of average partial effects in a class of "irregular" correlated random coefficient panel data models using different information of agents from subpopulations, so called "stayers" and "movers". Due to the irregularity, they get an estimator with slower than $\sqrt{n}$ convergence rate and the normal limiting distribution. Also, there is a literature on the testing of the CRC model. Basing on the equivalence of the CRC model and the generalized Roy model, Heckman and Schmierer (2010) derive two testable implications. One is the linearity structure of a conditional moment, and the other is the equivalence of different IV estimators. They construct test statistics and discuss their powers. Heckman, Schmierer, and Urzua (2010) consider similar implications and the tests of the CRC model.

Other related literature includes three aspects: random coefficient models, panel data models with unobserved heterogeneity, and models with a special regressor. All of these literature has been developed considerably in the last two decades. Random coefficient models have a long history. Swamy and Tavlas (2007) and Hsiao and Pesaran (2008) are good surveys for these models. For binary choice random coefficient models, Hoderlein (2009) considers a binary choice model with endogenous regressors under a weak median exclusion restriction. He uses a control function IV approach to identify the local average structural effect of the regressors on the latent variable, and derives $\sqrt{n}$ consistency and the asymptotic distribution of the estimator he proposed. He also proposes tests for heteroscedasticity, overidentification and endogeneity. Some of the recent papers are concerned with the distributions of the random coefficients. For example, Gautier and Kitamura (2013) consider the independent random coefficients binary choice models. Through the analysis of an ill-posed inverse integral equation, they construct an easy to implement nonparametric estimator of density function of the random coefficients using Fourier–Laplace series on spheres. Fox et al. (2011, 2012) discuss the identification and estimation of the random coefficients logit model based on the conditional independence conditions and the parametric distribution assumption of the error term. Other recent papers include Arellano and Bonhomme (2012), Fox and Gandhi (2010) and Hoderlein et al. (2010). All of these papers require independence or conditional independence conditions on the random coefficients.

Among the recent developments of panel data models, the nonseparable panel data models is an indispensable part. Chernozhukov et al. (2009, 2013) investigate quantile and average effects in nonseparable panel models. Evdokimov (2010) discusses the identification and estimation of a nonparametric panel data model with nonseparable unobserved heterogeneity. He obtains point identification and estimation via conditional deconvolution. Hoderlein and White (2012) give nonparametric identification in nonseparable panel data models with generalized fixed effects. Gayle (2013) considers the identification and estimation of a class of single-index panel data models with unknown link function and correlated unobserved individual effects.

The identification of discrete choice model is different from linear models. The framework we adopt in this paper for the identification of the average slope in binary response CRC panel data models is the special regressor method, which assumes the existence of a special regressor with additional information. Proposed by Lewbel (1998, 2000), this method has been exploited extensively in different settings. It is an effective way for identification and estimation under heterogeneity and endogeneity. Honoré and Lewbel (2002) use this method to study a binary choice fixed effects model which allows for general predetermined explanatory variables and give a $\sqrt{n}$ consistent semiparametric estimator. Khan and Lewbel (2007) investigate a truncated regression model using this method and propose a $\sqrt{n}$ consistent and asymptotically normal estimator. Dong and Lewbel (2015) give a good survey for this method.

The rest of the paper is organized as follows. In Section 2, we discuss the identification conditions for the binary response CRC panel data models. The estimator and the $\sqrt{n}$-normality result are given in Section 3. We develop a test statistic to test the correlations between random coefficients and regressors in Section 4. We conduct extensive simulations in Section 5 to show the finite sample advantage of our estimators. Section 6 concludes the paper. All of proofs are relegated to the Appendix.

2. Identification of a binary response CRC panel model

The identification of the binary response model is different from the linear models. We can identify the coefficients if we are willing to assume that the unobserved random terms have known distributions, and this will allow us to estimate the model by conditional maximum likelihood method. However, if we do not assume the distribution of the unobserved terms, the identification becomes problematic. We need to impose additional restrictions on the dependence structure between the regressors and the unobservables.

One way to identify the model is to transform the model to a single-index model, which can be estimated nonparametrically. However, the single-index model only admits limited heterogeneity, see Powell et al. (1989), Ichimura (1993), Klein and Spady (1993), Härdle and Horowitz (1996) and Newey and Ruud (2005).

Another way of identification is based on the conditional quantile restrictions. Manski (1985, 1988) give the identification conditions in this type for the binary response models. A sufficient condition for the identification of the coefficients is the median independence between the error and the regressors. He also suggests the conditional maximum score estimator to estimate the model. However, the limiting distribution is not standard which is derived by Kim and Pollard (1990). Horowitz (1992) modifies the maximum score estimator to a smoothed maximum score estimator and gets the asymptotic normal distribution. The convergence rates of maximum score estimators are less than $\sqrt{n}$. Chamberlain (2010) shows that the consistent estimation at the $\sqrt{n}$ convergence rate is possible only when the errors have logistic distributions without other additional assumptions.

Please cite this article in press as: Gao, Y., et al., Binary response correlated random coefficient panel data models. Journal of Econometrics (2015), http://dx.doi.org/10.1016/j.jeconom.2015.03.008