Journal of Choice Modelling 28 (2018) 108–123

Contents lists available at ScienceDirect

Journal of Choice Modelling

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

Estimation of an unbalanced panel data Tobit model with interactive effects

Xiaoqing Ye [a], Juan Xu [b,*], Xiangjun Wu [c]

[a] School of Mathematics and Statistics, South-Central University for Nationalities, Wuhan, PR China
[b] School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan, PR China
[c] School of Economics, Hangzhou Dianzi University, Hangzhou, PR China

ARTICLE INFO

Keywords:
Panel data Tobit model
Interactive effects
Missing data
Maximum likelihood
ECM

ABSTRACT

Unbalanced panel data or panel data with missing observations are common in empirical research. In this paper, we consider an unbalanced panel data Tobit model with interactive effects, and provide an estimator based on the iteration of Tobit factor analysis and maximum likelihood estimation. Monte Carlo studies are carried out to investigate the finite sample performance of the proposed method in comparison with other candidate methods. The results show that the finite sample performance of the proposed method is satisfactory under different Mont Carlo designs. We also apply our method to study female labor supply using an unbalanced panel data set from the Chinese Family Panel Studies (CFPS).

1. Introduction

The Tobit model proposed by Tobin (1958) has been widely used in empirical studies in a variety of fields such as labor and industrial organization. A key feature of the model is that the dependent variable can be observed only in a limited range. Generally, the parameters of the Tobit model can be estimated by maximum likelihood (ML) (see e.g. Amemiya, 1985) or by two-step methods (see e.g. Heckman, 1979). For cross sectional data, conventional Tobit model has been well studied; see e.g. Amemiya (1973). Multivariate Tobit models, which consider both censoring and simultaneity, are also fully discussed; see e.g. Amemiya (1974), Nelson and Olsen (1978), and Chen and Zhou (2011). Besides, some authors introduce spatial correlation into Tobit models; see e.g. Flores-Lagunes and Schnier (2012), Qu and Lee (2012, 2013), Xu and Lee (2015). In time series setting, de Jong and Herrera (2011) establish a formal asymptotic theory for the Tobit model.

Recently, as panel data sets have become widely available and very popular, a growing literature on panel data Tobit models has been developed; see e.g. Honoré (1992), Honoré (1993), Kyriazidou (1997), Honoré et al. (2000), Hu (2002), Greene (2004), Honoré and Hu (2004), Li and Zheng (2008), Khan et al. (2016), to name a few.

An advantage of panel data is that researchers can control for unobservable heterogeneity. Most often, the heterogeneity is assumed to enter the panel regression additively (see e.g. Hsiao, 2014). Nowadays, panel data models with interactive effects or cross sectional dependence have attracted considerable interest. Such cross-correlations can arise for a variety of reasons, such as omitted common factors, spatial spillovers, contagion and so on. In the presence of such dependence, traditional panel estimators such as fixed or random effects can result in misleading inference and even inconsistent estimators (Phillips and Sul, 2003). Andrews (2005) shows that, if regressors are correlated with unobserved common factors that might be causing the error cross sectional dependence, traditional panel

* Corresponding author.
E-mail address: xujuan@zuel.edu.cn (J. Xu).

https://doi.org/10.1016/j.jocm.2018.04.004
Received 9 May 2017; Received in revised form 29 March 2018; Accepted 25 April 2018

estimators may be inconsistent. Despite these, panel data models with interactive effects are still the models of choice for many theoretical and applied studies. As it allows more flexible modeling of heterogeneity across individuals and over time than traditional models and provides an effective way to model cross sectional dependence, and as Hsiao (2014) shows that this model setup also eliminates the source of bias. In fact, the factor structure of interactive effects $\lambda_i^{'}f_t$ is sufficiently general and renders traditional additive individual and time effects as a special case, for example, let $\lambda_i = (\alpha_i, 1)^{'}, f_t = (1, \theta_t)^{'}$, then $\lambda_i^{'}f_t = \alpha_i + \theta_t$.

There are many examples of panel data models that may require interactive effects or multifactor error structures. In macroeconomics, for example, individual countries' economic growth rates could depend on world-wide supply shocks f_t (such as oil shocks, technology shocks), and the common shocks have heterogeneous impacts across countries through the different factor loadings λ_i. In finance, and especially in asset pricing models assuming time-varying risk premia, f_t is a vector of common factors (such as systematic risk) and λ_i is the exposure to the risks. In labor economics, such as the Mincer equation, λ_i is a vector of unobservable individual traits such as talent, skill and f_t reflects the time-varying payoff to the individual traits. In economics of education, when considering the effect of class size and socioeconomic class composition on educational attainment, λ_i can be associated with motivation and ability to absorb knowledge when the student is listening to lectures or reading, and f_t can be interpreted as measuring teacher quality (Harding and Lamarche, 2014). Moreover, the factor structure provides a tractable way to model cross-section dependence. In the factor structure $\lambda_i^{'}f_t$, each cross section shares the same f_t, thus, cross-correlation emerges.

For linear panel data models with interactive effects, the theoretical properties of the estimators are well established, and the recent literature on the estimation methods is extensive and growing rapidly. There are two main strands of literature for estimating large static panel data models where both cross-section dimension N and time dimension T are large. One strand of the estimation method is the Gaussian quasi-maximum likelihood estimator (QMLE); see Bai (2009), Greenaway-McGrevy et al. (2012), Bai and Li (2014), Moon and Weidner (2015, 2017), Bai et al. (2015). The other strand is the common correlated effects (CCE) estimator; see Pesaran (2006), Kapetanios et al. (2011), Pesaran and Tosetti (2011), Su and Jin (2012), Harding and Lamarche (2014), Chudik and Pesaran (2015), Baltagi et al. (2016), Boneva et al. (2016), Everaert and De Groote (2016). In small T case, the estimation method is mainly generalized method of moments (GMM) (Ahn et al., 2001, 2013; Robertson and Sarafidis, 2015). For the dynamic model, the literature is relatively rare, see Holtz-Eakin et al. (1988), Bai (2013), Lu and Su (2016), Hidalgo and Schafgans (2017) and Shi and Lee (2017). The extension of traditional fixed effects to interactive effects substantially increases the flexibility of controlling for unobserved heterogeneity, but existing estimation approaches are mainly designed for linear models and do not offer the possibility of estimating nonlinear models, which may be of interest to applied researchers.

This paper proposes an unbalanced panel data Tobit regression estimator for a model with interactive effects, allowing the common factors and factor loadings to be arbitrarily correlated with the independent variables. Although much progress has been made in panel data Tobit models, knowledge of these models with interactive effects and missing observations is very limited. Our paper also complements the recent literature on nonlinear panel data models with interactive effects. For example, Chen et al. (2014) develop a comprehensive estimation method for a wide range of nonlinear panel data models, including static and dynamic Probit, Logit and Poisson models. Chen (2016) proposes an expectation-maximization (EM) procedure to estimate panel Probit model with interactive effects. Boneva and Linton (2017) extend the CCE estimator to settings where outcomes are discrete. Moon et al. (2017) incorporate interactive effects into the Berry, Levinsohn and Pakes (Berry et al., 1995) random coefficients discrete choice demand model, and propose a two-step least squares-minimum distance (LS-MD) procedure. Xue et al. (2018) combine the projection method and the special regressor method to estimate a binary choice model with interactive effects.

The literature above on both linear and nonlinear models are almost always assuming balanced panels, i.e. that all individuals in the sample are observed over the same period of time. However, in most practical applications, data sets are almost never balanced and unbalanced panel or missing data is pervasive. For example, due to some economic units electing to drop out of the sampling process, attrition arises; the other case is that individuals do not disappear from the panel but certain variables are unobserved for at least some time periods. Considering this, in this paper, we not only introduce interactive effects into traditional panel data Tobit models, but also consider the effects of missing observations on the consistency of the estimators.

The remainder of this paper is organized as follows. The econometric model is presented in Section 2. Section 3 describes the estimation methodology and discusses related estimators for unbalanced panel data Tobit model with interactive effects. Section 4 reports the finite sample results of Monte Carlo simulations and Section 5 applies our estimator to investigate the female labor supply in Shanghai, China. Section 6 concludes.

2. Model setup

We consider the following panel data Tobit model with N cross-sectional units and T time periods

$$y_{it} = \begin{cases} x_{it}^{'}\beta + u_{it}, & y_{it}^{*} > c \\ c, & y_{it}^{*} \le c \end{cases}, i = 1, 2, \dots, N; t = t_i = t_i(1), t_i(2), \dots, t_i(T_i) \tag{1}$$

$$y_{it}^{*} = x_{it}^{'}\beta + u_{it} \tag{2}$$

$$u_{it} = \lambda_i^{'}f_t + \varepsilon_{it} \tag{3}$$

where y_{it} is the dependent variable, the latent variable y_{it}^{*} is only observed when it is greater than a censoring variable c, x_{it} is a $K \times 1$