Journal of Econometrics 188 (2015) 301–312

Contents lists available at ScienceDirect

Journal of Econometrics

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

Semiparametric single-index panel data models with cross-sectional dependence

Chaohua Dong [a,c], Jiti Gao [a], Bin Peng [b,*]

[a] Monash University, Australia

[b] Economics Discipline Group, University of Technology Sydney, Australia

[c] Southwestern University of Finance and Economics, China

ARTICLE INFO

Article history:
Received 3 March 2014
Received in revised form
27 January 2015
Accepted 6 June 2015
Available online 18 June 2015

JEL classification:
C13
C14
C23
C51

Keywords:
Asymptotic theory
Closed-form estimation
Nonlinear panel data model
Orthogonal series expansion method

ABSTRACT

In this paper, we consider a semiparametric single-index panel data model with cross-sectional dependence and stationarity. Meanwhile, we allow fixed effects to be correlated with the regressors to capture unobservable heterogeneity. Under a general spatial error dependence structure, we then establish some consistent closed-form estimates for both the unknown parameters and the link function for the case where both cross-sectional dimension (N) and temporal dimension (T) go to infinity. Rates of convergence and asymptotic normality are established for the proposed estimates. Our experience suggests that the proposed estimation method is simple and thus attractive for finite-sample studies and empirical implementations. Moreover, both the finite-sample performance and the empirical applications show that the proposed estimation method works well when the cross-sectional dependence exists in the data set.

1. Introduction

Single-index models have been studied by both econometricians and statisticians in the past twenty years or so and cover many classic parametric models (e.g. linear model and logistic model) by using a general function form $g\left(x'\beta\right)$ (e.g. Chapter 2 of Gao (2007)). Recently, researchers start considering single-index models under panel data set-up (c.f. Chen et al., 2013a,b). For most of the published work on semiparametric single-index models, the estimation is based on a nonparametric kernel method, which may be sensitive to initial values due to multi-modality or flatness of a curve in practice. Chen et al. (2013b) use this technique to investigate a partially linear panel data model with fixed effects and cross-sectional independence. In their paper, a consistent parameter estimator is achieved with the rate of convergence $\sqrt{NT}$ (N and T here and hereafter are cross-sectional dimension and temporal dimension, respectively), but, due to the identification requirements, they have to impose extra restrictions on the fixed effects. Alternatively, one can use sieve estimation techniques to implement a two-step procedure (e.g. profile method or iterative method).

To the best of our knowledge, consistent closed-form estimates have not been established for this type of semiparametric single-index model in the literature. In this paper, we aim at establishing consistent closed-form estimates for a semiparametric single-index panel data model with both cross-sectional dependence and stationarity for the case where both N and T go to infinity. The estimation procedure proposed below allows us to avoid certain computational issues and is therefore easy to implement. The estimation techniques proposed in this paper can also be extended to the multi-factor structure model. For example, under certain restrictions similar to those in Su and Jin (2012), a semiparametric single-index extension of Pesaran (2006) can be achieved. Furthermore, we add fixed effects to the model and do not impose any particular assumptions on them, and therefore they can be correlated with the regressors to capture unobservable heterogeneity. Compared to Chen et al. (2013b), our set-up is more flexible on the fixed effects.

In this paper, we assume that all the regressors and error terms can be cross-sectionally correlated. As covered in Assumption 1 of

* Corresponding author.
E-mail address: Bin.Peng@uts.edu.au (B. Peng).

http://dx.doi.org/10.1016/j.jeconom.2015.06.001
0304-4076/

Section 3, we impose a general spatial correlation structure to link the cross-sectional dependence and stationary mixing condition together. As a result, some types of spatial error correlation can easily be covered by the assumptions given in Section 3 (c.f. Chen et al., 2012a,b). This set-up is more flexible than that considered by Chen et al. (2013b). As Andrews (2005) and Bai (2009) discuss, common shocks (e.g. global finance crisis) exist in many economic phenomena and cause serious forecasting biases, and an important characteristic is that they induce a correlation among individuals. Thus, it is vital for us to have such models that can capture this type of "global" cross-sectional dependence.

We specifically use Hermite polynomial series orthogonal expansion to estimate the unknown link function. This technique has been widely used in the econometrical literature, for example, by Gallant and Nychka (1987) and Gallant and Tauchen (1989), among others. In general, this method is referred to as sieve estimation. Very detailed literature reviews on the Hermite polynomials and the sieve estimation can be seen in Nevai (1986) and Chen (2007) respectively. However, a further technique gadget on the Hermite polynomials brings a benefit to the expansion of the link function (see Lemma B.1 in Appendix). As a result, a semiparametric model is rephrased completely as a parametric model such that we are able to derive a closed-form estimate for the index vector, and because of this the requirement for the parameter space is lowered to the minimum level, comparing with the profile method where usually θ_0 is stipulated to be an inner point of a compact convex set Θ.

In summary, this paper makes the following contributions:

1. it proposes a semiparametric single-index panel data model to simultaneously accommodate cross-sectional dependence, stationarity and unobservable heterogeneity;
2. it establishes simple and consistent closed-form estimates for the unknown index vector, and consequently there is no restriction on the parameter space;
3. it establishes both rates of convergence and asymptotic normality results for the estimates under a general spatial error dependence structure; and
4. it evaluates the proposed estimation method through the use of both simulated and real data examples.

The structure of this paper is as follows. Section 2 proposes our model and discusses the main idea. Section 3 constructs a closed-form estimate for a vector of unknown parameters of interest and introduces assumptions for the establishment of asymptotic consistency and normality results. In Section 4, we recover the unknown link function and evaluate the rate of convergence. In Section 5, we do Monte Carlo experiments which particularly verify whether the fixed effect dependence affects the proposed estimation and provide an empirical case study by looking into the demand of the United States (US) for cigarettes. Section 6 concludes this paper with some comments. The key proofs are given in Appendix B. Some other proofs, verifications and relevant discussions are given in a supplementary document of this paper (see Appendix C).

Throughout the paper, we will use the following notation: $\otimes$ denotes the Kronecker product; I_k denotes the identity matrix with dimension k; i_k signifies the $k\times 1$ vector $(1,\ldots,1)'$; $M_P = I_k - P(P'P)^{-1}P'$ denotes the projection matrix generated by matrix $P = P_{k\times l}$ with full column rank; A^- denotes the Moore–Penrose inverse of the matrix A; $\to_P$ and $\to_D$ stand for convergence in probability and convergence in distribution, respectively; $\|\cdot\|$ denotes the Euclidean norm; $\lfloor a\rfloor$ means the largest integer not exceeding a.

2. Semiparametric single-index panel data models

A semiparametric single-index panel data model is specified as follows:

$$y_{it} = g\left(x_{it}'\theta_0\right) + \gamma_i + e_{it}, \quad i = 1,\ldots,N,\ t = 1,\ldots,T, \tag{2.1}$$

where y_{it} is a scalar process, x_{it} is a $(d\times 1)$ explanatory variable, e_{it} is an error process and the link function $g(w) \in L^2(\mathbb{R}, \exp(-w^2/2))$ is unknown. Here, $L^2(\mathbb{R}, \exp(-w^2/2)) = \{g(w) : \int_{\mathbb{R}} g^2(w)\exp(-w^2/2)dw < \infty\}$ is a Hilbert space. We use $\{\gamma_i\}$ to capture fixed effects in this model, which is allowed to be correlated with the regressors. Under the current set-up, our main interest is to consistently estimate the index vector $\theta_0 = (\theta_{01},\ldots,\theta_{0d})'$ and the link function $g(\cdot)$ for the case where both N and T go to infinity.

To meet the identification requirements (c.f. Ichimura (1993) and Horowitz (2009)), we assume that $\|\theta_0\| = 1$ and $\theta_{01} > 0$. For the link function $g(\cdot)$, we expand it by the Hermite polynomials into an orthogonal series and approximate it by a truncated series.

The Hermite polynomial system $\{H_m(w),\ m = 0, 1, 2, \ldots\}$ is a complete orthogonal system in the Hilbert space $L^2\left(\mathbb{R}, \exp(-w^2/2)\right)$ and each element is defined by

$$H_m(w) = (-1)^m \cdot \exp\left(w^2/2\right) \cdot \frac{d^m}{dw^m}\exp\left(-w^2/2\right). \tag{2.2}$$

The orthogonality of the system reads $\int_{\mathbb{R}} H_m(w) H_n(w) \exp\left(-w^2/2\right) dw = m!\sqrt{2\pi}\delta_{mn}$, where δ_{mn} is the Kronecker delta. Define further that $h_m(w) = \frac{1}{\sqrt{m!}}H_m(w)$, so that for any $g(w) \in L^2\left(\mathbb{R}, \exp\left(-w^2/2\right)\right)$ we can expand it in terms of $h_m(w)$ as follows:

$$g(w) = \sum_{m=0}^{\infty} c_m h_m(w),$$
$$c_m = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} g(w)h_m(w)\exp\left(-w^2/2\right)dw. \tag{2.3}$$

Based on the above expansion, one is already able to use a profile method or an iterative estimation method to estimate θ_0 and the link function (e.g. Dong et al. (2014)). Since neither of these two methods results in a closed-form estimator, numerical estimates are often sensitive to the initial values in practice due to multi-modality or flatness of a curve. Instead, we further expand $h_m(x_{it}'\theta_0)$ by Lemma B.1 of the Appendix as follows.

$$g\left(x_{it}'\theta_0\right) = \sum_{m=0}^{k-1} c_m h_m\left(x_{it}'\theta_0\right) + \sum_{m=k}^{\infty} c_m h_m\left(x_{it}'\theta_0\right) \tag{2.4}$$
$$= \sum_{m=0}^{k-1}\sum_{|p|=m} a_{mp}(\theta_0)\,\mathscr{H}_p(x_{it}) + \delta_k\left(x_{it}'\theta_0\right), \tag{2.5}$$

where positive integer k is truncation parameter and

$$\delta_k\left(x_{it}'\theta_0\right) = \sum_{m=k}^{\infty} c_m h_m(x_{it}'\theta_0), \qquad a_{mp}(\theta_0) = \sqrt{\binom{m}{p}}\, c_m\theta_0^p,$$
$$\binom{m}{p} = \frac{m!}{\prod_{j=1}^{d} p_j!},$$
$$\theta_0^p = \prod_{j=1}^{d}\theta_{0j}^{p_j}, \qquad \mathscr{H}_p(x_{it}) = \prod_{j=1}^{d} h_{p_j}\left(x_{it,j}\right),$$
$$x_{it} = \left(x_{it,1},\ldots,x_{it,d}\right)',\ p = (p_1,\ldots,p_d)',$$
$$|p| = p_1 + \cdots + p_d$$

and p_j's for $j = 1,\ldots,d$ are non-negative integers.

The expansion (2.5) allows us to separate the covariate x_{it} and the coefficient θ_0, so the closed-form estimator can be established