ARTICLE IN PRESS

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

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

Nonparametric fixed effects model for panel data with locally stationary regressors

Youquan Pei, Tao Huang *, Jinhong You *

Department of Statistics and Management, Shanghai University of Finance of Economics, 777 Guoding Road, Shanghai, 200433, China

ARTICLE INFO

Article history:
Received 9 July 2016
Received in revised form 16 March 2017
Accepted 30 June 2017
Available online xxxx

JEL classification:
C12
C13
C14
C23

Keywords:
Panel data models
Fixed effect
Locally stationary
Local linear estimation
Hypothesis testing

ABSTRACT

We develop methods for inference in nonparametric time-varying fixed effects panel data models that allow for locally stationary regressors and for the time series length T and cross-section size N both being large. We first develop a pooled nonparametric profile least squares dummy variable approach to estimate the nonparametric function, and establish the optimal convergence rate and asymptotic normality of the resultant estimator. We then propose a test statistic to check whether the bivariate nonparametric function is time-varying or the time effect is separable, and derive the asymptotic distribution of the proposed test statistic. We present several simulated examples and two real data analyses to illustrate the finite sample performance of the proposed methods.

1. Introduction

Panel data arise frequently in finance and various areas of economics such as industrial organization and labor economics. Based on whether or not the individual effects are correlated with the explanatory variables, panel data models can be divided into two classes. One is the random effects panel model, in which the individual effects are random and uncorrelated with the explanatory variables, and the other is the fixed effects panel model, in which the individual effects either are not random or are random but correlated with the explanatory variables. In general, fixed effects panel data modeling is more robust than random effects panel data modeling; see Baltagi (2008) for more details.

Various parametric models and corresponding inference methods have been developed for panel data. See, for example, Hsiao (2014) and Baltagi (2008), and references therein. While parametric models are useful for analyzing panel data and providing parsimonious explanations of the relationships between the response variable and the explanatory variables, they often face the risk of modeling bias. To relax the assumptions of parametric forms, various nonparametric and semiparametric fixed effects panel data models have been developed to reduce modeling bias, including those of Baltagi et al. (2002), Fan et al. (2005), Su and Ullah (2006), Cai and Li (2008), Henderson et al. (2008), Sun et al. (2009), Li et al. (2011), Chen et al. (2012) and Dong et al. (2015), among others.

There is a rich literature on parametric and nonparametric/semiparametric fixed effects panel data models; however, most of the studies reported there have focused on modeling the "large-N, small-T" case (where T is the time series length and N is the cross section size), thus assuming that it is reasonable to take the explanatory variables to be stationary. Over the last few years, it has become increasingly common to assume that both T and N can be large (Chen et al., 2012; Fan et al., 2016, 2015; Ke et al., 2015). Unfortunately, for large T, (global) stationarity is a restrictive assumption that is sometimes hard to justify. When the observation time T tends to infinity, nonstationary behavior plays an important role in diverse fields. For example, we re-explore the well-known Cigar data set (Bada and Liebl, 2012), which contains the cigarette consumption per capita of $N = 46$ American states from 1963 to 1992 ($T = 30$) as well as the income per capita and cigarette prices. We plot the time series of logarithms of Consumption_{it} and $\text{Price}_{it}/\text{cpi}_t$ for each state in Fig. 1, where cpi_t is the consumer price index.

Fig. 1 shows that neither the price index nor the consumption index constitutes a stationary process. We further use the KPSS method (Kwiatkowski et al., 1992) to test whether the time series

* Corresponding authors.
E-mail addresses: huang.tao@mail.shufe.edu.cn (T. Huang), you.jinhong@mail.shufe.edu.cn (J. You).

https://doi.org/10.1016/j.jeconom.2017.06.023
0304-4076/

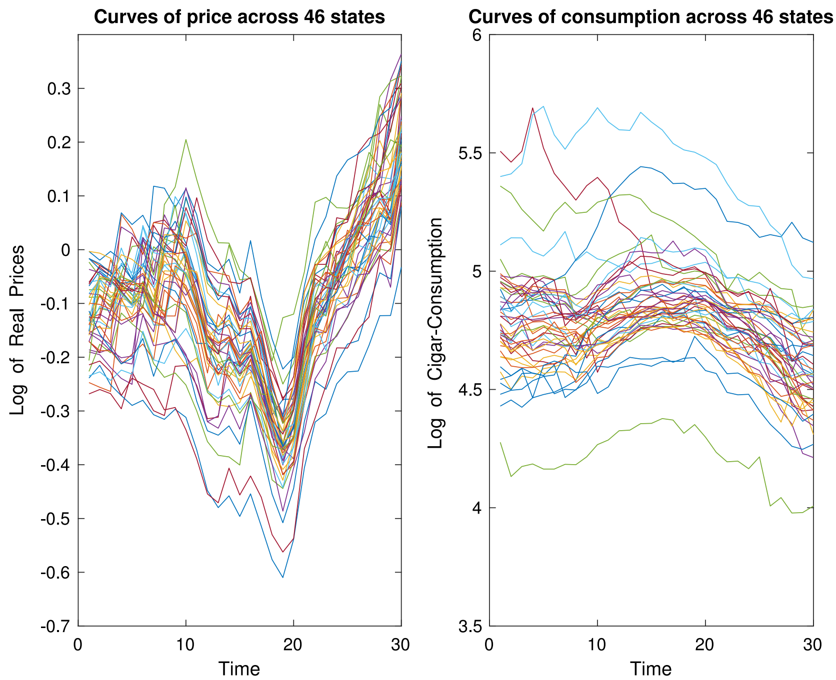

Fig. 1. Time series of $\log(\text{Price}_{it}/\text{cpi}_t)$ and $\log(\text{Consumption}_{it})$, $i = 1, \ldots, 46$.

are stationary, and the resulting p values are 0.011 and 0.010 for $\log(\text{Price}_{it}/\text{cpi}_t)$ and $\log(\text{Consumption}_{it})$, respectively, which confirms that the two time series are indeed nonstationary.

In practice, in order to model nonstationary time series, the real data at hand are often preprocessed to eliminate trends by differencing or detrending. As a result, the models used reveal only the underlying dynamics of the detrended time series, and it is difficult to interpret the evolutionary nature of the original data. It seems more attractive to model the original time series directly. One way to model nonstationary behavior is provided by the theory of locally stationary processes introduced by Dahlhaus (1997). Locally stationary time series models have attracted considerable interest in the recent literature, due to the fact that the dependence characteristics of time series can change as time evolves. As Dette et al. (2011) noted, several models have been proposed for locally stationary processes, including time-varying AR(p) models and time-varying ARMA(p, q) models. Vogt (2012) have studied nonparametric time series models with a time-varying regression function and locally stationary covariates. In the present paper, we shall focus on modeling "large-N, large-T" panel data by assuming the explanatory variables to be locally stationary. Hereinafter, we adopt the definition of locally stationarity proposed by Vogt (2012), which has an explicit explanation and an asymptotic property that is easy to derive.

We introduce a nonparametric fixed effects panel data model that can be regarded as a natural extension of trending panel data models with time-varying coefficients. This model is given by

$$Y_{it} = m\left(X_{it}, \frac{t}{T}\right) + \alpha_i + e_{it}, \quad i = 1, \ldots, N,\ t = 1, \ldots, T, \qquad (1.1)$$

where the covariates $X_{it} = (X_{it,1}, \ldots, X_{it,p})^\tau$ have dimension p, the α_i are the unobserved individual fixed effects satisfying $\sum_{i=1}^N \alpha_i = 0$ and may be correlated with X_{it} via some unknown structure, and the e_{it} are independent and identically distributed (i.i.d.) random errors satisfying $\mathrm{E}(e_{it}|X_{it}) = 0$ and $\mathrm{E}(e_{it}^2|X_{it}) = \sigma^2$. The regressors $\{X_{it}\}$ are assumed to be locally stationary, and the nonparametric regression function $m(\cdot, \cdot)$ is allowed to change smoothly over time. As discussed by Vogt (2012), the function $m(\cdot, \cdot)$ does not depend on the real time t but rather on the rescaled time t/T, which allows additional information to be obtained on the shape of m locally around a fixed time point t as the sample size increases.

The model (1.1) is quite general, and it can characterize the dynamic influence of explanatory variables on the responses. In particular, it contains the following nonparametric trending panel data model:

$$Y_{it} = m_1(X_{it}) + f\left(\frac{t}{T}\right) + \alpha_i + e_{it},$$
$$i = 1, \ldots, N, \quad t = 1, \ldots, T, \qquad (1.2)$$

where $m_1(\cdot)$ is an unknown function and $f(\cdot)$ is an unknown time trend function or a time-specific effect such as the calendar effect that is common across individuals. Obviously, the model investigated in Chen et al. (2012) is a special case of (1.2). If the regression function $m(\cdot, \cdot)$ in (1.1) is time-invariant, we then have

$$Y_{it} = m_1(X_{it}) + \alpha_i + e_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T. \qquad (1.3)$$

The model (1.3) has been extensively studied in the literature; see Henderson et al. (2008), among others.

In this paper, we develop a pooled nonparametric profile least squares dummy variable approach to estimate the unknown function $m(\cdot, \cdot)$ in (1.1), and establish the optimal convergence rate and asymptotic normality of the resultant estimator. In addition, an important issue is how the true model is to be correctly identified. Zhang and Wu (2011) considered testing parametric assumptions of trends in time series models. Vogt (2015) proposed a kernel-based test statistic to test for structural change in time-varying nonparametric regression time series models. Zhang and Wu (2015) studied the estimation and model selection problem for the time-varying nonparametric time series model. To the best of our knowledge, almost all the analogous testing problems are focused on time series models, while in this paper, we shall develop a test statistic to check whether the bivariate nonparametric function is time-varying or the time effect is separable, and we shall derive the asymptotic distribution of the proposed test statistic in the general framework (1.1) for panel data models with fixed effects.

Please cite this article in press as: Pei Y., et al., Nonparametric fixed effects model for panel data with locally stationary regressors. Journal of Econometrics (2017), https://doi.org/10.1016/j.jeconom.2017.06.023.