Journal of Econometrics 191 (2016) 176–195

Contents lists available at ScienceDirect

Journal of Econometrics

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

Estimation of heterogeneous panels with structural breaks[☆]

Badi H. Baltagi[a,b,*], Qu Feng[c], Chihwa Kao[a]

[a] Department of Economics and Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, NY 13244-1020, USA

[b] Department of Economics, University of Leicester, University Road, Leicester LE17 6EE, UK

[c] Division of Economics, School of Humanities and Social Sciences, Nanyang Technological University, HSS-04-48, 14 Nanyang Drive, Singapore 637332, Singapore

ARTICLE INFO

Article history:
Received 9 February 2014
Received in revised form
5 March 2015
Accepted 25 March 2015
Available online 31 October 2015

JEL classification:
C23
C33

Keywords:
Heterogeneous panels
Cross-sectional dependence
Structural breaks
Common correlated effects

ABSTRACT

This paper extends Pesaran's (2006) work on common correlated effects (CCE) estimators for large heterogeneous panels with a general multifactor error structure by allowing for unknown common structural breaks. Structural breaks due to new policy implementation or major technological shocks, are more likely to occur over a longer time span. Consequently, ignoring structural breaks may lead to inconsistent estimation and invalid inference. We propose a general framework that includes heterogeneous panel data models and structural break models as special cases. The least squares method proposed by Bai (1997a, 2010) is applied to estimate the common change points, and the consistency of the estimated change points is established. We find that the CCE estimator have the same asymptotic distribution as if the true change points were known. Additionally, Monte Carlo simulations are used to verify the main results of this paper.

1. Introduction

For panel data models, the presence of cross-sectional dependence due to unobservable common factors or spatial spillover effect is a major concern in estimation and inference. It could lead to invalid inference and inconsistent estimators, see Lee (2002) and Andrews (2005).[1] Several tests for cross-sectional dependence in panel data models have been proposed in the literature. These include Pesaran (2004, 2012), Ng (2006), Pesaran et al. (2008), Sarafidis et al. (2009), Chen et al. (2012), Hsiao et al. (2012), Baltagi et al. (2011, 2012), Halunga et al. (2011), Juhl (2011), and Su and Zhang (2011), to mention a few. To deal with cross-sectional dependence in panels, two general estimation methods have been proposed including spatial estimation methods (Anselin, 1988; Kelejian and Prucha, 1999; Kapoor et al., 2007; Lee, 2007; Lee and Yu, 2010, to name a few), and factor models (see Pesaran, 2006; Bai, 2009, to name a few).

In particular, Pesaran (2006) develops common correlated effects (CCE) estimators for large heterogeneous panels with a general multifactor error structure by least squares using augmented data. The common correlated effects (factors) can be asymptotically partialled out by means of the cross-sectional average of the dependent variable and the individual-specific regressors when the cross-section dimension is large. Kapetanios et al. (2011) show that the CCE estimator can be extended to the case of nonstationary unobserved common factors. Additionally, the CCE approach is also shown to be applicable to situations of spatial and other forms of weak cross-sectional dependent errors (Pesaran and Tosetti, 2011; Chudik et al., 2011), and heterogeneous dynamic panel data models with weakly exogenous regressors (Chudik and Pesaran, 2013).[2]

☆ The authors would like to thank the editor Han Hong, the associate editor and three anonymous referees for their constructive comments and suggestions. We are also grateful to helpful comments from Jushan Bai, Jiti Gao, Kazuhiko Hayakawa, Dukpa Kim, Tatsushi Oka, Richard Smith, Liangjun Su, and participants of a seminar at National University of Singapore and from 2013 Asian Meeting of the Econometric Society, 2014 Tsinghua Econometrics Conference, 2014 China Meeting of the Econometric Society, 2014 International Panel Data Conference. Financial support from the MOE AcRF Tier 1 Grant M4011314 at Nanyang Technological University is gratefully acknowledged.

* Corresponding author at: Department of Economics and Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, NY 13244-1020, USA.

E-mail addresses: bbaltagi@maxwell.syr.edu (B.H. Baltagi), qfeng@ntu.edu.sg (Q. Feng), dkao@maxwell.syr.edu (C. Kao).

1 Baltagi and Pirotte (2010) show that tests of hypotheses based on standard panel data estimators that ignore spatial dependence lead to misleading inference.

2 In a panel with unobserved common factors, Banerjee and Carrion-i-Silvestre (2011) suggest a test for panel cointegration based on a pooled CCE estimator of the coefficients.

http://dx.doi.org/10.1016/j.jeconom.2015.03.048

However, this literature assumes that the slope coefficients are constant over time. This implicit assumption is common in the literature on panel data models with large time dimension, see for example, Kao (1999), Phillips and Moon (1999), Hahn and Kuersteiner (2002), Alvarez and Arellano (2003), Phillips and Sul (2007), Pesaran and Yamagata (2008), Hayakawa (2009), to name a few. Due to policy implementation or technological shocks, structural breaks are possible especially for panels with a long time span. Consequently, ignoring structural breaks may lead to inconsistent estimation and invalid inference.

This paper extends Pesaran's (2006) heterogeneous panels by allowing for unknown common structural breaks in the slopes. This is useful for example when global technological or financial shocks affect all markets or firms at the *same* time. Since the framework of heterogeneous panels is fairly general and includes popular panel data models as special cases, it allows us to examine the issue of structural breaks in panel data models in a less restrictive way.

By considering both cross-sectional dependence and structural breaks in a general panel data model, this paper also contributes to the change point literature in several ways. First, it extends Bai's (1997a) time series regression model to heterogeneous panels, showing that the consistency of estimated change points can be achieved with the information along the cross-sectional dimension. This result confirms the findings of Bai (2010) and Kim (2011). Second, it also enriches the analysis of common breaks of Bai (2010) and Kim (2011) in a panel mean-shift model and a panel deterministic time trend model by extending them to a regression model using panel data. This makes it possible to allow for structural breaks and cross-sectional dependence in empirical work using panel regressions. In particular, our methods can be applied to regression models using large stationary panel data, such as country-level panels and state/provincial-level panels.[3]

Regarding estimating common breaks in panels, Feng et al. (2009) and Baltagi et al. (forthcoming) also show the consistency of the estimated change point in a simple panel regression model. Hsu and Lin (2012) examine the consistency properties of the change point estimators for nonstationary panels. More recently, Qian and Su (2014) and Li et al. (2014) study the estimation and inference of common breaks in panel data models with and without interactive fixed effects using Lasso-type methods. In terms of detecting structural breaks in panels, some recent literature includes Horváth and Hušková (2012) in a panel mean shift model with and without cross-sectional dependence, De Wachter and Tzavalis (2012) in dynamic panels, and Pauwels et al. (2012) in heterogeneous panels, to name a few.

The paper is organized as follows. Section 2 introduces heterogeneous panels with a common structural break. Section 3 starts with a simple heterogeneous panel model that ignores the unobservable common correlated effects. This model can be regarded as a direct extension of Bai's (1997a) time series regression model to a panel setup. The least squares estimation proposed by Bai (1997a) is applied. With the main results established in Section 3, the discussion of the general model with common correlated effects is presented in Section 4. Section 5 briefly discusses the case of multiple common breaks. In Section 6, Monte Carlo simulations are used to verify the consistency of the estimated common change point for both models considered. Section 7 provides concluding remarks. The mathematical proofs are relegated to the Appendix.

[3] Some empirical examples include Fleisher et al. (2010) using Chinese provincial-level panel data, and Huang (2009) using country-level panels, to mention a few.

2. Heterogeneous panels with a common structural break

In a heterogeneous panel data model:

$$y_{it} = x_{it}'\beta_i + e_{it}, \quad i = 1, \ldots, N;\ t = 1, \ldots, T, \tag{1}$$

x_{it} is a $p \times 1$ vector of explanatory variables, and the errors are cross-sectionally correlated, modelled by a multifactor structure

$$e_{it} = \gamma_i' f_t + \varepsilon_{it}, \tag{2}$$

where f_t is an $m \times 1$ vector of unobserved factors and γ_i is the corresponding loading vector. ε_{it} is the idiosyncratic error independent of x_{it}. However, x_{it} could be affected by the unobservable common effects f_t. Projecting x_{it} on f_t, we obtain

$$x_{it} = \Gamma_i' f_t + v_{it}, \quad i = 1, \ldots, N;\ t = 1, \ldots, T, \tag{3}$$

where Γ_i is an $m \times p$ factor loading matrix. v_{it} is a $p \times 1$ vector of disturbances. Due to the correlation between x_{it} and e_{it}, OLS for each individual regression could be inconsistent. Thus, Pesaran (2006) develops the CCE estimator of β_i by least squares using augmented data.[4]

In this paper, we allow for structural breaks to occur in some or all components of the slopes β_i.[5] Following Bai (2010) and Kim (2011), a structural break at a common unknown date k_0 is assumed. This could be due to a macro policy implementation or a technological shock that affects all markets or firms at the same time. More formally,

$$y_{it} = x_{it}'\beta_i(k_0) + e_{it}, \quad i = 1, \ldots, N;\ t = 1, \ldots, T, \tag{4}$$

where some or all components of $\beta_i(k_0)$ are different before and after the date k_0.[6] Following Bai (1997a), this structural break model can be written as

$$y_{it} = \begin{cases} x_{it}'\beta_i + e_{it}, & t = 1, \ldots, k_0, \\ x_{it}'\beta_i + z_{it}'\delta_i + e_{it}, & t = k_0 + 1, \ldots, T, \end{cases} \tag{5}$$

$i = 1, \ldots, N$, where $z_{it} = R'x_{it}$ denotes a $q \times 1$ subvector of x_{it} with $R' = (0_{q\times(p-q)}, I_q)$. I_q is the $q \times q$ identity matrix with $q \le p$. The case where $q < p$ denotes a partial change model, while the case where $q = p$ is for a pure change model. Pauwels et al. (2012) propose a testing procedure for k_0 in this setting.

Substituting $z_{it} = R'x_{it}$ in (5), we obtain

$$\beta_i(k_0) = \beta_i + R\delta_i \cdot 1\{t > k_0\} = \begin{cases} \beta_{1i} = \beta_i, & t = 1, \ldots, k_0, \\ \beta_{2i} = \beta_i + R\delta_i, & t = k_0 + 1, \ldots, T, \end{cases}$$

so that $\beta_{2i} - \beta_{1i} = R\delta_i$, and δ_i denotes the slope jump for i. When $\delta_i = 0$ there is no structural break in the slope.

The case of multiple break points will be discussed in Section 5. In the next two sections, we consider the simple case of one common break as in model (5). Compared with the heterogeneous

[4] For simplicity, observed common effects like seasonal dummies are not included in (1), but they can be easily handled as in Pesaran (2006).

[5] Pesaran (2004) discusses testing for cross-sectional dependence in a heterogeneous panel model with structural breaks. Kapetanios et al. (2011) examine the performance of the CCE estimator in case of a structural break in the mean of the unobserved factors using Monte Carlo experiments.

[6] As shown in Section 4, to apply the CCE approach, the common break assumption is required. In Kim (2014), the common break assumption is generalized to handle heterogeneous responses to a common shock that follow a common distribution. Liao and Wang (2012) also assume a common distribution, instead of a common break date, and estimate individual-specific structural breaks and their cross-sectional distribution using Bayesian methods. In addition, Li et al. (2011) consider a time-varying coefficient panel data model where the slope coefficient is allowed to be different for each time period, e.g., $\beta(t)$, but common for all individuals.