Journal of Econometrics 204 (2018) 66–85

Contents lists available at ScienceDirect

Journal of Econometrics

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

Testing for common breaks in a multiple equations system

Tatsushi Oka [a,*], Pierre Perron [b]

[a] Department of Econometrics and Business Statistics, Monash University, 900 Dandenong Road, Caulfield East, Victoria 3145, Australia
[b] Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, USA

ARTICLE INFO

Article history:
Received 31 May 2016
Received in revised form 31 May 2017
Accepted 10 January 2018
Available online 16 February 2018

JEL classification:
C32

Keywords:
Change-point
Segmented regressions
Break dates
Hypothesis testing
Multiple equations systems

ABSTRACT

The issue addressed in this paper is that of testing for common breaks across or within equations of a multivariate system. Our framework is very general and allows integrated regressors and trends as well as stationary regressors. The null hypothesis is that breaks in different parameters occur at common locations and are separated by some positive fraction of the sample size unless they occur across different equations. Under the alternative hypothesis, the break dates across parameters are not the same and also need not be separated by a positive fraction of the sample size whether within or across equations. The test considered is the quasi-likelihood ratio test assuming normal errors, though as usual the limit distribution of the test remains valid with non-normal errors. Of independent interest, we provide results about the rate of convergence of the estimates when searching over all possible partitions subject only to the requirement that each regime contains at least as many observations as some positive fraction of the sample size, allowing break dates not separated by a positive fraction of the sample size across equations. Simulations show that the test has good finite sample properties. We also provide an application to issues related to level shifts and persistence for various measures of inflation to illustrate its usefulness.

1. Introduction

Issues related to structural change have been extensively studied in the statistics and econometrics literature (see Csörgö and Horváth, 1997; Perron, 2006 for comprehensive reviews). In the last twenty years or so, substantial advances have been made in the econometrics literature to cover models at a level of generality that makes them relevant across time-series applications in the context of unknown change points. For example, Bai (1994, 1997) studies the least squares estimation of a single change point in regressions involving stationary and/or trending regressors. Bai and Perron (1998, 2003) extend the testing and estimation analysis to the case of multiple structural changes and present an efficient algorithm. Hansen (1992) and Kejriwal and Perron (2008) consider regressions with integrated variables. Andrews (1993) and Hall and Sen (1999) consider nonlinear models estimated by generalized method of moments. Bai (1995, 1998) studies structural changes in least absolute deviation regressions, while Qu (2008), Su and Xiao (2008) and Oka and Qu (2011) analyze structural changes in regression quantiles. Hall et al. (2012) and Perron and Yamamoto (2014, 2015) consider structural changes in linear models with endogenous regressors. Studies about structural changes in panel data models include Bai (2010), Kim (2011), Baltagi et al. (2016) and Qian and Su (2016) for linear panel data models and Breitung and Eickmeier (2011), Cheng et al. (2016), Corradi and Swanson (2014), Han and Inoue (2015) and Yamamoto and Tanaka (2015) for factor models.

The literature on structural breaks in a multiple equations system includes Bai et al. (1998), Bai (2000) and Qu and Perron (2007), among others. Their analysis relies on a common breaks assumption, under which breaks in different basic parameters (regression coefficients and elements of the covariance matrix of the errors) occur at a common location or are separated by some positive fraction of the sample size (i.e., asymptotically distinct).[1] Bai et al. (1998) assume a single common break across equations for a multivariate system with stationary regressors and trends as well as for cointegrated systems. For the case of multiple common breaks, Bai (2000) analyzes vector autoregressive models for stationary variables and Qu and Perron (2007) cover multiple system equations, allowing for more general stationary regressors and arbitrary restrictions across parameters. Under the framework of Qu and Perron (2007), Kurozumi and Tuvaandorj (2011) propose model selection procedures for a system of equations with multiple common breaks and Eo and Morley (2015) consider a confidence set for the common break date based on inverting the

* Corresponding author.
E-mail addresses: tatsushi.oka@monash.edu (T. Oka), perron@bu.edu (P. Perron).

[1] The concept of common breaks here is quite distinct from the notion of co-breaking or co-trending (e.g., Hatanaka and Yamada, 2003; Hendry and Mizon, 1998). In this literature, the focus is on whether some linear combination of series with breaks do not have a break, a concept akin to that of cointegration.

https://doi.org/10.1016/j.jeconom.2018.01.003
0304-4076/

likelihood ratio test. In this literature, it has been documented that common breaks allow more precise estimates of the break dates in multivariate systems. Given unknown break dates, however, an issue of interest for most applications concerns the validity of the assumption of common breaks.[2] To our knowledge, no test has been proposed to address this issue.

Our paper addresses three outstanding issues about testing for common breaks. First, we propose a quasi-likelihood ratio test under a very general framework.[3] We consider a multiple equations system under a likelihood framework with normal errors, though the limit distribution of the proposed test remains valid with non-normal, serially dependent and heteroskedastic errors. Our framework allows integrated regressors and trends as well as stationary regressors as in Bai et al. (1998) and also accommodates multiple breaks and arbitrary restrictions across parameters as in Qu and Perron (2007). Thus, our results apply for general systems of multiple equations considered in existing studies. A case not covered in our framework is when the regressors depend on the break date. This occurs when considering joint segmented trends and this issue was analyzed in Kim et al. (2017).

Second, we propose a test for common breaks not only across equations within a multivariate system, but also within an equation. As in Bai et al. (1998), the issue of common breaks is often associated with breaks occurring across equations, whereas one may want to test for common breaks in the parameters within a regression equation, whether a single equation or a system of multiple equations are considered. More precisely, the null hypothesis of interest is that some subsets of the basic parameters share one or more common break dates, so that each regime is separated by some positive fraction of the sample size. Under the alternative hypothesis, the break dates are not the same and also need not be separated by a positive fraction of the sample size, or be asymptotically distinct.

Third, we derive the asymptotic properties of the quasi-likelihood and the parameter estimates, allowing for the possibility that the break dates associated with different basic parameters may not be asymptotically distinct. This poses an additional layer of difficulty, since existing studies establish the consistency and rate of convergence of estimators only when the break dates are assumed to either have a common location or be asymptotically distinct, at least under the level of generality adopted here. Moreover, we establish the results in the presence of integrated regressors and trends as well as stationary regressors. This is by itself a noteworthy contribution. These asymptotic results will allow us to derive the limit distribution of our test statistic under the null hypothesis and also facilitate asymptotic power analyses under fixed and local alternatives. We can show that our test is consistent under fixed alternatives and also has non-trivial local power.

There is one additional layer of difficulty compared to Bai and Perron (1998) or Qu and Perron (2007). In their analysis, it is possible to transform the limit distribution so that it can be evaluated using a closed form solution and thus critical values can be tabulated. Here, no such solution is available and we need to obtain critical values for each case through simulations. This involves simulating the Wiener processes with consistent parameter estimates and evaluating each realization of the limit distribution with and without the restriction of common breaks. While it is conceptually straightforward and quick enough to be feasible for common applications, the procedure needs to be repeated many times to obtain the relevant quantities and can be quite computationally intensive. This is because we need to search over many possible combinations of all the permutations of the break locations for each replication of the simulations. To reduce the computational burden, we propose an alternative procedure based on the particle swarm optimization method developed by Eberhart and Kennedy (1995) with the Karhunen–Loève representation of stochastic processes. Our simulation results suggest that the test proposed has reasonably good size and power performance even in small samples under both computation procedures. Also, we apply our test to inflation series, following the work of Clark (2006) to illustrate its usefulness.

The remainder of the paper is as follows. Section 2 introduces the models with and without the common breaks assumption and describes the estimation methods under the quasi-likelihood framework. Section 3 presents the assumptions and asymptotic results including the asymptotic null distribution and asymptotic power analyses. Section 4 examines the finite sample properties of our procedure via Monte Carlo simulations. Section 5 presents an empirical application and Section 6 concludes. Appendix A contains all the proofs.

2. Models and quasi-likelihood method

In this section, we first introduce models for a multiple equations system with and without common breaks. Subsequently, we describe the quasi-likelihood estimation method assuming normal errors and then propose the quasi-likelihood ratio test for common breaks. For illustration purpose, we also discuss some examples.

As a matter of notation, "$\xrightarrow{p}$" denotes convergence in probability, "$\xrightarrow{d}$" convergence in distribution and "$\Rightarrow$" weak convergence in the space $D[0,\infty)$ under the Skorohod topology. We use $\mathbb{R}$, $\mathbb{Z}$ and $\mathbb{N}$ to denote the set of all real numbers, all integers and all positive integers, respectively. For a vector x, we use $\|\cdot\|$ to denote the Euclidean norm (i.e., $\|x\| = \sqrt{x'x}$), while for a matrix A, we use the vector-induced norm (i.e., $\|A\| = \sup_{x\neq 0}\|Ax\|/\|x\|$). Define the L_r-norm of a random matrix X as $\|X\|_r = (\sum_i\sum_j E|X_{ij}|^r)^{1/r}$ for $r \geq 1$. Also, $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$ for any $a, b \in \mathbb{R}$. Let $\circ$ denote the Hadamard product (entry-wise product) and let $\otimes$ denote the Kronecker product. Define $\mathbb{1}_{\{\cdot\}}$ as the indicator function taking value one when its argument is true, and zero otherwise and e_i as a unit vector having 1 at the ith entry and 0 for the others. We use the operator $\mathrm{vec}(\cdot)$ to convert a matrix into a column vector by stacking the columns of the matrix and the operator $\mathrm{tr}(\cdot)$ to denote the trace of a matrix. The largest integer not greater than $a \in \mathbb{R}$ is denoted by $[a]$ and the sign function is defined as $\mathrm{sgn}(a) = -1, 0, 1$ if $a > 0$, $a = 0$ or $a < 0$, respectively.

2.1. The models with and without common breaks

Let the data consist of observations $\{(y_t, x_{tT})\}_{t=1}^T$, where y_t is an $n \times 1$ vector of dependent variables and x_{tT} is a $q \times 1$ vector of explanatory variables for $n, q \in \mathbb{N}$ with a subscript t indexing a temporal observation and T denoting the sample size. We allow the regressors x_{tT} to include stationary variables, time trends and integrated processes, while scaling by the sample size T so that the order of all components is the same. In what follows, we consider

$$x_{tT} = \left(z_t', \varphi(t/T)', T^{-1/2}w_t'\right)'.$$

2 The common breaks assumption is also used in the literature on panel data (e.g. Bai, 2010; Kim, 2011; Baltagi et al., 2016). In this paper, we consider a multiple equations system in which the number of equations are relatively small, and thus panel data models are outside our scope. However, testing for common breaks in a system with a large number of equations is an interesting avenue for future research.

3 One may also consider other type of tests, such as LM-type tests. The literature on structural breaks, however, documents that even though LM-type tests have simple asymptotic representations, they tend to exhibit poor finite sample properties with respect to power. Thus, this paper focuses on the LR test (see Deng and Perron, 2008; Kim and Perron, 2009; Perron and Yamamoto, 2016, for instance).