Journal of Econometrics 188 (2015) 166–181

Contents lists available at ScienceDirect

Journal of Econometrics

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

Structural-break models under mis-specification: Implications for forecasting

Bonsoo Koo[a,*], Myung Hwan Seo[b]

[a] Department of Econometrics and Business Statistics, Faculty of Business and Economics, Monash University, PO Box 11E, Clayton Campus, VIC 3800, Australia

[b] Department of Economics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 151-742, Republic of Korea

ARTICLE INFO

Article history:
Received 19 April 2013
Received in revised form
8 December 2014
Accepted 6 March 2015
Available online 30 April 2015

JEL classification:
C13
C22
C53

Keywords:
Structural break
Forecasting
Mis-specification
Cube-root asymptotics

ABSTRACT

This paper revisits the least squares estimator of the linear regression with a structural break. We view the model as an approximation to the true data generating process whose exact nature is unknown but perhaps changing over time either continuously or with some jumps. This view is widely held in the forecasting literature and under this view, the time series dependence property of all the observed variables is unstable as well. We establish that the rate of convergence of the estimator to a properly defined limit is at most the cube root of T, where T is the sample size, which is much slower than the standard super consistent rate. We also provide an asymptotic distribution of the estimator and that of the Gaussian quasi likelihood ratio statistic for a certain class of true data generating processes. We relate our finding to current forecast combination methods and propose a new averaging scheme. Our method compares favourably with various contemporary forecasting methods in forecasting a number of macroeconomic series.

1. Introduction

Structural breaks have been observed in many economic time series and economic models (Stock and Watson, 1996). Documented examples include interest rates (Garcia and Perron, 1996), GDP (Ben-David and Papell, 1998; McConnell and Perez-Quiros, 2000) and labour productivity (Hansen, 2001). Consequently, various aspects of econometric analyses of structural break models have been investigated throughout the literature. For relevant surveys, see Bhattacharya (1994), Stock (1994), van Dijk et al. (2002) and Perron (2006).

A break, also known as a change-point, is often associated with a change in parameter values of the underlying regression model along an observable variable and the change involves a jump or discontinuity in the regression function. As observed by Bai and Perron (1998), Hansen (2000), and Perron and Qu (2006), among many others, this generates a convenient oracle property: an estimator for the location of the break converges to a true break point faster than estimators for other parameters converge, and they are asymptotically independent of each other. This means that distribution theory for structural break models can be established as if break dates were known *a priori*, once break dates are consistently estimated. This property extends to nonparametric regression models (Delgado and Hidalgo, 2000). Consequently, standard estimation and inference procedures involve estimation of break dates, followed by estimation of, and inference for, other model parameters conditional on these estimated change-points.

In practice, however, we are not certain whether structural break models correctly specify the true underlying DGP or not. Therefore, economic and statistical models for structural breaks are subject to mis-specification. The case where an estimated model has a smaller number of breaks than the true number has been analysed in Chong (1995) and Bai and Perron (1998). Extensions to the possible mis-specification in the regression function in addition to the mis-specification in the number of breaks have been considered by Chong (2003) and Bai et al. (2008). In these works, the breaks in the true regression functions occur only finite times with jumps and as a consequence the oracle property of the break estimates is preserved despite mis-specification. Thus, we view this type of mis-specification as weak mis-specification.

This paper assumes that the true regression function may not be linear and changes constantly over time. The change is mostly

* Corresponding author. Tel.: +61 3 9905 0547; fax: +61 3 9905 5474.
E-mail addresses: bonsoo.koo@monash.edu (B. Koo), myung.h.seo@gmail.com (M.H. Seo).

http://dx.doi.org/10.1016/j.jeconom.2015.03.046
0304-4076/

continuous over time (after time rescaling) but can have a finite number of jumps as well. In this scenario, the estimated change point in the linear regression with one break does not need to converge to one of the jumps in the true regression function, if any. We call this case as strong mis-specification. For instance, the linear regression model with one break may be approximating the locally stationary model. Thus, we extend the literature to another interesting problem. We conjecture that if the estimated change point converges to one of the true unknown jump points, the asymptotics should be analogous to that of Bai et al. (2008).

We show that the asymptotic property of the least squares estimator changes dramatically from the previous literature and provide conditions for such asymptotic results. In particular, the asymptotic independence, that is, the oracle property does not hold any more and the convergence rate of the estimator is at most $T^{-1/3}$, where T is the sample size, and it can be even slower. It appears that the order p of the L^p norm in the near epoch dependence affects the convergence rate, see Section 3.1. We also derive an asymptotic distribution when the true model is the time-varying coefficient model as in Robinson (1989). The asymptotic distribution is characterized by the maximizer of a nonzero mean Gaussian process. We also show that a rescaled quasi likelihood ratio statistic for the change point is asymptotically nuisance parameter free.

Indeed, many works on forecasting with structural break models have viewed the structural instability in economic time series and models as ongoing and its exact nature as unknown. For instance, see Clements and Hendry (1998), a recent review by Rossi (2012), and other works cited in this paragraph. Under the premise of model uncertainty, it is often argued that forecasting based solely on the post-break data is not necessarily optimal and alternative methods could outperform the traditional post-break forecasting method. Furthermore, the jump point in the estimated break model does not need to be a discontinuity point of the constantly changing true data generating scheme. Then, the literature naturally considered certain forecast averaging methods, for example, as in Pesaran and Timmermann (2007). In particular, they include averaging over different estimation windows (Pesaran and Pick, 2011), the optimal and robust weighting forecasting approach (Pesaran et al., 2013), reverse ordered CUSUM weighting (Pesaran and Timmermann, 2002), and more weighting on the recent data (Tian and Anderson, 2014), and weighted averaging between the models with and without breaks based on a Mallows criterion (Hansen, 2009).

Our asymptotic result provides an asymptotically justifiable way to construct the intervals, on which averaging is conducted, under very general assumptions on the data generating process. Furthermore, it highlights the reason why reducing the variance is important in forecasting. Averaging is expected to reduce the variance in the forecast and Breiman (1996) showed that the aggregation based on the bootstrap reduces the estimation variance in the prediction meaningfully under random sampling, when the underlying model is unstable. However, our asymptotic result suggests that the bootstrap may not be optimal as it is not consistent under the cube root type asymptotic experiment, see e.g. Abrevaya and Huang (2005). This is in line with Bühlmann and Yu (2002) and we extended their analysis to time series data.[1] Furthermore, unlike Bühlmann and Yu, even the subsampling is not valid due to the nonstationarity of the data in the current work. While it is intuitively appealing to assign different weights to different break points within the estimated interval, the current work focuses on the construction of the interval and leave the weight issue as a future research topic. It is worthwhile to note that a bigger interval is not necessarily better than a shorter one as it introduces more bias when the true data generating process is constantly time varying.

Lastly, we provide an empirical study in which we compare various forecasting methods with our forecasts averaging, which is based on inverting the quasi likelihood ratio statistic. In particular, we examine out-of-sample forecasting ability using the extensive macroeconomic time series data from Stock and Watson (2009). For the comparative study, we adopt two tests for predictive ability proposed by Giacomini and White (2006). We also compare forecasting methods using other measures such as relative prediction mean squared errors and percentage improvement. We found that our forecasting scheme compares favourably with the other forecasting methods in all the scenarios.

The remainder of this paper is organized as follows. Section 2 introduces the model and the estimation procedure for the unknown parameters. Section 3 provides conditions for asymptotic theory and develops distribution theories. In Section 4, we propose our forecasting method which incorporates our newly developed distribution theory after presenting two different Monte Carlo simulation studies. Empirical application of our forecasting procedure to various time series data is provided in Section 4.2.1. Section 5 concludes. The mathematical proofs are relegated to Appendix.

2. Structural break model under mis-specification

This section revisits the classical linear regression model with a possible break. That is,

$$y_{t,T} = x_{t,T}'\beta_1 1(\tau \le \gamma) + x_{t,T}'\beta_2 1(\tau > \gamma) + e_{t,T} \tag{1}$$

where $1(\cdot)$ is an indicator function, $\tau = t/T$ and unknown parameters $\theta = (\beta', \gamma)' \in \Theta$ with $\beta = (\beta_1', \beta_2')'$. In particular, $\gamma \in \Gamma$, which is a closed interval in $(0, 1)$. The regressors, $x_{t,T} \in \mathbb{R}^p$, may contain lags of the dependent variable and lagged explanatory variables. The array notation is used to allow for general types of processes. Prime denotes transpose. Some elements of β_2 could be zero. When there are no zeros in β_2, this is a pure structural break model. Let $\delta = \beta_2 - \beta_1$. If $\delta = 0$, the parameter γ is not identified.

The standard least squares estimator $\hat{\theta}$ of θ minimizes the sum of squared residuals

$$\begin{aligned}\mathbb{S}_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T}[y_{t,T} - x_{t,T}'\beta_1 1(\tau \le \gamma) - x_{t,T}'\beta_2 1(\tau > \gamma)]^2 \\ &= \frac{1}{T}\sum_{t=1}^{\lfloor \gamma T \rfloor}(y_{t,T} - x_{t,T}'\beta_1)^2 + \frac{1}{T}\sum_{t=\lfloor \gamma T \rfloor + 1}^{T}(y_{t,T} - x_{t,T}'\beta_2)^2, \end{aligned} \tag{2}$$

where $\lfloor t \rfloor$ is the biggest integer less than or equal to t. That is,

$$\hat{\theta} = \operatorname*{argmin}_{\theta \in \Theta} \mathbb{S}_T(\theta).$$

For a given $\gamma \in \Gamma$, we can obtain the concentrated sum of squared residuals,

$$\begin{aligned}\mathbb{S}_T(\gamma) &\equiv \mathbb{S}_T(\hat{\beta}_1(\gamma), \hat{\beta}_2(\gamma), \gamma) \\ &= \frac{1}{T}\sum_{t=1}^{\lfloor \gamma T \rfloor}(y_{t,T} - x_{t,T}'\hat{\beta}_1(\gamma))^2 \\ &\quad + \frac{1}{T}\sum_{t=\lfloor \gamma T \rfloor + 1}^{T}(y_{t,T} - x_{t,T}'\hat{\beta}_2(\gamma))^2, \end{aligned} \tag{3}$$

where $\hat{\beta}_1(\gamma)$ and $\hat{\beta}_2(\gamma)$ are the OLS estimates in the two subsamples. Then,

$$\hat{\gamma} = \operatorname*{argmin}_{\gamma \in \Gamma} \mathbb{S}_T(\gamma). \tag{4}$$

[1] In a related work, Seo (2014) studies the mis-specification in the threshold regression with dependent data. However, it considers stationary data as the change occurs due to an observable covariate. In addition, subsampling can work in his case whereas it does not work in our case.