Detecting Structural Breaks in Foreign Exchange Markets by using the group LASSO technique*

Mikio Ito

Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan

February 8, 2022

Abstract: This article proposes an estimation method to detect breakpoints for linear time series models with their parameters that jump scarcely. Its basic idea owes the group LASSO (group least absolute shrinkage and selection operator). The method practically provides estimates of such time-varying parameters of the models. An example shows that our method can detect each structural breakpoint’s date and magnitude.

Keywords: Vector Error Correction Model; Time-Varying Parameters; group LASSO; Structural Break.

JEL Classification Numbers: C22; L16.

*The author reported the previous version in the 96th meeting of Western Economic Association International in 2021. The author received useful suggestions from Professor Tatsuma Wada in the meeting. Mr. Akio Sashida suggested several possible events that might trigger breaks within my sample period.
1 Introduction

Recent development in data science has enabled us to handle 'big' datasets, including high dimensional time series; such data have been increasingly common in economics and finance. At the same time, we can adopt models with more than ten thousand parameters to estimate, as this paper addresses. A sparse regression technique based on the least absolute shrinkage and selection operator (LASSO) or a similar one is typical to extract useful information from such big data. (See Tibshirani et al. (2015) for general discussion on LASSO.)

There exists recent literature about the theory and application of the LASSO technique to time-series datasets. Articles by Michailidis and d’Alché Buc (2013) and Basu et al. (2015) study vector autoregressive model using the LASSO technique. Chan et al. (2014) adopt a group LASSO model to solve the problem in a variable selection for a structural break autoregression model. Ito (2019) shows a method to estimate state parameters that jump rarely based on the group LASSO.

On the other hand, recent economists use econometric models with parameters varying over time when they are interested in structural breaks of economic or financial systems. For example, Primiceri (2005) studies a macroeconomic model with parameters varying over time by adopting a Bayesian technique to estimate a state-space model. The author and his coworkers apply a conventional least square method to estimate time-varying parameters of time series models such as a VAR model. (See Ito et al. (2021).)

This article proposes an estimation method to detect breakpoints for linear time series models with their parameters that jump scarcely. One usually supposes non-Gaussian processes like the combination of Gaussian and Poisson ones for the models; the author supposes no statistical specific noises in his estimation. His basic idea owes a sparse regression technique to regard them as linear regression models with possibly quite many parameters. Its most significant point is that scarcely changes in state variables are equivalent to a sparse structure of solutions for a transformed linear equation system obtained from the linear time series model by differencing parameters. The transformation allows the researchers to employ group LASSO techniques when they estimate quite many parameters supposed to jump.

The article’s organization is as follows. Section 2 reviews a basic idea. It shows how to transform a linear time series model into another with differenced parameters allowing group LASSO techniques. Finally, the section proposes the practical method to estimate a time series model with multiple parameters scarcely jumping over time. Section 3 shows an application of how we detect several structural changes in foreign exchange markets by using our method. In practice, the author employs the technique for the same data that he and his co-authors studied the foreign exchange markets in Ito et al. (2021). The result shows the periods in which structural changes happened in the markets in a sharper way than Ito et al. (2021). Section 4 concludes and shows several points for future research tasks.
2 Linear Time series Model with parameters that jump scarcely

There is literature about the estimation of time-varying coefficients of economic models. P. Swamy has published many articles on broadly theoretical frameworks providing several models with time-varying parameters since the 1970’s, for example, Swamy et al. (2010). However, we focus on a narrower class of models. Suppose $m \times n$ matrix $X_t$ and $m$-dimensional vector $y_t$ are given at each $t$.

$$y_t = X_t \beta_t + u_t, \quad (t = 1 \cdots T),$$

(1)

where $\beta_t$ is a $n$-dimensional vector coefficient to be estimated. Let $t$ denote time. Thus $T$ is the sample size when the data is set.

We can regard the above model as a linear regression model whose each coefficient $\beta$ varies over time. At this time, we do not have to consider $u_t$'s are gaussian. We can rewrite (1) in the following matrix form.

$$\begin{bmatrix}
y_1 \\
y_2 \\
y_3 \\
\vdots \\
y_T
\end{bmatrix} = 
\begin{bmatrix}
X_1 & O \\
X_2 & O \\
X_3 & \cdots \\
O & \cdots \\
O & X_T
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\vdots \\
\beta_T
\end{bmatrix}
+ 
\begin{bmatrix}
u_1 \\
u_2 \\
u_3 \\
\vdots \\
u_T
\end{bmatrix},$$

(2)

The data matrix in the regression (2) is $mT \times nT$ and the coefficient vector is of $nT$ dimension. Thus, usual regression solver such as OLS is inapplicable since the coefficient matrix in (2) is singular.

Now, we suppose that (1) is a regression model whose coefficients $\beta_t$'s quite rarely change. Then, we rewrite the linear equation system (1) into another linear system with respect to $\Delta \beta_t$ in place of $\beta_t$ to focus on sparsity. The following relations are my key idea.

$$\begin{align}
\beta_1 &= \beta_0 + \Delta \beta_1 \\
\beta_2 &= \beta_0 + \Delta \beta_1 + \Delta \beta_2 \\
\vdots \\
\beta_T &= \beta_0 + \Delta \beta_1 + \Delta \beta_2 + \cdots + \Delta \beta_T,
\end{align}$$

(3)

where we regard $\beta_0$ as the starting value given of a sequence of $\Delta \beta$'s.

From (3), we obtain its matrix form.

$$\begin{bmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\vdots \\
\beta_T
\end{bmatrix} = 
\begin{bmatrix}
I & O & O & \cdots & O \\
I & I & O & \cdots & O \\
I & I & I & \cdots & O \\
I & I & I & \cdots & I
\end{bmatrix}
\begin{bmatrix}
\Delta \beta_1 \\
\Delta \beta_2 \\
\Delta \beta_3 \\
\Delta \beta_T
\end{bmatrix}
+ 
\begin{bmatrix}
\beta_0 \\
\beta_0 \\
\beta_0 \\
\vdots
\end{bmatrix}$$

(4)
One can set $\beta_0$ as the coefficient vector for the usual time-invariant regression. (Please see Appendix for detail.)

We apply the group LASSO method to the following equation.

$$r = \tilde{X} \Delta \beta + u,$$

where $r$ is the residuals of a VEC model with time invariant parameters and $\tilde{X}$ is the matrix as shown in the following two slides.

$$r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_T \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} - \begin{bmatrix} X_1 & O \\ X_2 & \vdots \\ \vdots & \vdots \\ O & X_T \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_0 \\ \vdots \\ \beta_0 \end{bmatrix},$$

and

$$\tilde{X} = \begin{bmatrix} X_1 & O \\ X_2 & \vdots \\ \vdots & \vdots \\ O & X_T \end{bmatrix} \begin{bmatrix} I & O & O & \cdots & O \\ I & I & O & \cdots & O \\ I & I & I & \cdots & O \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ I & I & I & \cdots & I \end{bmatrix},$$

Where $I$ is the $n \times n$ identity matrix.

At this point, we provide our breakpoint detecting procedure.

**Step 1** We apply a group Lasso solver to the equation (5) by grouping variables according to periods.

- Both $r$ and each column of $\tilde{X}$ should be "normalized" before the above operation.

**Step 2** The group Lasso solver finds a series of $\Delta \beta_t$ whose all components are zeros for almost all periods.

**Step 3** We regard the period such that $t$ such that $\Delta \beta_t \neq 0$ as a break point.

Note that we can obtain each $\beta_t$ through the following equation

$$\beta_t = \hat{\beta}_0 + \sum_{r=1}^{t} \Delta \beta_r.$$

We call the linear equation system (5) a difference form of the linear time series model. Note that the transformation above from the original model (1) to the different form is quite algebraic and that the error terms are identical. Thus, the linear equation (5) as a statistical model is equivalent to the original state-space model corresponding to
the equation (1). Irrespective of an estimation method, the above two statistical models provide essentially the same results. One should contemplate which kind of noises they choose, Gaussian or non-Gaussian, when they employ a linear time series model given data \( y \) observed.

As the author stressed in Ito (2019), when we address state-space models covering the linear time series model, we have assumed Gaussian noises in many applications. At the same time, their error terms are not always supposed to follow Gaussian noises. Intuitively, under the assumption of Gaussian noise for the models, the sequence of estimated parameters tends to be smooth. In contrast to Gaussian noises, under the assumption of non-Gaussian noise, for instance, the Poisson process for a state-space model, the sequence of estimated state variables tends to be almost invariant and jump scarcely.

Potentially supposing broader classes of errors, the author regards LASSO (or other sparse regression solvers) as a suitable solver for a linear time series model with parameters that jump in scarce timings. LASSO effectively solves the corresponding difference form (5) even if the number of parameters amounts to more than ten thousand, as is shown in the next section. In this case, the group LASSO or generalized LASSO is applicable. (See Yuan and Lin (2006) and Tibshirani (2011) for detail of the two LASSO.)

3 Application to Foreign Exchange Markets

The author shows an application in which the method in Section 2 possibly detects several structural changes in a real financial market. The data is the same as Ito et al. (2021). They used monthly nominal data from the Thomson Reuters Datastream on the spot and forward exchange rates for three developed countries (Canada, Japan, and the UK) from May 1990 to July 2015 taken. (See Ito et al. (2021) for detail) They estimated a time-varying VEC model’s parameters, supposing them varying like random walks over time. They argue that the estimated degree of market comovement has increased over sample periods. At the same time, the behavior is not monotonous. One can regard the degree as the speed of adjustment if the deviation from the long-run equilibrium exists. Figures 1 and 2 summarize their findings.
Figure 1: Degree of comovement in foreign exchange rates (1990-2015) (from Ito et al. (2021))

Figure 2: Changes in the Degree of comovement (1990-2015) (from Ito et al. (2021))
This section aims to show possible breakpoints sharper than Ito et al. (2021) and in what periods they occurred. Ito et al. (2021) suggests that the exchange markets vary over time and that the changes are gradual. Assuming there are some cointegrating relationships, we consider a vector error correction (VEC) model for $m$-vector time series $X_t$.

$$
\Delta X_t = \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_k \Delta X_{t-k} + \Pi_k X_{t-k} + \mu + u_t, \quad (8)
$$

where $\Delta X_t = X_t - X_{t-1}$, $\mu$ is a vector of intercepts, and $u_t$ is a vector of error terms. The Coefficient matrix $\Pi_k$ represents some long-run relationships among the components of $X_t$’s. Regarding (8) as a linear time series regression (2) and supposing $\Pi_k$ varying over time, we apply our group LASSO technique. At this point, we show a practical procedure applied in this section as follows.

**Step 1** First, we estimate a usual VEC model supposing any coefficient time-invariant after some cointegration test.

**Step 2** Using its loading matrix $\Pi_k = \alpha \beta$, we specify $\beta$ matrix, say, $\beta^*$.

**Step 3** Considering $\Pi_k$ time-varying, we estimate $\Delta \Pi_{k,t}$ for each period $t$.

**Step 4** We obtain $\Pi_{k,t}$ through the cumulated sum of $\Delta \Pi_{k,t}$’s.

**Step 5** Supposing $\Pi_{k,t} = \alpha_t \beta^*$, we calculate $\alpha_t, t = 1, \cdots$

**Step 6** We obtain our degree of comovement from $\alpha_t$’s in the same way as Ito et al. (2021).

We provide here practical information about our estimation. The sample size = 306; the number of variables = 6. According to Ito et al. (2021), we choose $k = 2$ in (8), the total number of parameters is 11016 when we attempt to obtain time-varying loading matrices to calculate our time-varying degree. We need to try many values of the hyper-parameter of the group LASSO model to find some breaks. It takes about 1800 seconds on my Mac mini (Late 2014) with Intel Core i7 (16 GB 1600 MHz DDR3 memory).

Figures 3 and 4 summarize this paper’s findings as Figures 1 and 2. Moreover, the results of breakpoint are summarized as follows.

**1st Impact** The degree is non-zero between May 1990 and July 1994 and attains a local maximum at December 1993.

**2nd Impact** The degree is non-zero between May 2007 and July 2011 and attains a local maximum at August 2008.

Notice that the first break possibly occurred before May 1990, and the left peak in the figure possibly reflects the UK pound crisis in September 1992 as well as the UK’s rapid economic recovery after the crisis. The right peak in the figure demonstrates Lehman Brothers Holdings Inc.’s bankruptcy on September 15, 2008.
Figure 3: Degree of comovement in foreign exchange rates (1990-2015)

Figure 4: Changes in the Degree of comovement (1990-2015)
4 Concluding Remarks

Here the author provides some remarks. We have shown a method to estimate linear time series models with time-varying parameters (TVP) that jump in very few periods. Using group LASSO allows us to detect breakpoints when addressing multivariate linear models. The robustness relies on not any statistical technique but a simple, sparse technique that guarantees sparse structure parameters. This paper’s result does not reflect the Asian financial crisis from July 1997. It could come from our choice of data: Canada, Japan, UK. Ito et al. (2021) also detect two peaks of the changes in the Degree of comovement while the "resolution" is lower than this paper. The two estimates of Ito et al. (2021) and this paper, which estimate is better? It depends on what information the users want.

Finally, we show some future tasks: more statistical inference more applications to evaluate our framework’s effectiveness. We should study correspondences between our detected breaks and possible actual shocks in practice. We are searching for better forecast performance by focusing on the periods after the last breakpoint by our technique.
References

Basu, S., Michailidis, G., et al. (2015), “Regularized estimation in sparse high-dimensional time series models,” The Annals of Statistics, 43, 1535–1567.

Chan, N. H., Yau, C. Y., and Zhang, R.-M. (2014), “Group LASSO for structural break time series,” Journal of the American Statistical Association, 109, 590–599.

Ito, M. (2019), “An Estimation Method for State Space Models using generalized LASSO Techniques,” Reported in 94th annual conference of WEAI.

Ito, M., Noda, A., and Wada, T. (2021), “Time-Varying Comovement of Foreign Exchange Markets: A GLS-Based Time-Varying Model Approach,” Mathematics, 9, 849.

Michailidis, G. and d’Alché Buc, F. (2013), “Autoregressive models for gene regulatory network inference: Sparsity, stability and causality issues,” Mathematical biosciences, 246, 326–334.

Swamy, P., Tavlas, G. S., Hall, S. G., and Hondroyiannis, G. (2010), “Estimation of parameters in the presence of model misspecification and measurement error,” Studies in Nonlinear Dynamics & Econometrics, 14.

Tibshirani, R., Wainwright, M., and Hastie, T. (2015), Statistical learning with sparsity: the lasso and generalizations, Chapman and Hall/CRC.

Tibshirani, R. J. (2011), The solution path of the generalized lasso, Stanford University.

Yuan, M. and Lin, Y. (2006), “Model selection and estimation in regression with grouped variables,” Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68, 49–67.
Appendix

One can regard $\beta_0$ as a prior if they are a Bayesian. But, I propose the least square estimate of the following linear regression model to obtain $\beta_0$ conveniently with our givens.

$$\begin{bmatrix}
y_1 \\
y_2 \\
y_3 \\
\vdots \\
y_T
\end{bmatrix} = \begin{bmatrix}
X_1 \\
X_2 \\
X_3 \\
\vdots \\
X_T
\end{bmatrix} \beta + \begin{bmatrix}
\varepsilon_1 \\
\varepsilon_2 \\
\varepsilon_3 \\
\vdots \\
\varepsilon_T
\end{bmatrix}. \quad (9)$$